DATA-DRIVEN SUB-GRID MODEL DEVELOPMENT FOR LARGE EDDY SIMULATIONS OF TURBULENCE

By

ROMIT MAULIK

Bachelor of Engineering in Mechanical Engineering.
Birla Institute of Technology
Mesra, India
2012

Master of Science in Mechanical & Aerospace Engineering
Oklahoma State University
Stillwater, OK
2015

Submitted to the Faculty of the
Graduate College of
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
May, 2019

DATA-DRIVEN SUB-GRID MODEL DEVELOPMENT
FOR LARGE EDDY SIMULATIONS OF TURBULENCE

Dissertation Approved:

Omer San, PhD.

Dissertation Advisor

Balaji Jayaraman, PhD.

Arvind Santhanakrishnan, PhD.

Ja Eun Ku, PhD.

ACKNOWLEDGMENTS

This dissertation is dedicated to my parents, my partner Choi, my adviser Dr. San and my cat Gauss. Thank you for your unflinching support and for your investments in my future. In addition, I would like to express my gratitude towards the members of my committee, Dr. Jayaraman, Dr. Santhanakrishnan and Dr. Ku for their guidance over the past few years. Much of what is in this document was made possible by productive interactions with my fellow graduate students in MAE. Thank you all once again.

I am also grateful for teaching opportunities provided to me by MAE. Financial assistance from the Oklahoma State University Graduate College, Oklahoma NASA EPSCOR, Nvidia Corporation, ASHRAE, APS, SIAM and the US DOE Office of Science are deeply appreciated.

Acknowledgments reflect the views of the author and are not endorsed by committee members or Oklahoma State University.

Name: ROMIT MAULIK

Date of Degree: May, 2019

Title of Study: DATA-DRIVEN SUB-GRID MODEL DEVELOPMENT FOR LARGE EDDY SIMULATIONS OF TURBULENCE

Major Field: MECHANICAL & AEROSPACE ENGINEERING

Abstract: Turbulence modeling remains an active area of research due to its significant impact on a diverse set of challenges such as those pertaining to the aerospace and geophysical communities. Researchers continue to search for modeling strategies that improve the representation of high-wavenumber content in practical computational fluid dynamics applications. The recent successes of machine learning in the physical sciences have motivated a number of studies into the modeling of turbulence from a data-driven point of view. In this research, we utilize physics-informed machine learning to reconstruct the effect of unresolved frequencies (i.e., small-scale turbulence) on grid-resolved flow-variables obtained through large eddy simulation. In general, it is seen that the successful development of any data-driven strategy relies on two phases - learning and a-posteriori deployment. The former requires the synthesis of labeled data from direct numerical simulations of our target phenomenon whereas the latter requires the development of stability preserving modifications instead of a direct deployment of learning predictions. These stability preserving techniques may be through prediction modulation - where learning outputs are deployed via an intermediate statistical truncation. They may also be through the utilization of model classifiers where the traditional L_2-minimization strategy is avoided for a categorical cross-entropy error which flags for the most stable model deployment at a point on the computational grid. In this thesis, we outline several investigations utilizing the aforementioned philosophies and come to the conclusion that sub-grid turbulence models built through the utilization of machine learning are capable of recovering viable statistical trends in stabilized a-posteriori deployments for Kraichnan and Kolmogorov turbulence. Therefore, they represent a promising tool for the generation of closures that may be utilized in flows that belong to different configurations and have different sub-grid modeling requirements.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER I

Introduction

Turbulent flows are encountered in a variety of systems, both natural and engineering, involving a wide range of spatial and temporal scales. In order to capture all the scales of a turbulent flow, all the scales are required to be resolved in a direct numerical simulation (DNS), where the full spectra of turbulence are resolved down to the Kolmogorov scale. However, the extreme computational and storage expenses associated with a DNS for any application of practical interest precludes their widespread use for anything beyond the simplest of problems. Therefore a significant amount of attention is devoted to the development of large eddy simulation (LES) techniques which aim to capture the large energy containing motions of any energetic flow while modeling sub-grid quantities (Lesieur and Metais, 1996; Pope, 2004; Sagaut, 2006). LES hypothesizes the existence of solutions in a transformed space which is low-pass spatially-filtered which consequently represent the large scale motions which contain most of the energy in a turbulent flow. However, the numerical evolution of the LES governing equations is fraught with inaccuracies associated with the discarding of the finer scales due to the highly non-linear nature of turbulence with multiscale interactions in wavenumber space. Therefore, a direct evolution of the LES equations needs to be coupled with some measure of small-scale modeling. This dissertation will investigate data-driven methods for estimating small-scale contributions in LES.

The chief interest in pursuing LES research is the greater insight it provides into the physics of a particular flow phenomena as against the more widely-used Reynolds-Averaged Navier-Stokes equations (RANS) which model almost all relevant scales of

the flow. This is not to imply that RANS is 'obsolete': in most flows of engineering interest RANS is almost always the tool of choice for the practitioner (Hanjalic, 2005) and it is widely understood that a greater integration of CFD into an engineering design workflow needs a leveraging of both RANS and LES approaches in a hybrid fashion (Spalart, 1997; Fröhlich and Von Terzi, 2008). However, the main focus of this research will be to develop novel closure strategies for LES.

There are two main ideologies in the turbulence modeling community. Both approaches are devised with the end goal of modeling *sub-grid* quantities. To be particular, the goal of these techniques is to model the effect of these unresolved quantities on the larger scales as accurately as possible. This is in accordance with well-established turbulence theory which has confirmed that there are strong non-linear interactions between different scales through the formation of an energy cascade (Kolmogorov, 1941*b*,*a*; Orszag, 1970; Kraichnan, 1967). Therefore, the primary test for a sub-grid model is whether it is able to faithfully reproduce the spectral scaling expected in two- and three-dimensional turbulence. The functional approach to turbulence modeling specifies the use of an *eddy-viscosity* as a function of grid-resolved variables and is the predominant approach for LES in practical applications. Eddy-viscosity approaches are consistent with Kolmogorov's ideas about the energy spectrum of three-dimensional (3D) homogeneous isotropic turbulence where energy is injected into the flows at large scales and is gradually transferred by a non-linear cascade to the smaller scales until viscous dissipation becomes dominant (Frisch, 1995). Most functional sub-grid models are derived from the oft-used Smagorinsky approach (Smagorinsky, 1963) which computes an eddy-viscosity from the resolved strain-rate tensor and characteristic length scales (generally determined by grid size) with the additional specification of a Smagorinsky constant (C_s). However, while the Smagorinsky constant is proven to have a particular value for homogeneous and isotropic turbulence ($C_s \approx 0.18$), its application to various practical flows has shown that it is not single-valued. This is re-

peating phenomenon in many *static* eddy-viscosity models (for instance the popular models specified in Vreman (2004) and Nicoud and Ducros (1999)) which necessitate coefficient tweaking for good agreement with truth. Consequently there have been several attempts to dynamically update the value of this constant (Germano et al., 1991; Lilly, 1992) through the solution of a least-squares problem at each time step between grid cut-off and a prespecified test-filter cut-off wavenumber by leveraging the concept of scale-similarity. We shall utilize comparisons to the dynamic implementation of the Smagorinsky eddy-viscosity kernel. However, this introduces another ad-hoc user defined parameter, namely, the choice of the filter and its characteristic width. We must also note that the functional form relating C_s to a 'turbulent eddy viscosity' has been generated from mixing length theory rather empirically. One aspect of this dissertation shall provide a framework to bypass the restrictive empirical assumptions involved in the calculation of the linear eddy-viscosity through the extraction of trends from high-fidelity DNS snapshots.

An alternative method of closure modeling focuses on bypassing the eddy-viscosity hypothesis through structural arguments. A structural argument may be based on scale-similarity (Bardina et al., 1980; Layton and Lewandowski, 2003) which attempts to augment the quadratic non-linearity the LES governing equations with an explicit low-pass spatial filtering procedure. Structural arguments may also be based on the hypothesis that the LES space admits solutions that are continuous low-pass spatially filtering counterparts of DNS. In this scenario, an inverse filter is assumed to exist and is utilized for reconstructing the effect of the finer frequencies. This forms the core assumption of the approximate deconvolution (AD) method proposed by (Stolz and Adams, 1999). While there are limitations to the hypothesis utilized in structural arguments, the chief being the assumption of isomorphism between DNS and LES spaces (necessitating a finer grid than what is required for the eddy-viscosity techniques), they are attractive as they bypass the Bousinessq hypothesis (i.e., a tur-

bulence stress tensor that is linear in the velocity gradient). In order to deploy the structural methods accurately, one must specify a low-pass spatial filtering operator with a specific transfer function. It is generally seen that the choice of this filter is crucial for a-posteriori closure accuracy and requires tuning similar to that necessary for functional coefficients. Traditional structural applications have prescribed Gaussian-type filters for this iterative procedure which is a relatively strong (and often incorrect) assumption when homogeneous and isotropic turbulence is observed. Therefore, the utilization of these methods in practical deployments necessitates coupling with functional techniques (Habisreutinger et al., 2007; Chow et al., 2005) which damp out errors from the previously mentioned Gaussian assumption. While this does not eliminate the primary drawback we are attempting to address (i.e. the specification of user-defined parameters or filters), it is often observed that the strengths of both approaches complement each other to mitigate weaknesses. A significant component of this research shall also study the devising of data-driven filters that are not limited by the Gaussian assumption to engineer a bypass to the limitations of the isomorphic assumption.

As mentioned in the previous paragraph - a widely utilized strategy for successful closure modeling in LES is to combine structural and functional arguments. The core argument for this method is that grid-support relevant losses may be accounted for using functional arguments and areas of the flow field which are relatively well resolved may be tackled by structural techniques. One can immediately notice a limitation to this method. Basically, both methods are always *active* at any given location in a flow field and therefore contaminate their underpinning hypotheses. This gives us an opportunity to predict areas in a flow field that may be better suited to a certain turbulence modeling strategy. This dissertation will also study the use of classifiers to blend turbulence models adeptly. The chief promise in these studies will be the advantage of using the large body of turbulence modeling literature as-is

for complementing literature. In later chapters, we study techniques that switch as well as blend between models for improved statistical predictions of two-dimensional turbulence. We also utilize classifier based techniques to switch between central and upwinded derivative calculations (with the latter utilized for imparting numerical dissipation to the flow field). Note that the use of upwinded schemes for turbulence modeling is denoted *implicit* LES and is another popular strategy for modeling the (primarily) dissipative nature of the finer scales in a turbulent field (Grinstein et al., 2007).

This brings us to the primary motivation for the undertaking of this research. Over the past decade, advances in data collection and increasing access to computational resources have led to a revolution in the use of data-driven techniques for the solution of intractable inverse problems. One such problem is that of turbulence where the phenomenon of scale separation causes infeasible computational demands even for the most simple systems. This behavior is shared by all non-linear partial differential equations and necessitates the utilization of multiple modeling approximations such as sub-grid modeling for LES for tractable compute times. With this contextual background, in this document, we introduce hybrid modeling (physics-informed machine learning) methodologies for determining (or combining) sub-grid models without any phenomenological assumptions (in the spirit of structural models). This is accomplished by the use of artificial neural networks (ANNs) to establish data-driven maps between coarse-grained LES fields and truth extracted from associated DNS but without the use of any explicit filter or model-form coefficient. These optimal maps between grid-resolved and sub-grid quantities (or between the former and the optimal choice of turbulence model) are obtained by supervised learning from sub-sampled direct numerical simulation (DNS) data and are deployed in both a-priori and a-posteriori fashion for the LES of two-dimensional and three-dimensional turbulence. The utilization of DNS data also conforms to the ideology of optimal LES

based modeling strategies (Langford and Moser, 1999; Moser et al., 2009). To that end, we have identified the following components of this research direction. In general, the development of a data-driven closure is specified through

1. Generation of high-fidelity data snapshots through DNS. These snapshots are highly sub-sampled in space and time to reflect measurement deficiences.

2. Synthesizing LES data that mimics the loss of grid-support in LES space through spectrally accurate cut-off filtering. This data may be assumed to be that obtained from a *perfect* closure.

3. Identifying a relationship conduit between the perfect closure and the grid-resolved data. This is generally done through the identification of the commutation error in the non-linear term. This error is recast into an eddy-viscosity for functional models or a filter deficiency in structural techniques. This preserves the *physics-informed* nature of these learning strategies in that we steer clear of directly predicting an entire flow-field. Alternatively, we may also identify the conditional-probability of the most appropriate dissipation strategy for this synthetic data.

4. Using data-driven methods to learn a correction to this error. This may be through the learning of a low-pass spatial filter that mimics the loss of grid-support or the prediction of an effective eddy-viscosity. The learning may also be utilized as a classifier for the most optimal model or gradient discretization strategy among a series of choices.

5. Assessment through forward deployment of the learning in the presence of numerical errors associated with LES governing equation evolution. Assessments are generally made through statistical quantities such as probability density functions, spatially-averaged kinetic energy spectra and structure functions.

This dissertation will be devised with individual chapters corresponding to various studies that implement our approach to closure development using machine learning. Each chapter implements our ideology towards closure development in a standalone fashion (i.e, the points in the aforementioned list are adhered to as far as possible). One-line summaries of the various chapters in this document are given in the enumerated list below -

1. Chapter 2 assesses the utility of ANNs for estimating an effective inverse filtering operation from coarse-grained data in an a-priori study.

2. Chapter 3 extends the techniques introduced in chapter 1 by deploying data-driven forward and inverse filters in an a-posteriori study.

3. Chapter 4 studies the performance of deep ANN's for direct prediction of a perfect closure in both a-priori and a-posteriori studies.

4. Chapter 5 is similar to the previous chapter but the focus of the prediction is an eddy-viscosity that mimics the contribution of a perfect closure to preserve pressure, time and Galilean invariance.

5. Chapter 6 utilizes ANNs as classifiers for dynamically switching between AD, static Smagorinsky and no-model deployments (where the latter choice indicates no necessity for turbulence modeling) in a-posteriori assessments.

6. Chapter 7 is similar to chapter 6 but utilizes ANNs to dynamically switch between central and upwinded schemes for the quadratic non-linearity thereby imparting a numerical dissipation in accordance with turbulence modeling requirements.

CHAPTER II

A neural network approach for the blind deconvolution of turbulent flows

2.1 Abstract

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.

2.2 Introduction

In the past few decades, an exponential increase in computational power, algorithmic advances and experimental data collection strategies have seen an explosion in modeling efforts which leverage information obtained from physical data. The ability

This chapter is adapted from Maulik and San, J. Fluid Mech., 831, 151-181.

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 and 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 modeling 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 and Koumoutsakos, 2002; Gautier et al., 2015) but there has been recent interest in their utilization for sub-grid scale modeling for high Reynolds (Re) number flows. For instance, artificial neural networks (ANNs) have recently been utilized for the purpose of sub-grid scale model classification (Gamahara and Hattori, 2017) and Reynolds stress anisotropy tensor prediction (Ling, Kurzawski and Templeton, 2016; Wang, Wu and Xiao, 2017). Together with other machine learning based techniques such as Gaussian process regression (Raissi and Karniadakis, 2017, 2016; Zhang and Duraisamy, 2015; Wang et al., 2016), field inversion (Duraisamy et al., 2015; Parish and Duraisamy, 2016), symbolic regression (Weatheritt and Sandberg, 2016; Brunton et al., 2016) and several different algorithms (Ling and Templeton, 2015; Ling, Jones and Templeton, 2016; Tracey et al., 2013; Bright et al., 2013), data-driven approaches such as ANNs are poised to form the backbone of the next great leap in closure modeling for non-linear 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 and 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 modeling. In this work, we have developed a single-layer feed-forward ANN to identify a non-linear

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 behavior 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 behavior.

In particular, the approach outlined in our study may be considered to be analogous to the *approximate deconvolution* methodology (Stolz and 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 et al., 2001). A potential advantage to emphasize here is that a data-driven closure could be incorporated to *learn* the non-linear 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 et al., 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*b*) 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 clo-

sures which require an explicitly specified low-pass spatial filter (Layton and 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 artifact of data-localization.

2.3 Artificial Neural Networks

2.3.1 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)

$$\mathbf{S}^l = \mathbf{W}^l \mathbf{X}^{l-1}, \tag{2.1}$$

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

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

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

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

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 et al., 1998), however, only a small number of functions which are bounded, monotonically increasing and differentiable are used for this purpose.

2.3.2 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 et al. (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 Section 2.3.1. Our input matrix is

$$\mathbf{X}^0 = \begin{bmatrix} \mathbf{x}_1^0 & \mathbf{x}_2^0 & \dots & \mathbf{x}_{n_s}^0 \end{bmatrix}, \tag{2.4}$$

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

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

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

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

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

$$\mathbf{W}^2 = \begin{bmatrix} \mathbf{w}_1^2 & \mathbf{w}_2^2 & \dots & \mathbf{w}_Q^2 \end{bmatrix}, \tag{2.7}$$

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

$$\mathbf{S}^2 = \mathbf{W}^2\mathbf{H}^\intercal, \tag{2.8}$$

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

$$\mathbf{T} = \begin{bmatrix} \mathbf{t}_1 & \mathbf{t}_2 & \dots & \mathbf{t}_{n_s} \end{bmatrix}, \tag{2.9}$$

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

$$\mathbf{W}^2_{op}\mathbf{H}^\intercal = \mathbf{T}, \tag{2.10}$$

or by taking a transpose of both sides

$$\mathbf{H}\mathbf{W}^{2\intercal}_{op} = \mathbf{T}^\intercal, \tag{2.11}$$

which leads us to the following expression for the optimal weights

$$\mathbf{W}^{2\intercal}_{op} = \mathbf{H}^\dagger\mathbf{T}^\intercal. \tag{2.12}$$

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 and Hagan, 1997) in favor 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 et al., 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 $\mathbf{H}^\intercal = f(\mathbf{W}^1\mathbf{X}^0 + \mathbf{B}^1)$	▷ Tan-Sigmoid activation
4: Calculate pseudoinverse $\mathbf{H}^\dagger$	▷ Moore-Penrose pseudoinverse
5: Calculate $\mathbf{W}_{op}^{2\intercal} = \mathbf{H}^\dagger\mathbf{T}^\intercal$	▷ 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

2.4 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 and 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:

$$\tau_{ij} = \bar{u}_i\bar{u}_j - \overline{u_i u_j}, \tag{2.13}$$

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

$$\tau_{ij} = \bar{u}_i \bar{u}_j - \overline{u_i^* u_j^*}, \tag{2.14}$$

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:

$$u_i' = G(\mathbf{x}; \sigma) u_i, \tag{2.15}$$

which utilizes the following Gaussian kernel with standard deviation σ (held at a

default value of 1.0):

$$G(\mathbf{x};\sigma) = \frac{1}{(\sqrt{2\pi}\sigma)^d}\exp(-\frac{|\mathbf{x}|^2}{2\sigma^2}), \tag{2.16}$$

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:

$$u_i' = u_i + \mu\kappa, \tag{2.17}$$

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 are 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 2.1.

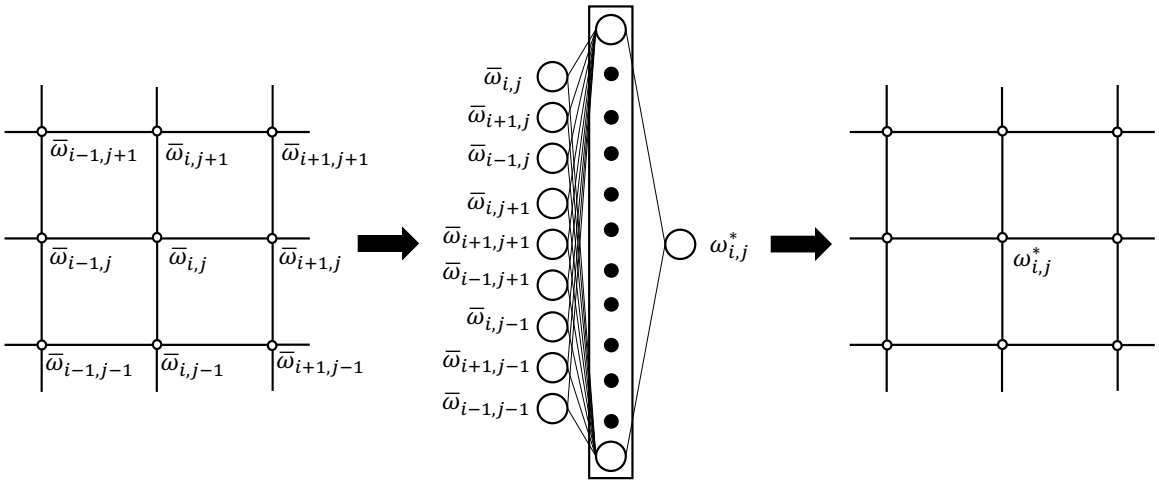


Figure 2.1: A schematic of the proposed blind deconvolution strategy for a two-dimensional test case where we aim to recover ω^* from $\bar{\omega}$. Note that each hidden neuron is associated with a unique bias.

2.5 Cross-Validation

An important assessment of a data-driven modeling 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.

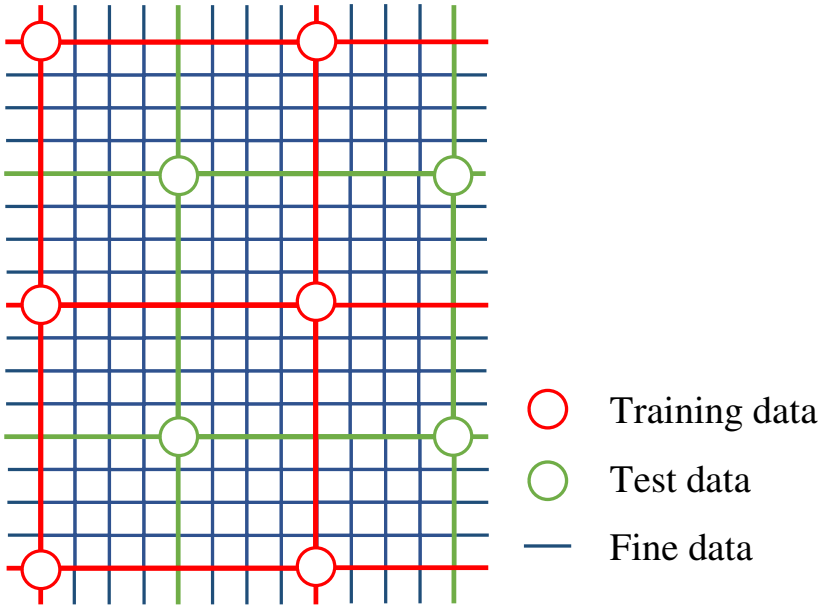


Figure 2.2: A schematic of the spatial shifting strategy for a simplified two-dimensional grid showing two different data sets. Several different data sets may be constructed using this technique which are numerically different physically similar.

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 2.2 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 upto 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 behavior demanded of 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

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 2.1: Cross-validation data sets for the proposed data-driven blind deconvolution closure.

to the network. Three testing data sets are generated in a similar manner with one of these data being filtered with the same filter radius and the others being filtered with a 10% larger and 10% smaller filter radius. 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 2.1. Note that these data sets are all generated from the same high fidelity solution field and correspond to a perfectly a-priori analysis. A concise summary of this cross-validation procedure is given in a flow chart in Figure 2.3.

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 Section 2.7.

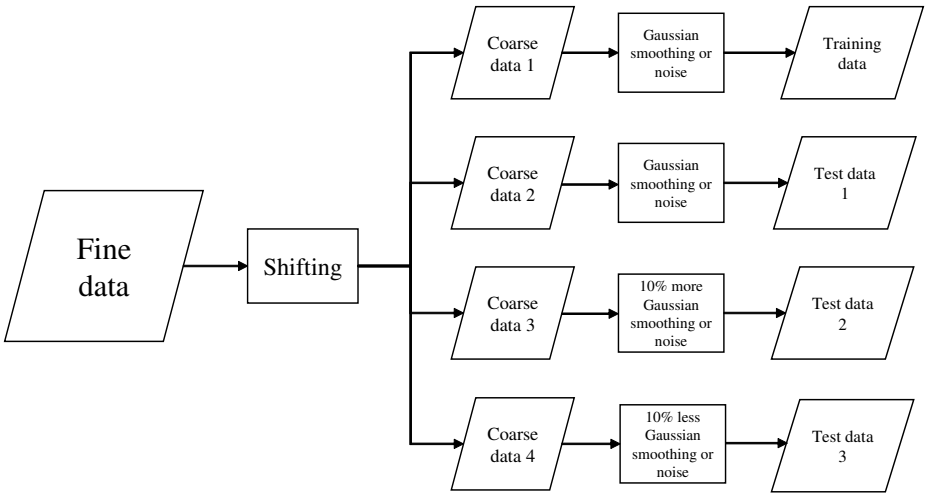


Figure 2.3: A flow chart detailing the generation of cross-validation data sets using the shifting operation as well as different perturbations. The network is trained using Algorithm 1 and is then tested on test data 1, 2 and 3.

2.6 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.

2.6.1 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 and 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} Kraich-

nan 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 neighbors (i.e., see Figure 2.1). Training times were (on average) of the order of 0.01 seconds for approximately 65,000 data samples.

Figure 2.4 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 behavior in a local neighboring 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 2.5) but the regularizing nature of the ELM training ensures that the cut-off length scale pile up of energy (or aliasing error) is successfully stabilized. This behavior 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.

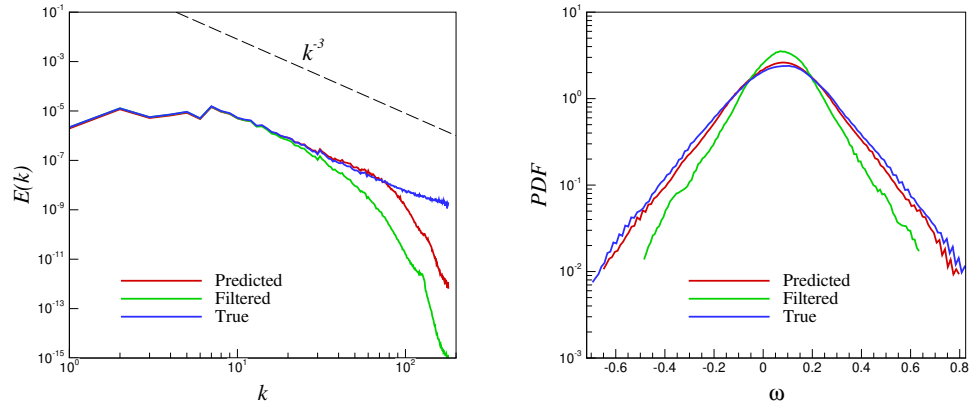
Kraichnan turbulence		
	Deconvolution	
Dataset	Filtered	Deconvolved
Test data 1	1.02×10^{-2}	4.38×10^{-3}
Test data 2	1.14×10^{-2}	5.73×10^{-3}
Test data 3	9.11×10^{-3}	3.71×10^{-3}
	Regularization	
Dataset	Noised	Regularized
Test data 1	4.03×10^{-2}	1.26×10^{-2}
Test data 2	4.88×10^{-2}	1.38×10^{-2}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

Table 2.2: 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.

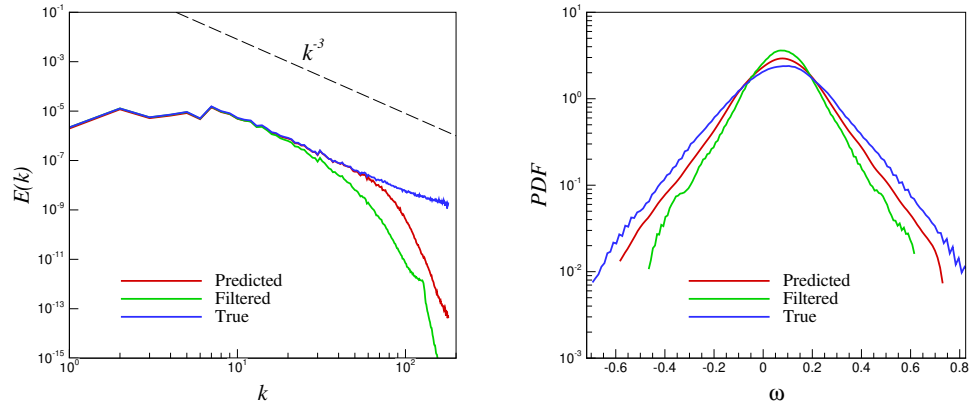
Figure 2.6 and Figure 2.7 visually represent the results of our proposed architecture. The enhancement of smaller features (due to a higher retention of the inertial range) in the case of the filtered inputs is clearly visible. On the other hand, a blurring effect can be discerned for the noisy perturbations 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 2.2 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.

2.6.2 Kolmogorov Turbulence

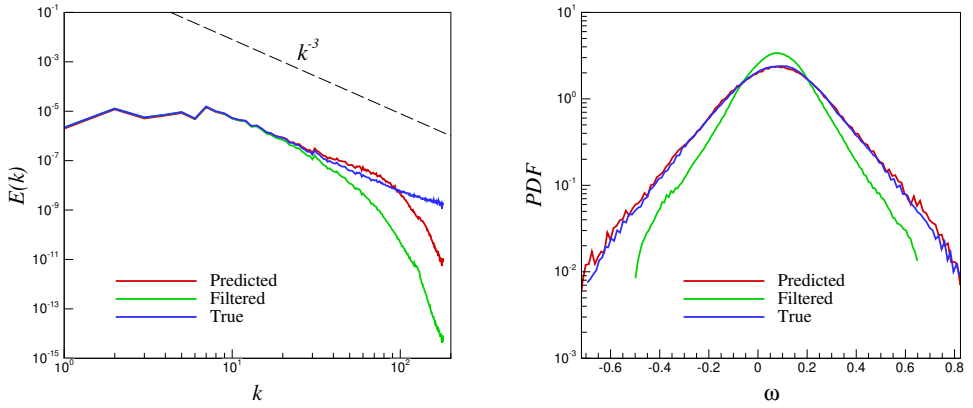
To further validate our claims about the potential of the proposed approach, the Taylor-Green vortex (TGV) problem was simulated in a periodic box l, following Bull and Jameson (2015), to exhibit the properties of decaying isotropic homogeneous 3D



(a) Test data 1

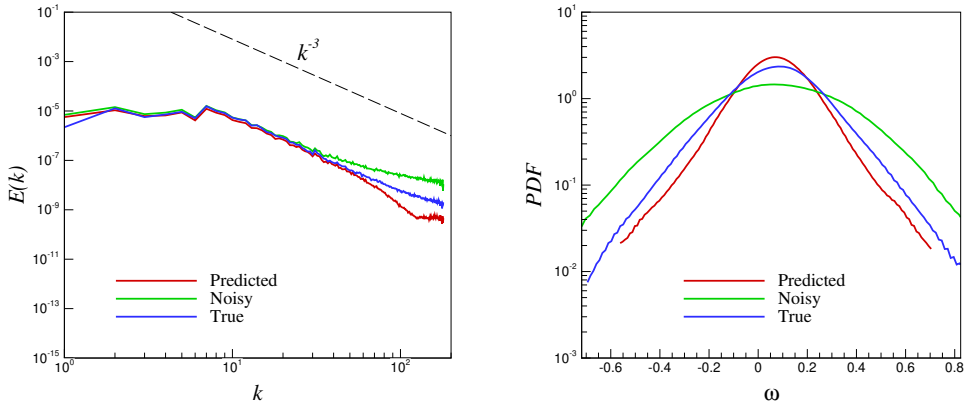


(b) Test data 2

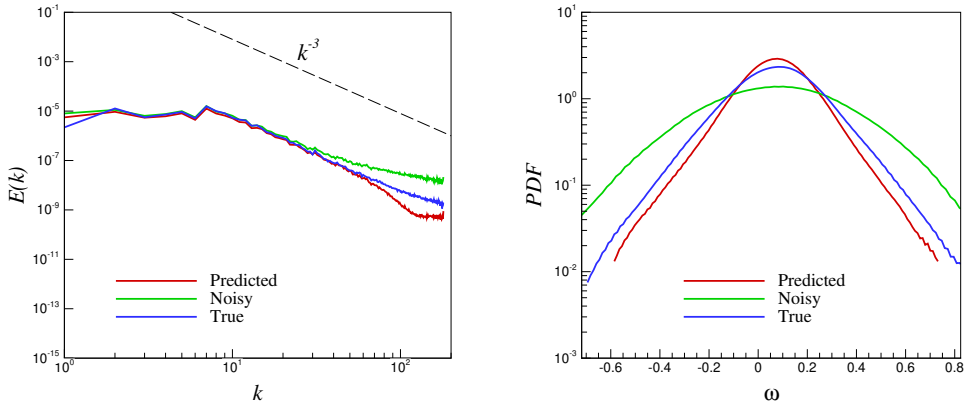


(c) Test data 3

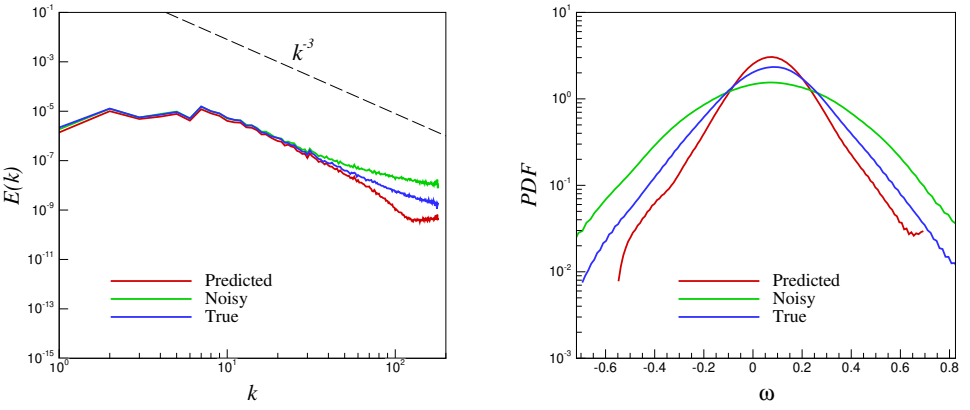
Figure 2.4: A-priori results of the kinetic energy spectra (left) and PDF of the vorticity (right) for Kraichnan turbulence. Results for three different deconvolution test data sets shown.



(a) Test data 1

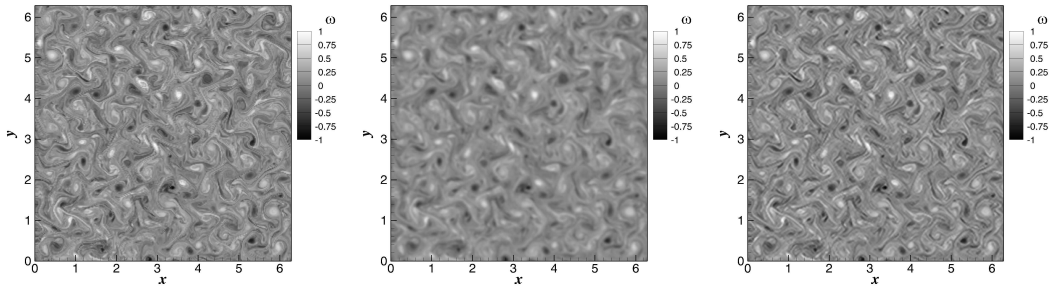


(b) Test data 2



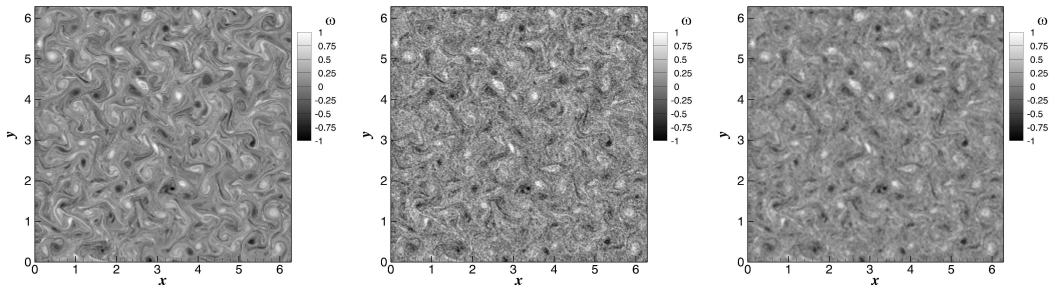
(c) Test data 3

Figure 2.5: A-priori results of the kinetic energy spectra (left) and PDF of the vorticity (right) for Kraichnan turbulence. Results for three different regularization test data sets shown.



(a) True (b) Filtered (c) Recovered

Figure 2.6: 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 proposed framework.



(a) True (b) Noisy (c) Recovered

Figure 2.7: 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 proposed framework.

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 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 2.8 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 non-linear in physical space as against the linear parameter space we attempt to explore. Essentially, increasing the filter radius by 10% leads to some physical behavior 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 behavior exhibited by our

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 2.9. 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 behavior 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.

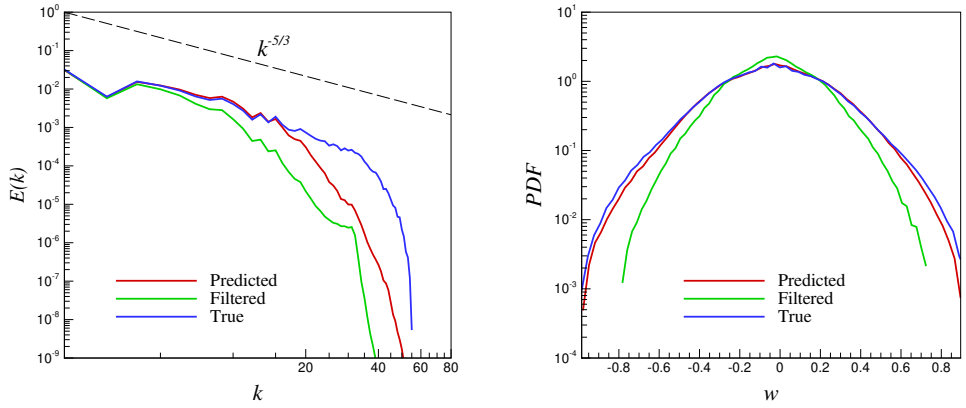
We clarify that the trends in the PDF of other primitive variables are similar. The blind deconvolution ability of the proposed methodology is shown in Figure 2.10 where one can clearly observe the presence of finer structures in the reconstructed field. Also, an examination of the noisy inputs in Figure 2.11 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 favor 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

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	Noised	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}

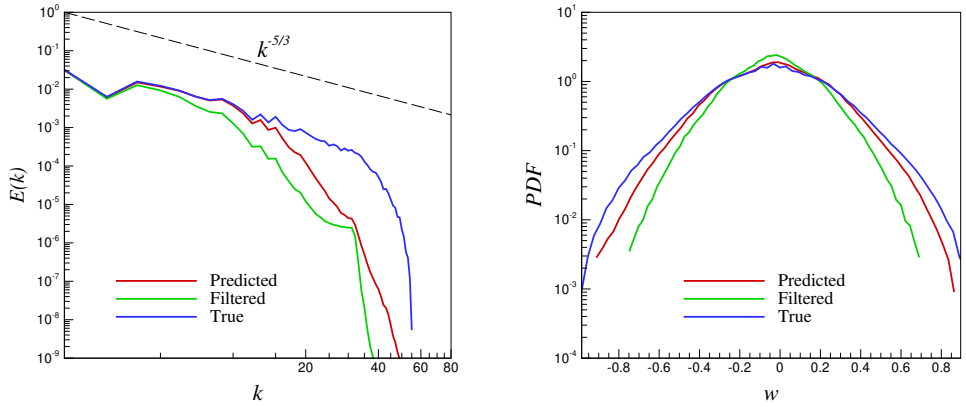
Table 2.3: Mean-squared-error values for deconvolved and regularized fields obtained from the proposed architecture. Data shown from the three-dimensional Kolmogorov turbulence test case. Note that the mean-squared-error values are obtained from the z component of the velocity field.

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 2.3 offers a quantitative insight into the performance of the proposed closure where it is once again observed that the framework is sufficiently capable in reducing the mean-squared-error of the perturbed fields for all testing datasets.

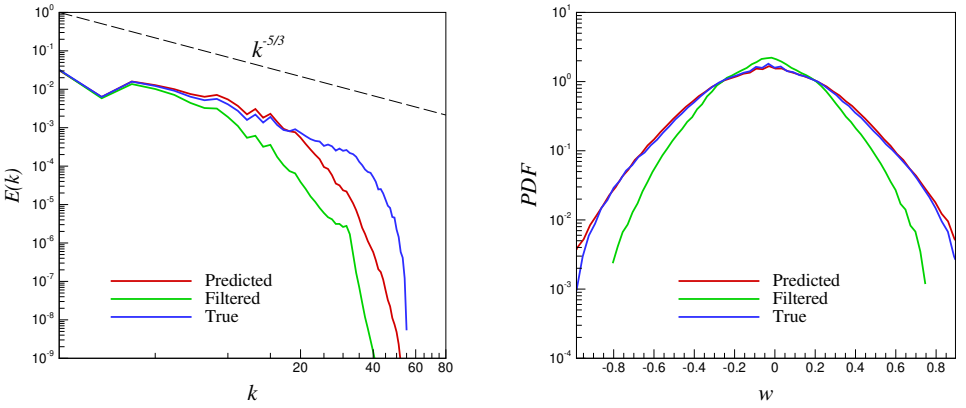
A comparison of the proposed closure with the state of the art structural closure models is shown in Figure 2.12 by using probability density functions for an assessment of deviatoric subfilter stress recovery performance. For the purpose of benchmarking we utilize a set of closures including the scale-similarity approach proposed by Bardina et al. (1980) (denoted SS), the approximate deconvolution methodology given by Stolz and Adams (1999) with 3 iterative deconvolutions (which forms the conceptual analog of our proposed architecture and is denoted as AD^3) as well as the scale-similarity approach proposed by Layton and Lewandowski (2003) which may be interpreted to be an approximate deconvolution methodology which 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



(a) Test data 1

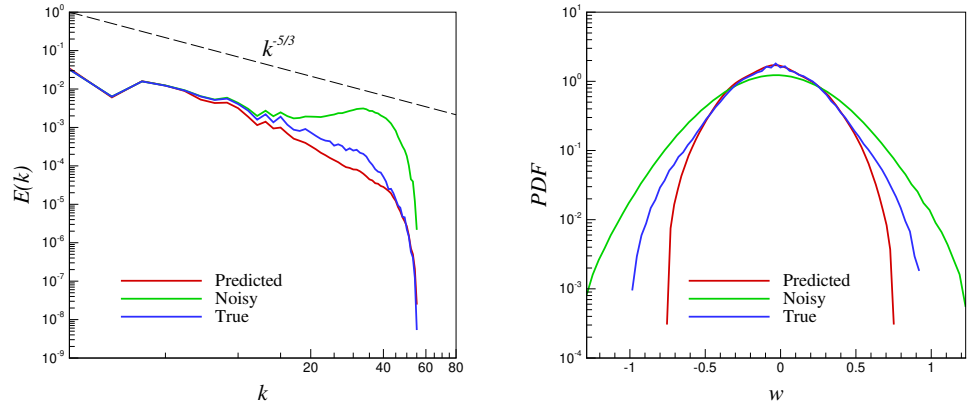


(b) Test data 2

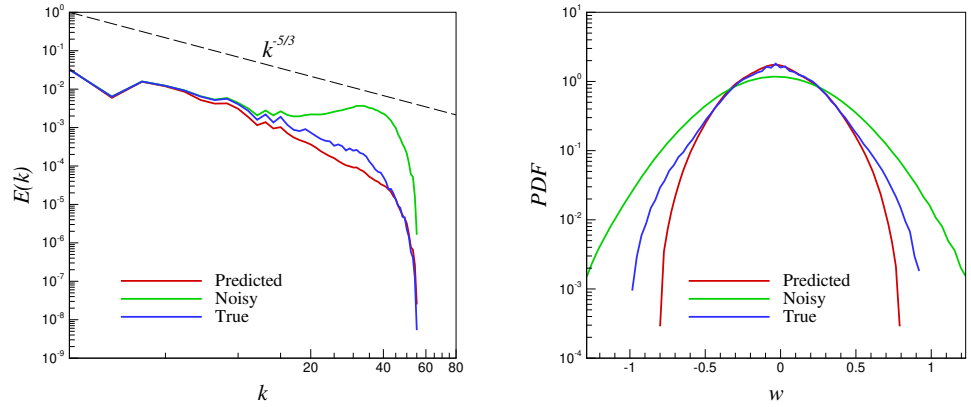


(c) Test data 3

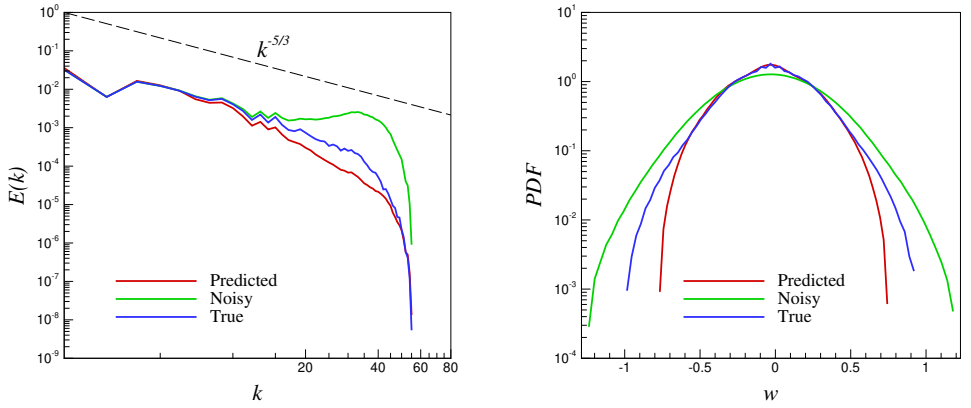
Figure 2.8: A-priori results of the kinetic energy spectra (left) and PDF of the z component of velocity (right) for Kolmogorov turbulence. Results for three different deconvolution test data sets shown.



(a) Test data 1

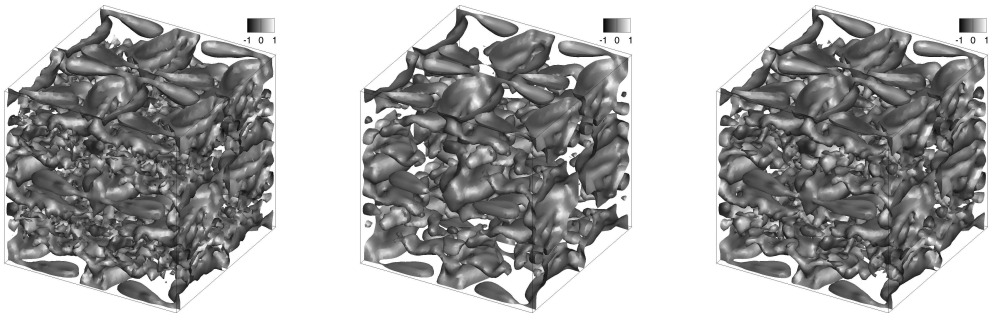


(b) Test data 2



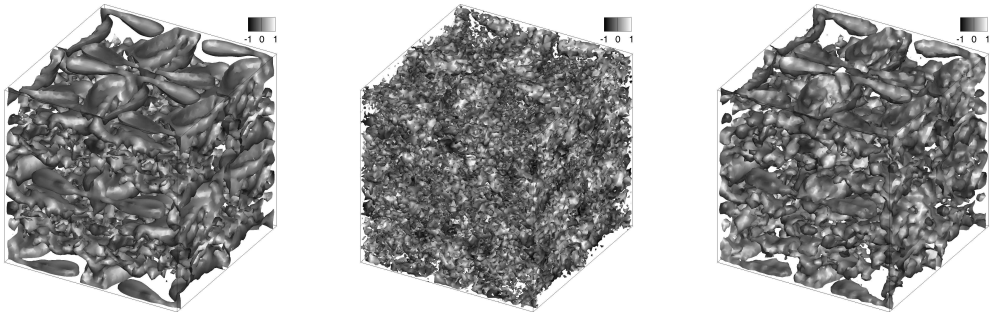
(c) Test data 3

Figure 2.9: A-priori results of the kinetic energy spectra (left) and PDF of the z component of velocity (right) for Kolmogorov turbulence. Results for three regularization different test data sets shown.



(a) True (b) Filtered (c) Recovered

Figure 2.10: A-priori results for velocity field recovery from low-pass spatially filtered perturbations for Kolmogorov turbulence. Isosurfaces for x-component of the velocity colored 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 proposed framework.



(a) True (b) Noisy (c) Recovered

Figure 2.11: A-priori results for velocity field recovery from noisy perturbation inputs for Kolmogorov turbulence. Isosurfaces for x-component of the velocity colored 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 proposed framework.

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

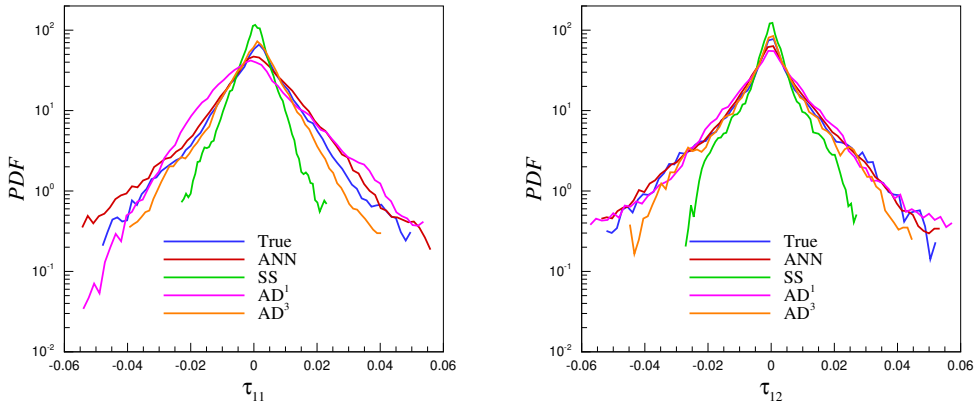
Table 2.4: Mean-squared-error values for deviatoric subfilter scale components with respect to the true subfilter scale stresses for Kolmogorov turbulence.

rather arbitrary, past studies (Maulik and San, 2018) have shown that a choice of the number of iterations between 3 and 5 is usually sufficient for satisfactory subfilter recovery.

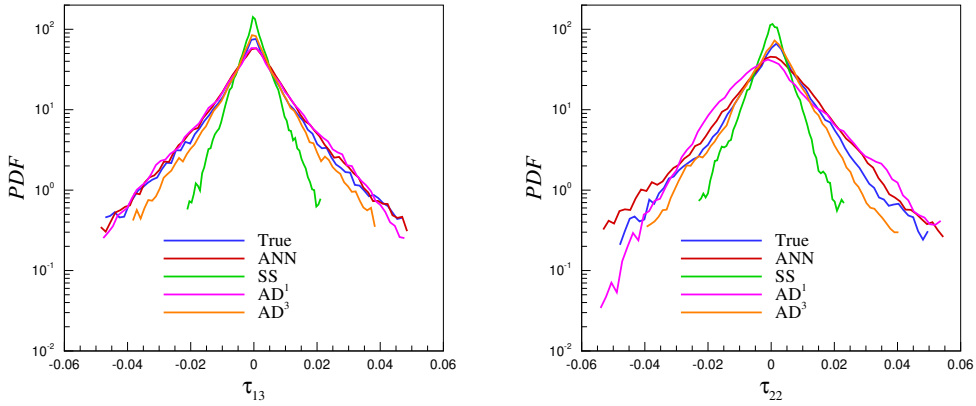
Our performance assessments indicate that the proposed architecture (denoted ANN) performs in a similar fashion to other widely utilized structural subfilter modeling strategies. The AD^3 approach can be seen to perform better (on average) than our proposed framework due to the fact that the specified 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 (both of which require the specification of a low-pass spatial filter). A quantitative characterization of the subfilter stress recovery is shown in Table 2.4 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.

2.6.3 Stratified Turbulence

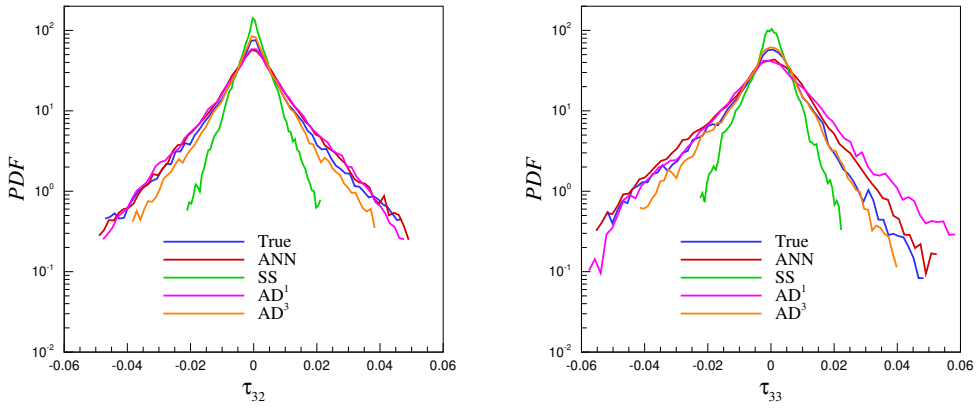
Another three-dimensional test case is given by a stratified turbulence data set obtained using 512^3 degrees of freedom. The difference between this test case and the Kolmogorov turbulence test case (TGV) is that this particular simulation is carried out using an implicit large eddy simulation of the inviscid Euler equations (see



(a) τ_{11} (b) τ_{12}



(c) τ_{13} (d) τ_{22}



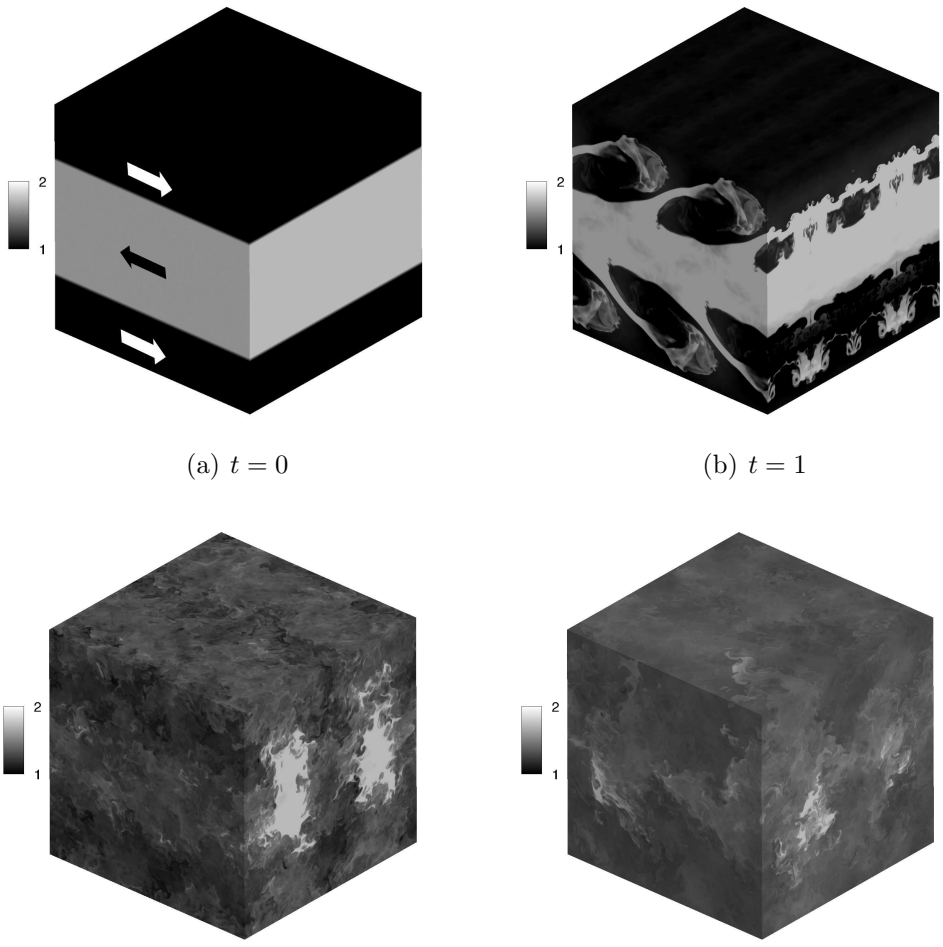
(e) τ_{32} (f) τ_{33}

Figure 2.12: 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.

Maulik and San (2017*b*) 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 (Bos and Bertoglio, 2006; Zhou et al., 2014; Sytine et al., 2000). 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 analog of the study presented in Maulik and San (2017*b*) to generate a Kelvin-Helmholtz instability (KHI) which eventually transitions through non-linear interactions to fully developed compressible turbulence. Figure 2.13 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 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 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 2.14 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



(a) $t = 0$ (b) $t = 1$

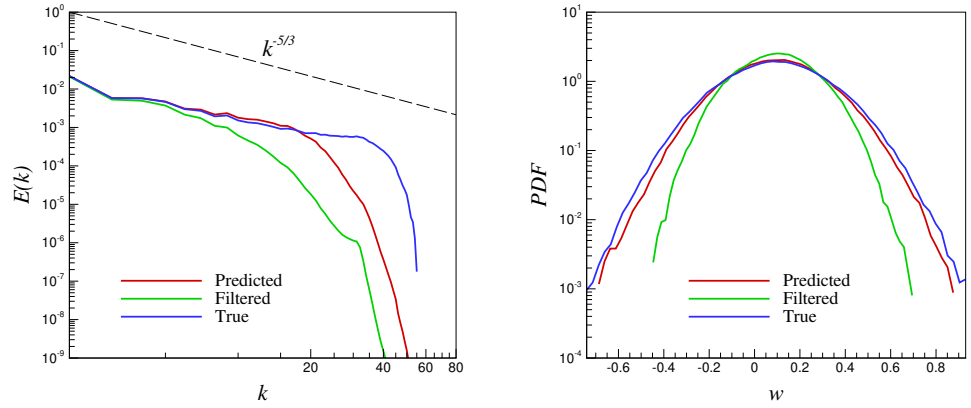
(c) $t = 3$ (d) $t = 5$

Figure 2.13: Evolution of density contours for the stratified turbulence problem through time.

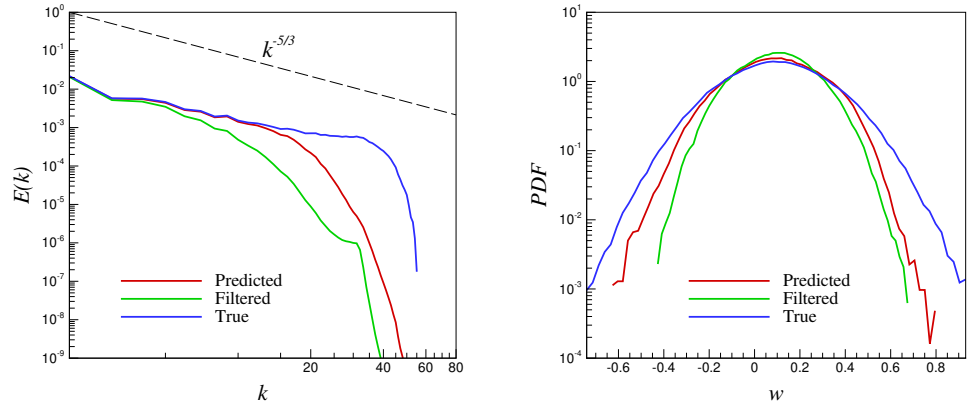
about the mean rather well (which leads to excellent inertial range capture).

The regularization performance of the proposed framework is tested for this particular test case with results described in Figure 2.15. 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 familiar reduction in tail capture accuracy. Aliasing errors are effectively removed as expected. Figures 2.16 and 2.17 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 2.5 where trends similar to those observed for the Kraichnan and Kolmogorov test cases are recovered.

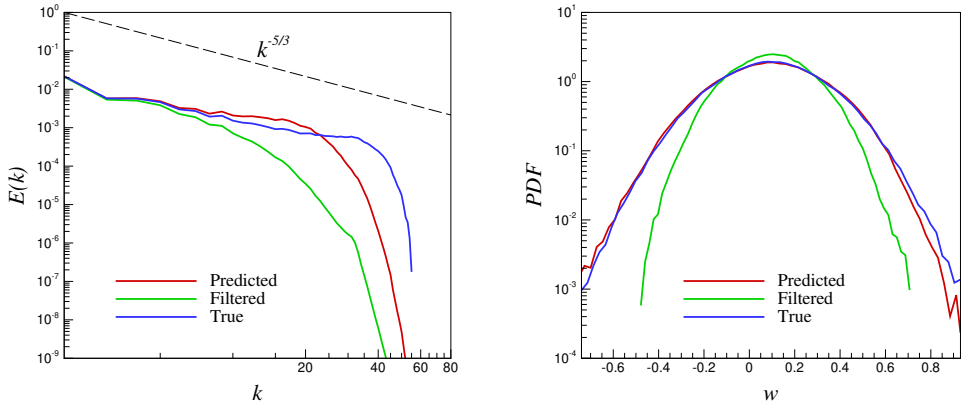
We outline a comparison with other structural closures in Figure 2.18 where trends similar to those observed in the Kolmogorov test case are witnessed. The AD^3 approach performs in a superior manner as expected due to the prior specification of the Gaussian filter for both smoothing and iterative deconvolution. 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 2.6 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.



(a) Test data 1

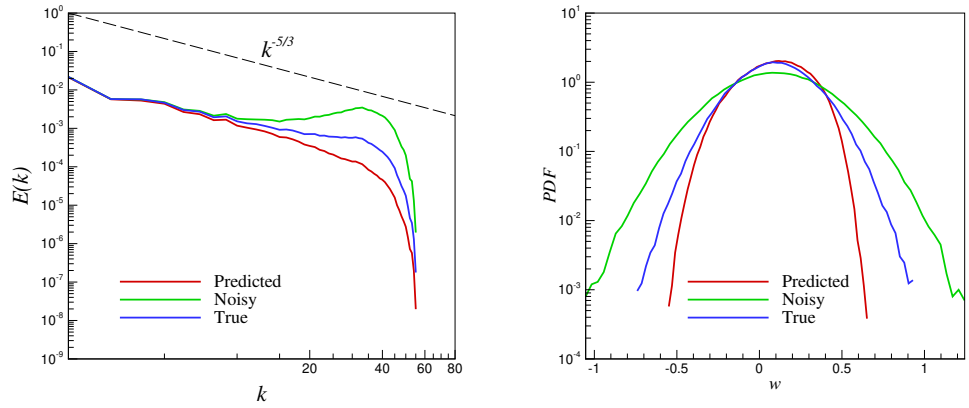


(b) Test data 2

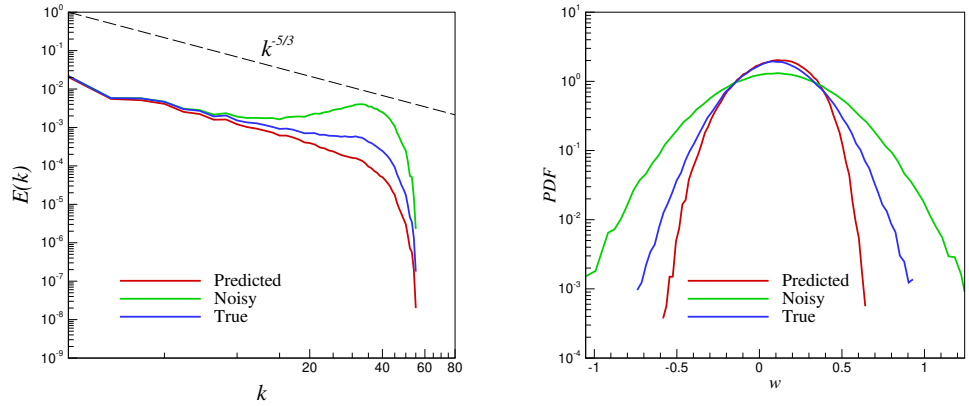


(c) Test data 3

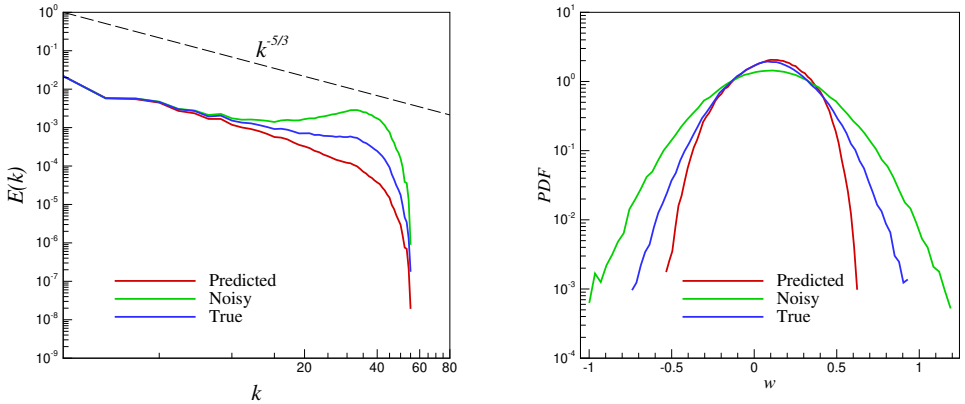
Figure 2.14: A-priori results of the kinetic energy spectra (left) and PDF of the z component of velocity (right) for stratified turbulence. Results for three different deconvolution test data sets shown.



(a) Test data 1

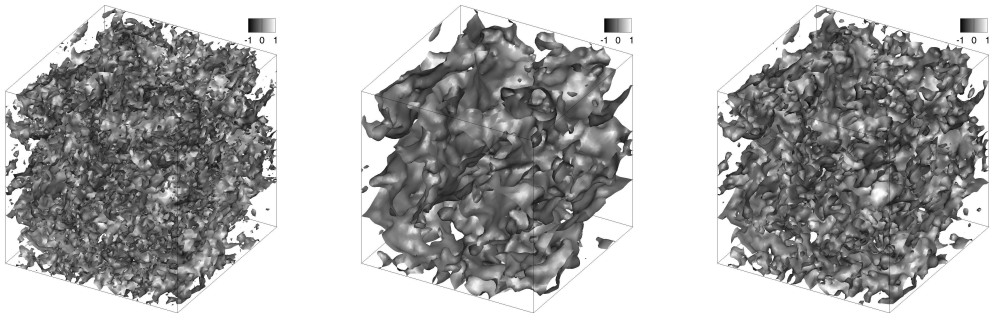


(b) Test data 2



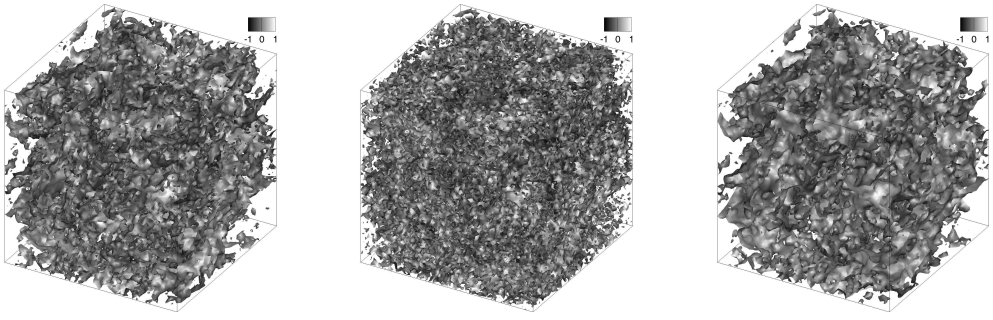
(c) Test data 3

Figure 2.15: A-priori results of the kinetic energy spectra (left) and PDF of the z component of velocity (right) for stratified turbulence. Results for three different regularization test data sets shown.



(a) True (b) Filtered (c) Recovered

Figure 2.16: 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 colored by z component are shown: (a) true coarse-grained fields, (b) coarse-grained fields with Gaussian smoothing, and (c) coarse-grained fields reconstructed using proposed framework.



(a) True (b) Noisy (c) Recovered

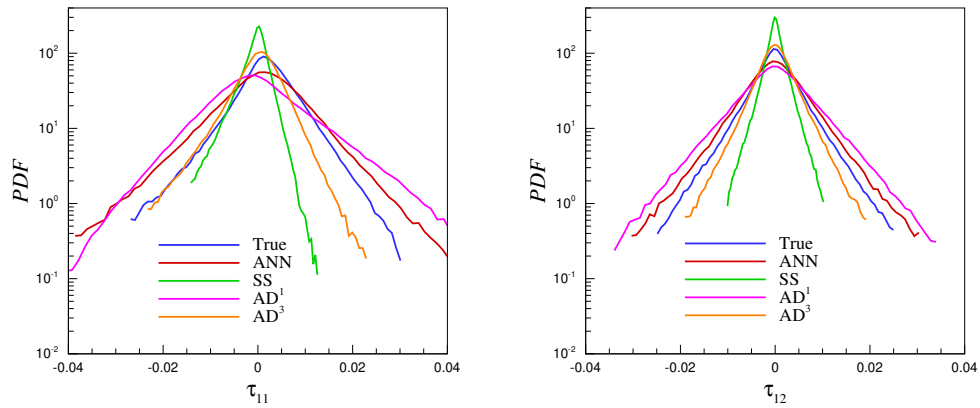
Figure 2.17: A-priori results for velocity field recovery from noisy perturbation inputs for stratified turbulence. Data shown for regularization test data 1. Isosurfaces for x component of the velocity colored by z component are shown: (a) true coarse-grained fields, (b) coarse-grained fields with Gaussian noise, and (c) coarse-grained fields reconstructed using proposed framework.

Stratified turbulence		
	Deconvolution	
Dataset	Noised	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	Noised	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}

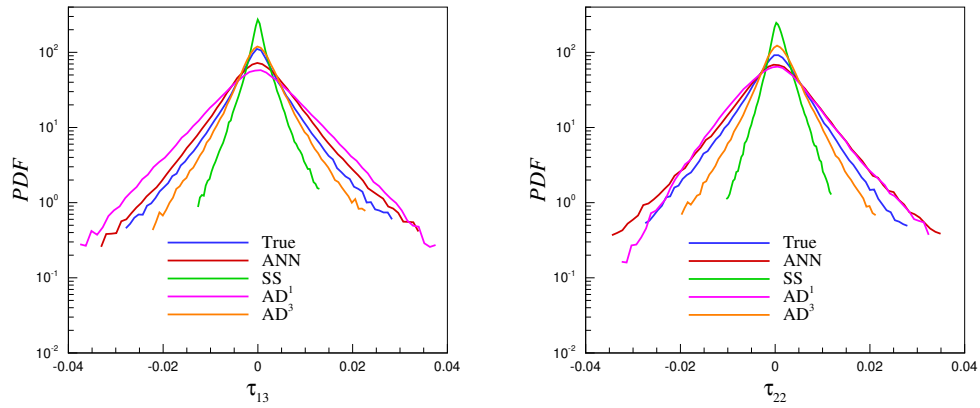
Table 2.5: Mean-squared-error values for deconvolved and regularized fields obtained from the proposed architecture. Data shown from the three-dimensional stratified turbulence test case. Note that the mean-squared-error values are obtained from the z component of the velocity field.

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

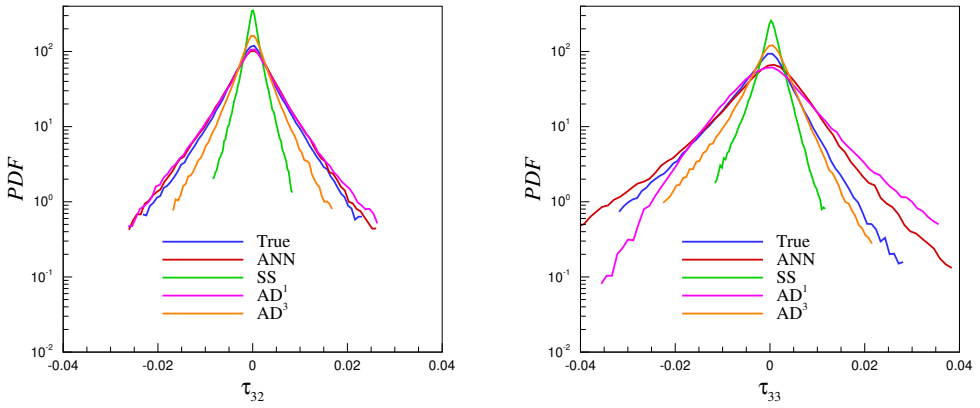
Table 2.6: Mean-squared-error values for deviatoric subfilter scale components with respect to the true subfilter scale stresses for stratified turbulence.



(a) τ_{11} (b) τ_{12}



(c) τ_{13} (d) τ_{22}



(e) τ_{32} (f) τ_{33}

Figure 2.18: 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.

2.7 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 modeling. 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 exhibiting distinctly different physics. For instance, a great utility of any data-driven modeling 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.

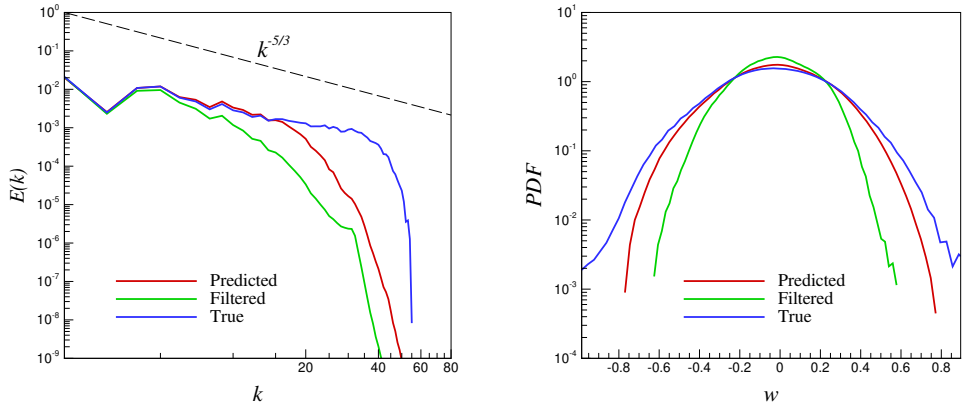
We first examine the ability of the closure modeling strategy by using training data from a high-fidelity Taylor-Green vortex simulation at $Re = 1600$ to recover true fields at $Re = 5000$ as shown in Figure 2.19. We observe that the framework is successfully able to replicate closure performance as shown in previous test cases. It tells us that the deconvolution (or regularization) ability has been learned without violating (as it appears) the underlying physics of homogeneous isotropic turbulence. Our next assessment shown in Figure 2.20 utilizes the closure modeling 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 2.21 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 our

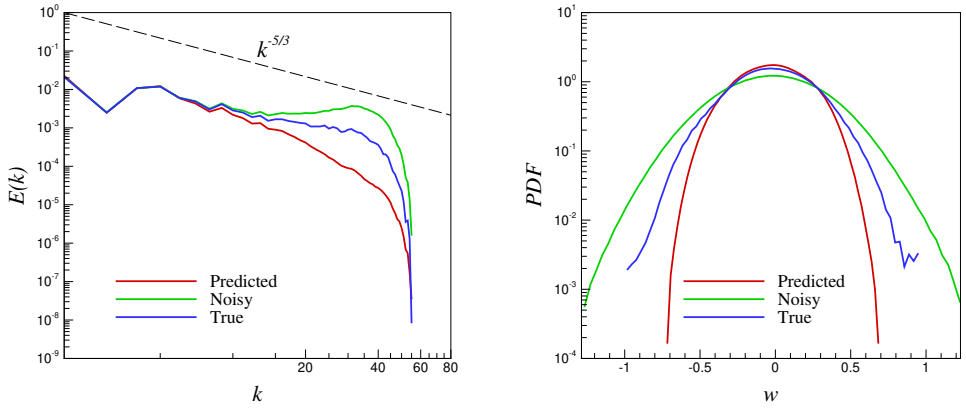
ANN. In this case we highlight that the only link between training and testing data is the $k^{-5/3}$ cascade and the magnitude of perturbations to the coarse-grained field. It can be seen that a remarkable inertial range recovery and regularization performance is still observed for both the low-pass spatially filtered and noised testing data. Analogously, in Figure 2.22, 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.

2.8 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 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

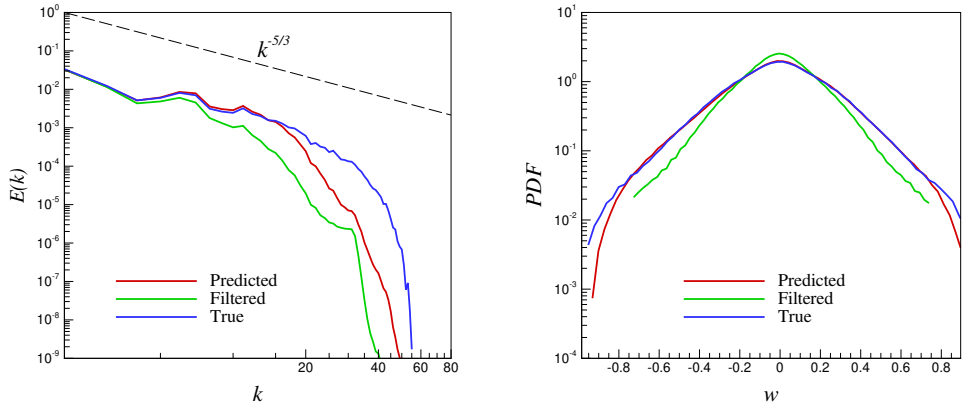


(a) Filtered Inputs

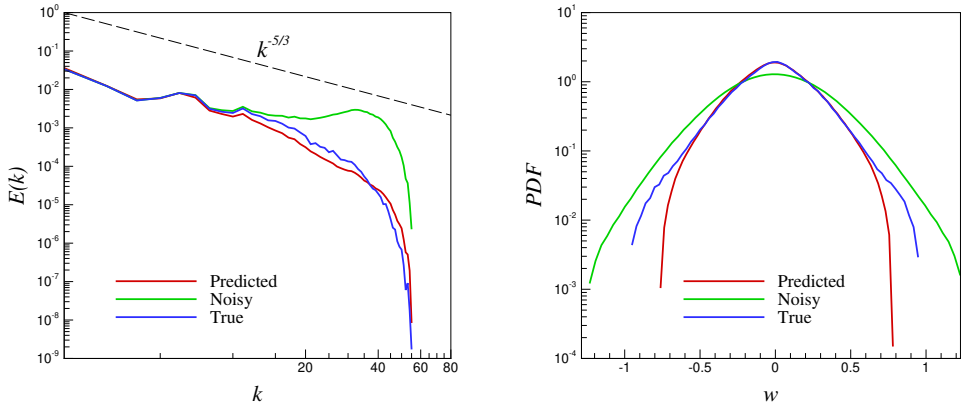


(b) Noised Inputs

Figure 2.19: A-priori results of the kinetic energy spectra (left) and PDF of the vorticity (right) for Kolmogorov turbulence. Here we utilize DNS data for the Taylor-Green vortex at $Re = 1600$ to reconstruct an approximation to the true field for $Re = 5000$.

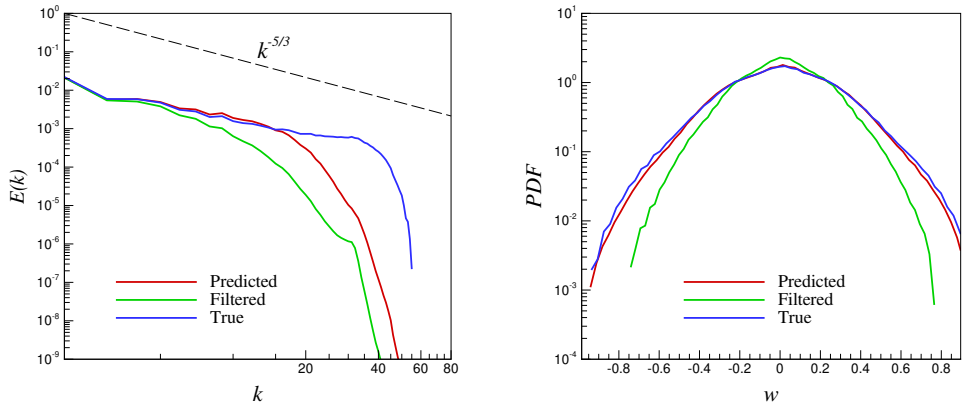


(a) Filtered Inputs

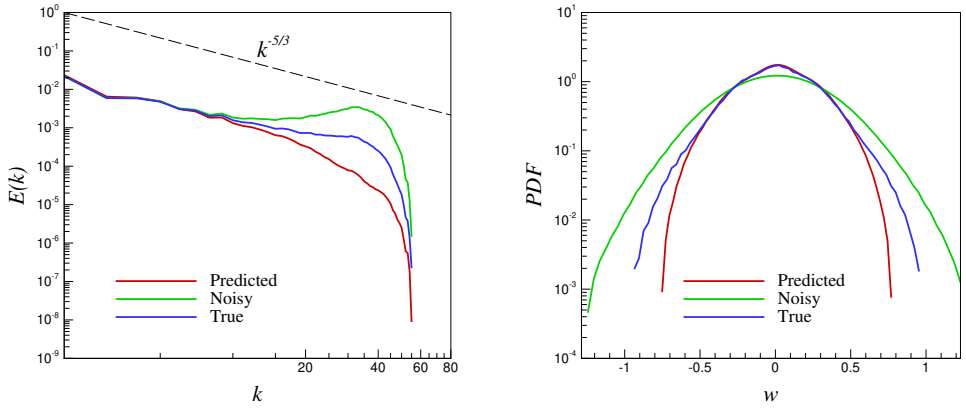


(b) Noised Inputs

Figure 2.20: A-priori results for the kinetic energy spectra (left) and PDF of the vorticity (right) for Kolmogorov turbulence. 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 $Re = 1600$.

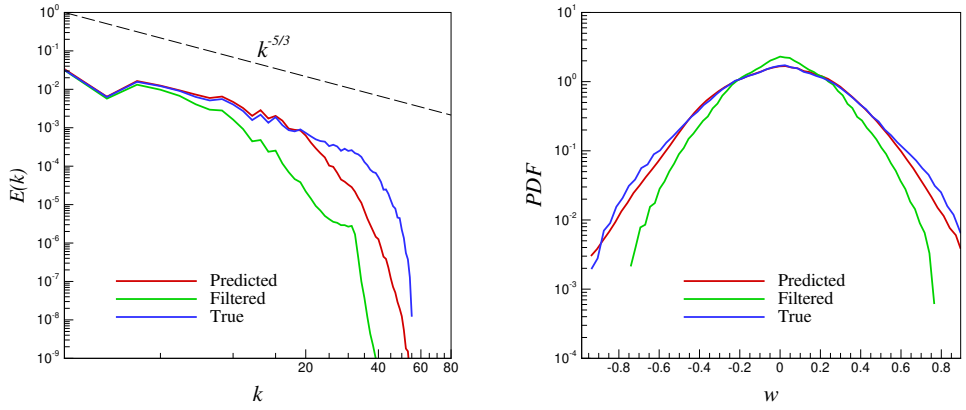


(a) Filtered Inputs

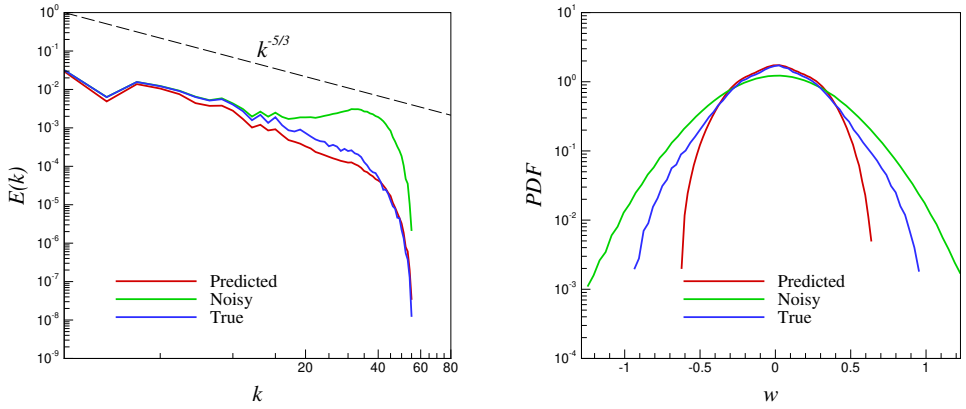


(b) Noised Inputs

Figure 2.21: A-priori results of the kinetic energy spectra (left) and PDF of the vorticity (right) stratified turbulence. 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.



(a) Filtered Inputs



(b) Noised Inputs

Figure 2.22: A-priori results of the kinetic energy spectra (left) and PDF of the vorticity (right) for Kolmogorov turbulence. Here we utilize high fidelity data from the stratified turbulence problem to train a prediction for the Taylor-Green vortex at $Re = 1600$.

of the proposed architecture remains localized to *physics* and not numerical artifact. 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 ability of the proposed approach for subfilter scale content recovery is benchmarked against several popular structural closure modeling 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 Section 2.7, 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 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 modeling framework. Basically, a data-driven model is *only* as good as the data it has been trained on and can only reproduce physical behaviors 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.

CHAPTER III

Data-driven deconvolution for large eddy simulations of Kraichnan turbulence

3.1 Abstract

In this article, we demonstrate the use of artificial neural networks as optimal maps which are utilized for convolution and deconvolution of coarse-grained fields to account for sub-grid scale turbulence effects. We demonstrate that an effective eddy-viscosity is predicted by our purely data-driven large eddy simulation framework without explicit utilization of phenomenological arguments. In addition, our data-driven framework precludes the knowledge of true sub-grid stress information during the training phase due to its focus on estimating an effective filter and its inverse so that grid-resolved variables may be related to direct numerical simulation data statistically. The proposed predictive framework is also combined with a statistical truncation mechanism for ensuring numerical realizability in an explicit formulation. Through this we seek to unite structural and functional modeling strategies for modeling non-linear partial differential equations using reduced degrees of freedom. Both a-priori and a-posteriori results are shown for a two-dimensional decaying turbulence case in addition to a detailed description of validation and testing. A hyperparameter sensitivity study also shows that the proposed dual network framework simplifies learning complexity and is viable with exceedingly simple network architectures. Our

This chapter is adapted from Maulik et al., Phys. Fluids, 30(12), 125109

findings indicate that the proposed framework approximates a robust and stable sub-grid closure which compares favorably to the Smagorinsky and Leith hypotheses for capturing the theoretical k^{-3} scaling in Kraichnan turbulence.

3.2 Introduction

Over the past decade, advances in data collection and increasing access to computational resources have led to a revolution in the use of data-driven techniques for the solution of intractable inverse problems (Mitchell et al., 1990; Shen et al., 2017; Guest et al., 2018; Duraisamy et al., 2019). One such problem is that of turbulence, the multiscale nature of which causes infeasible computational demands even for the most simple systems. This behavior is shared by all non-linear partial differential equations and necessitates the utilization of multiple modeling approximations for tractable compute times. One such modeling approach is that of large eddy simulation (LES) (Sagaut, 2006), which attempts to simulate the evolution of lower wavenumber modes of turbulence while the effects of higher wavenumber modes are modeled by an algebraic or differential equation. The procedure of modeling the finer scales is often denoted a *closure* due to the lack of knowledge about higher-order wavenumber interactions in the coarse-grained flow (Berselli et al., 2006) and remains a critical component of accurate computational modeling for many applications (Hickel et al., 2014; Yu et al., 2016; Zhou et al., 2018). From an LES point of view, the closure problem arises due to the fact that low-pass spatial filtering (due to coarse-graining and discrete numerical approximations) does not commute with the non-linear term.

Within the context of the Navier-Stokes equations, it is generally accepted that the finer scales are dissipative at the Kolmogorov length scales (Kolmogorov, 1941*b*) and therefore, most turbulence models seek to specify a sub-grid viscosity which mimics the dissipative behavior of the unsupported frequencies (Frisch, 1995). Most sub-grid models can be traced back to the seminal work of Smagorinsky (Smagorinsky, 1963),

where a model was proposed based on the concepts of an effective eddy viscosity determined by an a-priori specified mixing length and a $k^{-5/3}$ scaling recovery for the kinetic energy content in the wavenumber domain. Similar hypotheses have also been used for two-dimensional turbulence (Leith, 1968), for approximating the k^{-3} cascade in two-dimensional turbulence and generally have their roots in dimensional analysis related to the cascade of enstrophy. The two aforementioned models may be classified as functional due to the phenomenological nature of their deployment and represent the bulk of LES related turbulence models used in practical deployments.

In contrast, the structural approach to turbulence modeling utilizes no explicit specification of an eddy-viscosity and relies on an estimation of the low-pass spatial filtering nature of coarse-graining. With this approximate knowledge of the filter, arguments for scale-similarity (Bardina et al., 1980; Layton and Lewandowski, 2003) or approximate-deconvolution (AD) (Stolz and Adams, 1999) are utilized to reconstruct the true non-linear term. In case of scale-similarity, the non-linear interactions of flow components are estimated by utilizing a forward filtering operation to the grid-resolved variables, while in AD an inverse filter is estimated using iterative resubstitutions. However, structural techniques are limited due to the fact that they approximately recover sub-filter stresses alone and are not dissipative enough due to the neglect of sub-grid considerations. Therefore, they require the specification of an additional (usually functional) sub-grid model or the specification of a finer resolution where sub-grid terms are negligible (Germano, 2015). Further information about turbulence models and whether they may be classified as functional or structural may be found in Saugaut's excellent text (Sagaut, 2006).

A common thread that connects both functional and structural models is the a-priori specification of a model coefficient or a characteristic filter width or ratio. Consequently, the choice of such parameters become crucial in the a-posteriori performance of the deployed model. Crucially, literature has consistently shown that

the choice of these coefficients are not single-valued, particularly for off-nominal flow situations. One may refer to discussions by Galperin and Orszag (1993) and Canuto and Cheng (1997) for examples for the effect of varying eddy viscosity. The effect of characteristic filter widths and the order of deconvolution has also been explored by San et al. (2015) and by Schneiderbauer and Saeedipour (2018). With this contextual background, in this study, we introduce a hybrid modeling (physics-informed machine learning) methodology for determining sub-grid models without any phenomenological assumptions (in the spirit of structural models) but with sub-grid capture ability. This is accomplished by the use of artificial neural networks (ANNs) to establish data-driven maps between a-priori convolved and deconvolved fields but without the use of any explicit filter.

In recent times, data-driven techniques have become extremely popular for the spatio-temporal modeling of dynamical systems (Schmidt and Lipson, 2009; Bright et al., 2013; Xiao et al., 2015; Brunton et al., 2016; Schaeffer, 2017; Raissi et al., 2017; Mohan and Gaitonde, 2018; Raissi and Karniadakis, 2018; Rudy et al., 2018; San and Maulik, 2018; Wan et al., 2018; Kim et al., 2018; Muravleva et al., 2018; Jin et al., 2018). With respect to turbulence, some widely used strategies for inference include symbolic regression (Weatheritt and Sandberg, 2016, 2017*a*,*b*), where functional model-forms for RANS deployments were generated through optimization against high-fidelity data. Ma et al. (2015) utilized compressive-sensing based machine learning for closure of multiphase system.Gautier et al. (2015) utilized a genetic algorithm was utilized for regression tasks in a close-loop separation control deployment of a turbulent mixing layer. Other techniques incorporating Bayesian ideologies have also been used, for instance by Xiao et al. (2016) where an iterative ensemble Kalman method was used to assimilate prior data for quantifying model form uncertainty in RANS models. In Wang, Wu and Xiao (2017); Wang, Wu, Ling, Iaccarino and Xiao (2017) and Wu et al. (2018*a*), random-forest regressors were utilized for

RANS turbulence-modeling given DNS data. In Singh and Duraisamy (2016) and Singh et al. (2017), an ANN was utilized to predict a non-dimensional correction factor in the Spalart-Allmaras turbulence model through a field-inversion process. The field-inversion process was utilized to develop optimal a-priori estimates for the correction factor from experimental data. Bypassing functional formulations of a turbulence model (a focus of this study) was also studied from the RANS point of view by Tracey et al. (2015). Ling and Templeton (2015) utilized support vector machines, decision trees and random forest regressors for identifying regions of high RANS uncertainty. A deep-learning framework where Reynolds-stresses would be predicted in an invariant subspace was developed by Ling, Kurzawski and Templeton (2016). The reader is directed to a recent review by Duraisamy et al. (2019), for an excellent review of turbulence modeling using data-driven ideas.

As shown above, the use of machine learning ideologies and in particular ANNs has generated significant interest in the turbulence modeling community. This is motivated by the fact that a multilayered artificial neural network may be optimally trained to universally approximate any non-linear function (Hornik et al., 1989). Greater accessibility to data and the GPU revolution has also motivated the development of advanced ANN architectures for constrained learning and improved physical interpretability. Within the context of LES (and associated with the scope of this paper) there are several investigations into sub-grid modeling using data-driven techniques. In one of the first studies of the feasibility of mapping to unresolved stresses using grid resolved variables by learning from high-fidelity data, Sarghini et al. (2003) utilized ANNs for estimating Smagorinsky model-form coefficients within a mixed sub-grid model for a turbulent channel flow. This may be considered similar to the field-inversion procedure describe previously. ANNs were also used for wall-modeling by Milano and Koumoutsakos (2002) where it was used to reconstruct the near wall field and compared to standard proper-orthogonal-decomposition techniques. An al-

ternative to ANNs for sub-grid predictions was proposed by King et al. (2016) where a-priori optimization was utilized to minimize the L^2-error between true and modeled sub-grid quantities in a least-squares sense using a parameter-free Volterra series. Maulik and San (2017*a*) utilized an extreme-learning-machine (a variant of a single-layered ANN) to obtain maps between low-pass spatially filtered and deconvolved variables in an a-priori sense. This had implications for the use of ANNs for turbulence modeling without model-form specification. A more in-depth investigation has recently been undertaken by Fukami et al. (2018) where convolutional ANNs are utilized for reconstructing downsampled snapshots of turbulence. Gamahara and Hattori (2017) utilized ANNs for identifying correlations with grid-resolved quantities for an indirect method of model-form identification in turbulent channel flow. The study by Vollant et al. (2017) utilized ANNs in conjuction with optimal estimator theory to obtain functional forms for sub-grid stresses. In Beck et al. (2018), a variety of neural network architectures such as convolutional and recurrent neural networks are studied for predicting closure terms for decaying homogeneous isotropic turbulence. A least-squares based truncation is specified for stable deployments of their model-free closures. Model-free turbulence closures are also specified by Maulik et al. (2019), where sub-grid scale stresses are learned directly from DNS data and deployed in a-posteriori through a truncation for numerical stability. King et al. (2018) studied generative-adversarial networks and the LAT-NET (Hennigh, 2017) for a-priori recovery of statistics such as the intermittency of turbulent fluctuations and spectral scaling. A detailed discussion of the potential benefits and challenges of deep learning for turbulence (and fluid dynamics in general) may be found in the article by Kutz (Kutz, 2017).

While a large majority of the LES-based frameworks presented above utilize a least-squares error minimization technique for constructing maps to sub-grid stresses *directly*, this work represents a physics-informed implementation of sub-grid source

terms through the learning of convolutional and deconvolution maps between grid-resolved and unresolved fields. In other words, our framework is able to reproduce, approximately, a map related to the convolution associated with insufficient grid-support in LES implementations of the Navier-Stokes equations as well as its inverse. These optimal maps are obtained by supervised learning from subsampled direct numerical simulation (DNS) data and are deployed in an a-posteriori fashion for the LES of two-dimensional turbulence. In this manner, we unite the advantages of functional and structural modeling of turbulence in addition to precluding the use of any phenomenological arguments. Through this, we also aim to achieve a harmonious combination of first-principles based physics as well data-driven mechanisms for high accuracy. 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 and Moser, 1999; Moser et al., 2009; LaBryer et al., 2015; King et al., 2016; Pathak et al., 2018).

Therefore, this investigation represents an advancement of the concepts proposed by the authors previously (Maulik and San, 2017*a*), where solely the deconvolutional ability of artificial neural networks was investigated in an a-priori sense for sub-filter stresses. The adaptations proposed in our current study are targeted towards recovering the sub-grid component of the coarse-grained LES computation. In addition, we not only address the issue of a-priori sub-grid recovery with our proposed closure, but also demonstrate its robustness in a-posteriori deployment with associated numerical challenges. While the two-dimensional turbulence case is utilized for a proof-of-concept as well as for its geophysical implications where improved closure development is still sought extensively, our generalized framework may easily be scaled up to multiple dimensional non-linear partial differential equations. Our results indicate that the proposed framework provides for a robust sub-grid model with a dynamically computed effective eddy-viscosity within the structural modeling

ideology.

3.3 Turbulence modeling equations

We proceed with the introduction of our framework by outlining the governing equations for two-dimensional turbulence. These are given by the Navier-Stokes equations in the vorticity-streamfunction formulation. In place of a primitive variable formulation, our decaying turbulence problem is solved for using the temporal evolution of the following non-dimensionalized and coupled system of equations,

$$\begin{aligned} \frac{\partial \omega}{\partial t} + J(\omega, \psi) &= \frac{1}{Re} \nabla^2 \omega, \\ \nabla^2 \psi &= -\omega, \end{aligned} \tag{3.1}$$

where the velocity vector components may be recovered as

$$\begin{aligned} \frac{\partial \psi}{\partial y} &= u \\ \frac{\partial \psi}{\partial x} &= -v. \end{aligned} \tag{3.2}$$

The computational necessities of coarse-graining result in a grid-filtered system of equations

$$\begin{aligned} \frac{\partial \overline{\omega}}{\partial t} + J(\overline{\omega}, \overline{\psi}) &= \frac{1}{Re} \nabla^2 \overline{\omega} + \Pi, \\ \nabla^2 \overline{\psi} &= -\overline{\omega}, \end{aligned} \tag{3.3}$$

where overbarred quantities imply grid-resolved variables. A resulting unclosed term is obtained, ideally represented as

$$\Pi = J(\overline{\omega}, \overline{\psi}) - \overline{J(\omega, \psi)}. \tag{3.4}$$

The second term on the right-hand side of the above equation represents the primary target of approximation for the structural modeling mechanism. In contrast, the functional modeling procedure is to represent Π as an effective eddy-viscosity multiplied by Laplacian of the vorticity. In this study, we shall utilize a data-driven paradigm for approximating

$$\overline{J(\omega,\psi)} \approx \widetilde{J(\omega^*,\psi^*)}, \tag{3.5}$$

where asterisked quantities are those obtained by data-driven deconvolution and the tilde represents data-driven convolution. This procedure is similar to the AD mechanism which requires an a-priori low-pass spatial filter specification. Note that the proposed methodology effectively aims to approximate the operations of Fourier cut-off filtering and its inverse which is the primary reason why it blends the distinction between sub-filter and sub-grid recovery. The former is a potential limitation of the AD mechanism in its current implementation. Our approximate sub-grid model is thus given by

$$\tilde{\Pi} = J(\bar{\omega},\bar{\psi}) - \widetilde{J(\omega^*,\psi^*)}. \tag{3.6}$$

For the purpose of comparison we also introduce the Smagorinsky and Leith models which utilize algebraic eddy-viscosities for sub-grid stress calculation given by

$$\Pi_e = \nabla.\left(\nu_e \nabla\bar{\omega}\right), \tag{3.7}$$

where for the Smagorinsky model we have

$$\nu_e = (C_s\delta)^2|\bar{S}|, \tag{3.8}$$

and the Leith hypothesis states

$$\nu_e = (C_l\delta)^3|\nabla\bar{\omega}|. \tag{3.9}$$

Note that $|\bar{S}| = \sqrt{2S_{ij}S_{ij}}$ and $|\nabla\bar{\omega}|$ correspond to two commonly used kernels for eddy-viscosity approximations. Here, δ is generally assumed to be the characteristic mixing length taken to be the grid size. The online performance of our proposed framework shall be compared to these simple, but robust closures. We remark here that the standard procedure for closure in the vorticity-streamfunction formulation (relevant to two-dimensional simulations) is based on sub-grid vorticity source term modeling but our generalized procedure may be extended to the primitive variable approach as a source term in the Navier-Stokes momentum equations. For the convenience of the reader we also tabulate some of the notation that will be widely used in the rest of this article in Table 3.3. We note that the variables outlined in this table are all defined on a coarse(i.e, LES) grid. Details regarding the preparation of the data for our machine learning methods shall be outlined in subsequent sections.

Notation	Category
$\bar{a}$	Grid filtered (i.e, Fourier cut-off filtered) from DNS
a^c	Comb filtered (i.e, sub-sampled) from DNS
a^*	Data-driven deconvolved variable
$\tilde{a}$	Data-driven convolved variable

Table 3.1: A summary of filter and deconvolutional notation

3.4 Data-driven convolution and deconvolution

The ANN, also known as a multilayered perceptron, consists of a set of linear or non-linear mathematical operations on an input space vector to establish a map to an output space. Other than the input and output spaces, a ANN may also 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 multi-step 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 non-linearity into the eventual map. In contrast, no activation (also referred to as a 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 labeled data used for informing the optimal map between inputs and outputs. Mathematically, if our input vector $\mathbf{p}$ resides in a P-dimensional space and our desired output $\mathbf{q}$ resides in a Q-dimensional space, the ANN establishes a map $\mathbb{M}$ as follows:

$$\mathbb{M} : \{p_1, p_2, \ldots, p_P\} \in \mathbb{R}^P \rightarrow \{q_1, q_2, \ldots, q_Q\} \in \mathbb{R}^Q. \tag{3.10}$$

In this study, we utilize two maps which relate to convolution and deconvolution of fields with grid-resolved and sub-grid components respectively. We must caution the reader here that the maps are not assumed to transform between isomorphic spaces (considered a limitation of structural AD (Guermond et al., 2004; Germano, 2015)). This allows for the estimation of sub-grid loss due to coarse-graining the degrees of freedom in an LES deployment. In equation form, our optimal map $\mathbb{M}_1$

relates coarse-grained field stencils to their grid-filtered (i.e., Fourier cut-off filtered) counterparts and is given by

$$\mathbb{M}_1 : \{\omega^c_{i,j}, \omega^c_{i,j+1}, \omega^c_{i,j-1} \ldots, \omega^c_{i-1,j-1} \in \mathbb{R}^9 \rightarrow \{\tilde{\omega}\} \in \mathbb{R}^1. \tag{3.11}$$

where $\tilde{\omega}$ represents an approximation for $\bar{\omega}$.

Our second map, relates grid-filtered field stencils to their coarse-grained counterparts given by

$$\mathbb{M}_2 : \{\bar{\omega}_{i,j}, \bar{\omega}_{i,j+1}, \bar{\omega}_{i,j-1} \ldots, \bar{\omega}_{i-1,j-1} \in \mathbb{R}^9 \rightarrow \{\omega^*\} \in \mathbb{R}^1. \tag{3.12}$$

where ω^* represents an approximation for ω^c. Note that both maps are trained for optimal prediction using normalized inputs. Our normalization (approximately) rescales our data to zero mean and unit variance by using grid-resolved variable quantities. Therefore, both inputs and outputs to maps are normalized by quantities available dynamically and the deployment of the network does not require a-priori storage of training parameters. For instance, the normalization of $\bar{\omega}$ may be obtained by

$$\bar{\omega}^n = \frac{\bar{\omega} - \mu(\bar{\omega})}{\sigma(\bar{\omega})}, \tag{3.13}$$

where $\mu(a)$ and $\sigma(a)$ refer to the mean and variance of a field variable a. Similarly the normalization of ω^* is given by

$$\omega^{*^n} = \frac{\omega^* - \mu(\bar{\omega})}{\sigma(\bar{\omega})}. \tag{3.14}$$

In this manner, no a-priori training coefficients may be recorded. Trained maps using this normalization technique may thus be used for the convolution or deconvolution of any coarse-grained variable. A key facet of our proposed methodology is that our

trained maps are obtained only from vorticity data even though they need deployment for the deconvolution of the streamfunction as well as the convolution of the Jacobian. Successful sub-grid information recovery (described in the results section) shows that this data-independence in training can be related to a true learning of the filtering and deconvolution characteristics between coarse and fine grids.

The pseudocode for a deployment of our optimally trained maps is shown in Algorithm 3 where it can be seen that each time step (or sub-step) of an explicit flow evolution requires the specification of a data-driven approximation to the true Jacobian $J(\overline{\omega, \psi})$. In subsequent sections, we shall comment on the final a-posteriori constraining for ensuring numerical realizability. Figure 3.1 visually outlines the two networks deployed in this study.

Algorithm 3 Proposed framework deployment

1: Given trained maps $\mathbb{M}_1$ and $\mathbb{M}_2$
2: Given $\overline{\omega}$ and $\overline{\psi}$
3: Normalize $\overline{\omega}$ and $\overline{\psi}$ to get $\overline{\omega}^n$ and $\overline{\psi}^n$ respectively
4: Use $\mathbb{M}_2$ to obtain deconvolved variables ω^{n^*} and ψ^{n^*}
5: Rescale to physical domain to get ω^* and ψ^*
6: Calculate estimated coarse-grid Jacobian $J(\omega^*, \psi^*)$
7: Normalize Jacobian $J(\omega^*, \psi^*)$ to get $J(\omega^*, \psi^*)^n$
8: Use $\mathbb{M}_1$ to obtain convolved variables $\widetilde{J(\omega^*, \psi^*)^n}$
9: Rescale $\widetilde{J(\omega^*, \psi^*)^n}$ to physical domain to get $\widetilde{J(\omega^*, \psi^*)}$
10: Deploy turbulence model $\tilde{\Pi} = J(\bar{\omega}, \bar{\psi}) - \widetilde{J(\omega^*, \psi^*)}$ subject to post-processing for numerical stability given by Equation 3.15

As evident, implementation of the proposed framework requires multiple convolutional and deconvolutional passes over the grid-resolved variables and therefore we refer to this framework, from henceforth, as the data-driven convolutional and deconvolutional closure (DCD). Both our networks utilize one hidden layer along with the

Figure 3.1: A schematic of our data-driven mapping for convolution and deconvolution. Two separate ANNs are utilized for projection to and from deconvolved variable space.

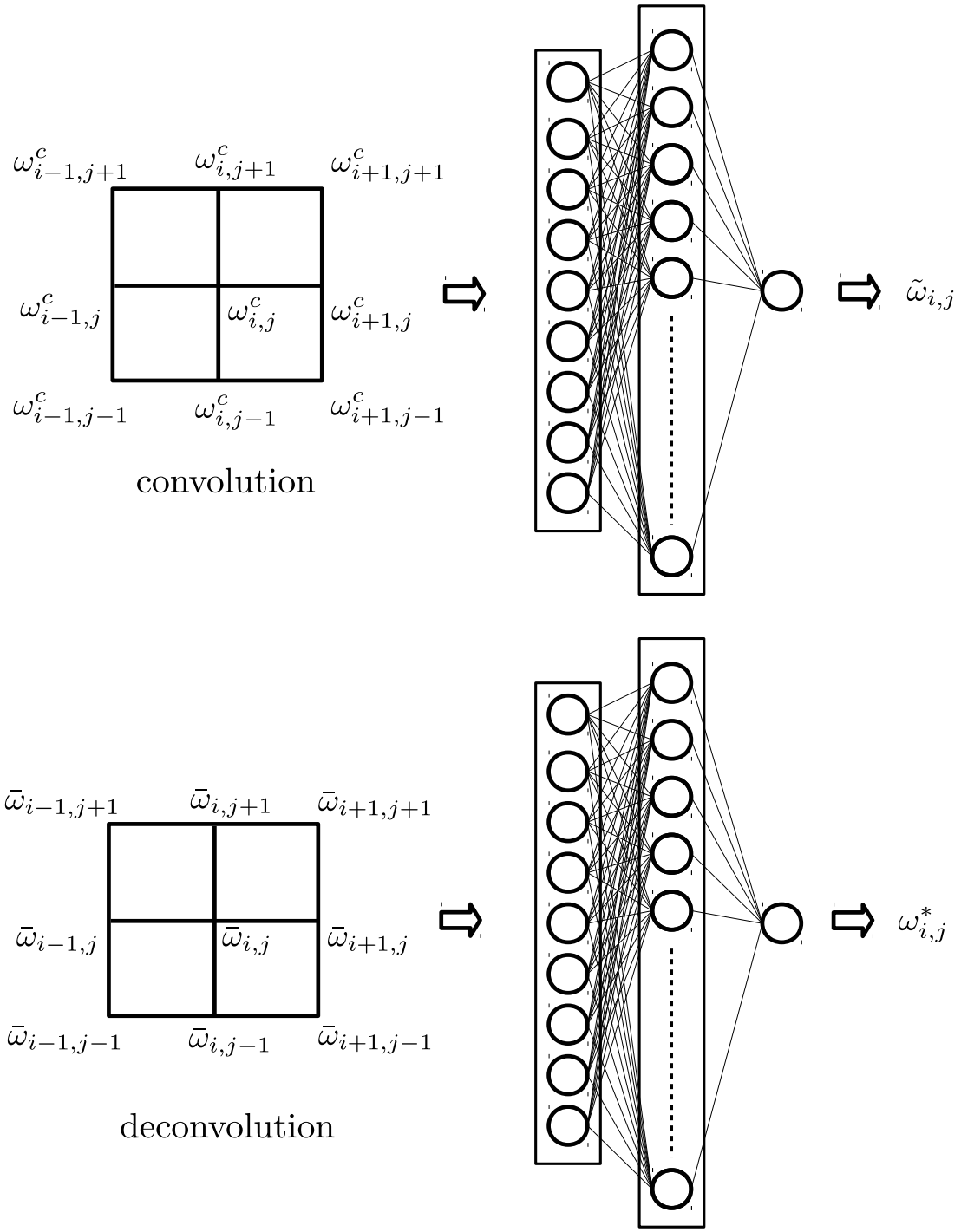
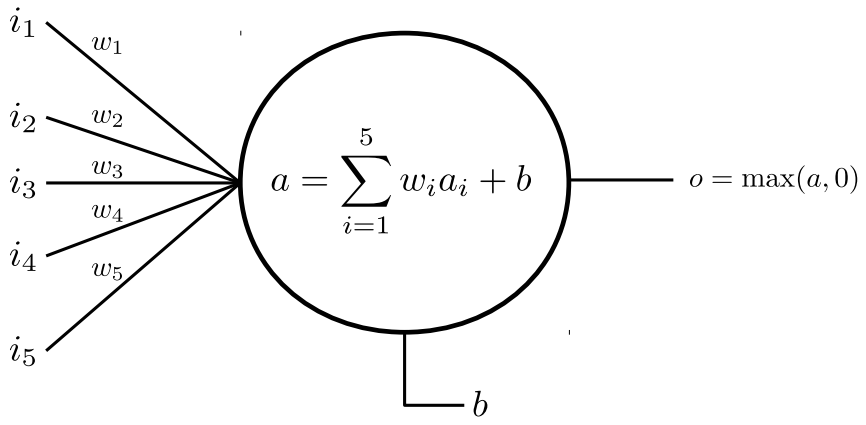


Figure 3.2: A schematic of our biasing and activation at each hidden layer neuron. Assuming five inputs from previous layer.



input and output layers. This hidden and output layers have a bias vector associated with it. For faster training, we utilize rectified linear activation functions (ReLU) for our hidden layer and a linear activation function for the output layer. Note that input data is not activated as it enters the network. Our hidden layer utilizes 100 unit cells (i.e., neurons) which are acted on by the ReLU transformation and biasing. The process of bias and activation at each neuron is displayed in Figure 3.2 and every neuron is fully connected to its previous layer (i.e, with incident inputs from all neurons from the previous layer). In subsequent sections, we outline a sensitivity study of our proposed ideology for varying architecture depths where it is proven that one-layered networks suffice for this particular problem.

3.5 Training and a-priori validation

For the purpose of generating our optimal maps discussed in the previous section, we utilize two supervised learnings with sets of labeled inputs and outputs obtained from direct numerical simulation (DNS) data for two-dimensional Kraichnan turbulence. We have utilized a second-order accurate energy-conserving Arakawa scheme for the non-linear Jacobian and second-order accurate finite-difference discretization

schemes for the Laplacian of the vorticity. The Poisson update is performed using a spectrally-accurate solver and the time-integration is performed by a third-order accurate TVD Runge-Kutta explicit method. Further details on the problem setup and the implementation of an energy and enstrophy conserving numerical method can be found by the authors' previous studies (San and Staples, 2012; Maulik and San, 2017*c*). Our grid-resolved variables (i.e., $\bar{\omega}$) are generated by a Fourier cut-off 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). Our subsampled variables (i.e., ω^c) are obtained by a comb filtering procedure where every eighth data point is retained.

We also emphasize on the fact that, while the DNS data generated multiple time snapshots of flow evolution, data was 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. Our DNS utilized an explicit formulation with a constant timestep of 0.0001 implying potential generation of 40000 snapshots out of which only 4 were selected at regular intervals for data harvesting. This represents a 0.01% utilization of total potential data during training which is particularly challenging for this unsteady problem. The generation of data sets at the coarse grained level is outlined in Algorithm 4.

We also note that the Reynolds number chosen for generating the training and validation data sets is given by $Re = 32000$ while deployment is tested for a higher Reynolds number of 64000 for both a-priori and a-posteriori assessment. We remind the reader here, map training is performed solely on the vorticity field despite the fact that trained maps are to be utilized for vorticity, streamfunction and the Jacobian. The generation of data sets at the coarse grained level is outlined in algorithm 4.

Algorithm 4 Data harvesting from DNS

1: Obtain DNS data for vorticity ω^{DNS} at $N^2 = 2048^2$

2: Comb filter to obtain ω^c from ω^{DNS} by sub-sampling every eighth point

3: Grid filter to obtain $\bar{\omega}$ from ω^{DNS}

4: Normalize $\bar{\omega}$ to $\bar{\omega}^n$ using Equations 3.13

5: Normalize ω^c to ω^{c^n} using Equation 3.14

6: ω^{c^n} and $\bar{\omega}^n$ are input and output pairs respectively for map $\mathbb{M}_1$ optimization, where we assume true output $\tilde{\omega}^n \approx \bar{\omega}^n$ according to Equation 3.5

7: $\bar{\omega}^n$ and ω^{c^n} are input and output pairs respectively for map $\mathbb{M}_2$ optimization, where we assume true output $\omega^{*^n} \approx \omega^{c^n}$

Two-thirds of the total dataset generated for optimization is utilized for training and the rest is utilized for test 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. The test data, however, is utilized to record the performance of the trained network on data it was not exposed to during training. Similar behavior in training and test losses would imply a well-formulated learning problem. The final ANN (obtained post-training) would be selected according to the best loss on the test data after a desired number of iterations which for this study was fixed at 50. The choice for a low number of iterations was observed by Pearson correlation values reaching 0.99 for both training and test data sets. We also note that the error-minimization in the training of the ANN utilized the Adam optimizer (Kingma and Ba, 2014) implemented in the open-source neural network training platform TensorFlow. We remark that while the networks may have learned the target maps from the data they are provided for training and test, validation would require an a-posteriori examination as detailed in the following section. We note here that data preprocessing as well as architectural modifications (for instance network depth, number of neurons and activation types) need further investigation

for improved generalization.

We first outline an a-priori study for the proposed framework where the optimal maps are utilized for predicting probability distributions for the true Jacobian i.e., $J(\overline{\omega, \psi})$. A pseudocode for the computation of this true Jacobian is outlined in Algorithm 5. In other words, we assess the turbulence model for a one snapshot prediction. This study is carried out for one of our data snapshots $t = 2$ but for both in and out-of-training data sets. We remark that the maps have previously been exposed to vorticity data from $Re = 32000$ only and our out-of-training data set is given by a similar flow scenario but at higher Reynolds number given by $Re = 64000$. One can thus make the argument for some transfer of learning between similar flow classes but with slight difference in physics. The performance of the framework is shown in Figure 3.3 where the framework predicts the density functions of the true Jacobian accurately for both sets of data. We also note that this study solely utilized a mean-squared-error minimization for the target variables without any physics-based regularization. A future study involving loss-functions devised with intuition from the Navier-Stokes equations would potentially aid in preserving invariance and symmetry properties between grid-resolved and deconvolved space. In addition, while the localized stencil based sampling for map deployments proposed here is amenable to deployment in structured grids, extension to arbitrary meshes would require the use of interpolation or graph convolutional kernels for unstructured information injection into the learning architecture.

Algorithm 5 True Jacobian $\overline{J(\omega, \psi)}$ from DNS

1: Obtain DNS data for vorticity ω^{DNS} and streamfunction ψ^{DNS} at $N^2 = 2048^2$

2: Calculate Jacobian on DNS grid i.e., $J(\omega^{DNS}, \psi^{DNS})$

3: Apply grid filter to $J(\omega^{DNS}, \psi^{DNS})$ in order to obtain $\overline{J(\omega, \psi)}$ at $N^2 = 256^2$.

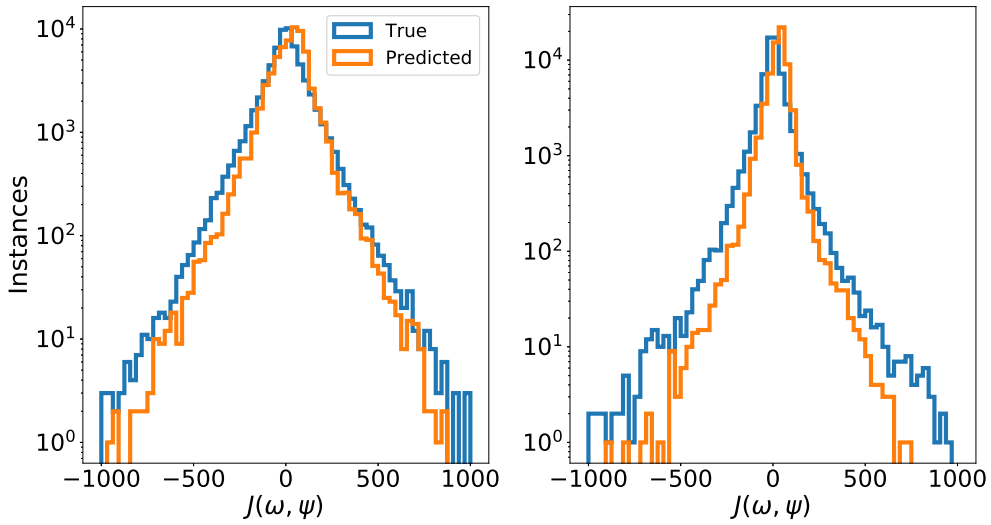


Figure 3.3: The prediction ability of the use of both forward and inverse maps in the calculation of the approximate underlying Jacobian $\widetilde{J(\omega^*, \psi^*)}$ for $Re = 32000$ (left) and $Re = 64000$ (right). The true Jacobian $\overline{J(\omega, \psi)}$ is also shown.

3.6 *A posteriori* testing

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 and 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 setup is particulary challenging due to the neglected effects of numerics in the a-priori training and testing. 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. Theoretical comparisons with Kraichnan turbulence (Kraichnan, 1967) and the expected k^{-3} cascade are also provided. In brief, we mention that the numerical implementation of the conservation laws are through second-order discretizations for all spatial quantities (with a kinetic-energy conserving Arakawa discretization for

the calculation of the non-linear 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 estimating $\tilde{\Pi}$ 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. For the purpose of numerical stability we ensure the following condition before deploying our framework

$$\Pi = \begin{cases} \tilde{\Pi}, & \text{if } (\nabla^2\bar{\omega})(\tilde{\Pi}) > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{3.15}$$

where the truncation explicitly ensures no negative numerical viscosities due to the deployment of the sub-grid model. We remind the reader that the Smagorinsky and Leith hypotheses explicitly specify positive eddy-viscosities that are obtained by absolute value quantities as given in Equations 3.8 and 3.9. An a-priori visual quantification of the truncation is shown in Figure 3.4 where quantities in the first and third quadrants are retained predictions and the others are discarded. A similar behavior is seen for both $Re = 32000$ and $Re = 64000$ data. This image also highlights the challenges of translating a-priori conclusions to a-posteriori implementations due to the requirement of numerical stability.

Figure 3.5 displays the statistical fidelity of coarse-grained simulations obtained with the deployment of the proposed framework for $Re = 32000$. 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 cut-off 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-

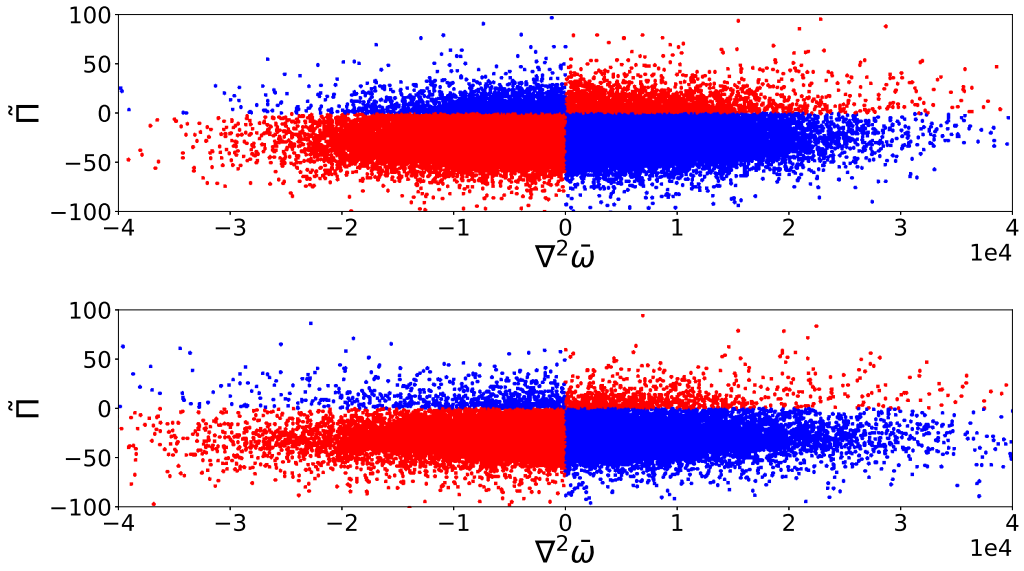


Figure 3.4: A visual assessment of the truncation of our numerical post-processing during deployment given by Equation 3.15. Blue points indicate truncated deployment for ensuring no negative viscosity and numerical stability. A-priori predictions for $Re = 32000$ (top) and $Re = 64000$ (bottom) shown.

grained flow by estimating the effect of sub-grid quantities and preserving trends with regards to the inertial range scaling. Figure 3.6 visually quantifies the effect of the stabilization imparted by the proposed framework. The reader may observe that the proposed framework recovers an excellent scaling behavior. This is similar to the performance obtained by deploying the Smagorinsky model at $C_s = 0.2$, a widely utilized parameteric choice obtained through prior numerical experimentation. The Leith performance at $C_l = 0.2$ is slightly under-dissipative. The reader may notice that an arbitrary choice of $C_s = C_l = 1.0$ leads to overdissipative performance of the eddy-viscosity closures. Our data-driven framework is thus more resistant to unnecessary dissipation. Note that choice of a higher eddy viscosity coefficient for two-dimensional turbulence has been detailed in previous literature (Cushman-Roisin and Beckers, 2011). Another quantification of the perfomance of the DCD closure is described in Figures 3.7 and 3.8 which juxtapose the varying performance of these

parameter-dependant eddy-viscosity hypothesis (i.e., Smagorinsky and Leith respectively) to the proposed data-driven approach. One can observe that an optimal selection of parameters (after a-posteriori examination) given by $C_l = 0.5$ for the Leith model recreates the performance of the proposed framework well as well. This implies that the proposed framework has learned a similar dissipative nature through a-priori optimization of a filter and its inverse. 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 and Orszag, 1993; Canuto and Cheng, 1997; Vorobev and Zikanov, 2008) with higher values specifically for geophysical flows. In comparison, the proposed framework has embedded the adaptive nature of dissipation into its map which is a promising outcome. Before proceeding, we note that default parameteric choices for the Smagorinsky and Leith models are given by $C_s = C_l = 0.2$.

For ensuring that the training is sufficiently generalized for this particular problem, we establish a suite of testing for the predictive performance and the numerical stability of our proposed framework. We first perform multiple forward simulations using the deployment of our proposed closure by utilizing a different random seed in the random-number generation required for the initial conditions at $Re = 32000$ (Maulik and San, 2017*c*). This is to ensure that there is no data memorization by our maps. We choose 24 random initial conditions and ensemble-average their kinetic energy spectra at the completion of the LES for our model as well as the Smagorinsky, Leith and no-model (i.e., UNS) coarse-grid runs. We have also included ensemble results from Smagorinsky and Leith deployments at higher values of $C_s = C_l = 1.0$ to describe the loss of fidelity at the lower wavenumbers in case of incorrect parameter specification. The resultant spectra are shown in Figure 3.9 where one can ascertain that the prediction quality of our framework remains identical regardless of varying initial conditions. This is promising as it validates our hypothesis that it is the

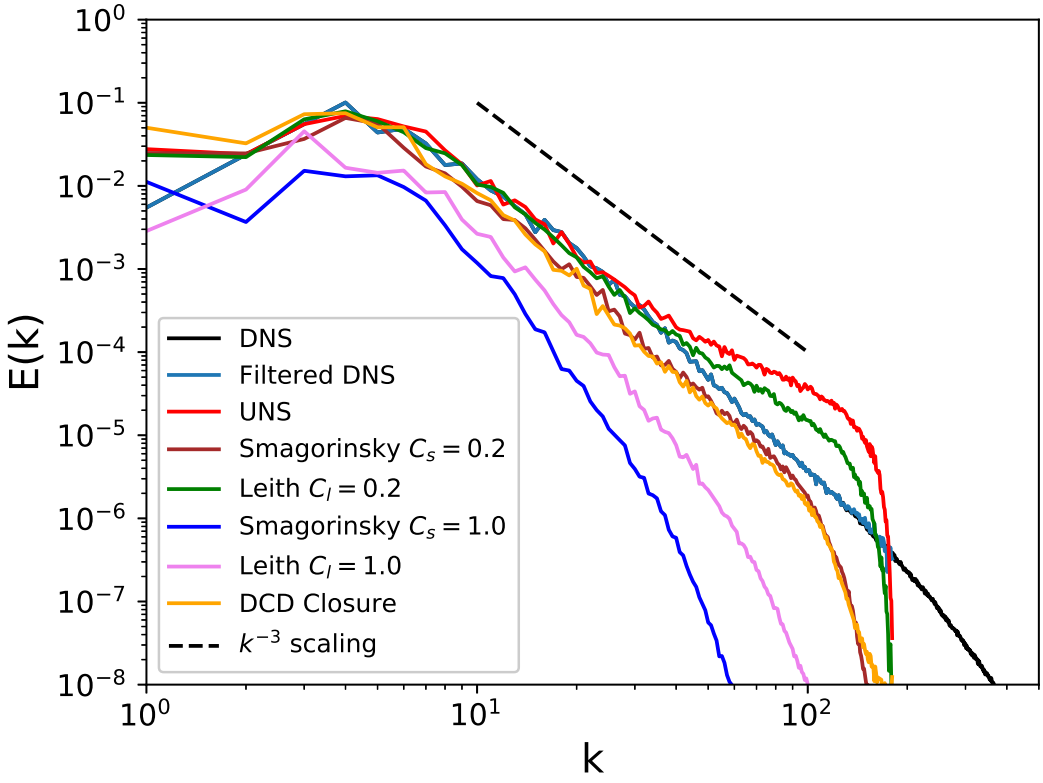


Figure 3.5: The a-posteriori performance of proposed framework for $Re = 32000$ in terms of energy spectra. At each step of sub-grid stress calculation, both forward and inverse maps are used for convolution and deconvolution in the estimation of the true underlying Jacobian.

smaller scales which are primarily affected by the proposed closure. We also demonstrate the utility of our learned map on an a-posteriori simulation for $Re = 64000$ data where similar trends are recovered as seen in statistical comparisons (Figure 3.10) and qualitative behavior (Figure 3.11). This also demonstrates an additional stringent validation of the data-driven model for ensuring generalization.

We also seek to compare the performance of the proposed framework against the dynamic formulation of the Smagorinsky and Leith models (Germano et al., 1991) modified for the vorticity and streamfunction formulation as described by Maulik and San (2017*c*) where a least-squares optimization problem is solved at two scales of resolution for an optimal value of the Smagorinsky and Leith coefficients calculated in

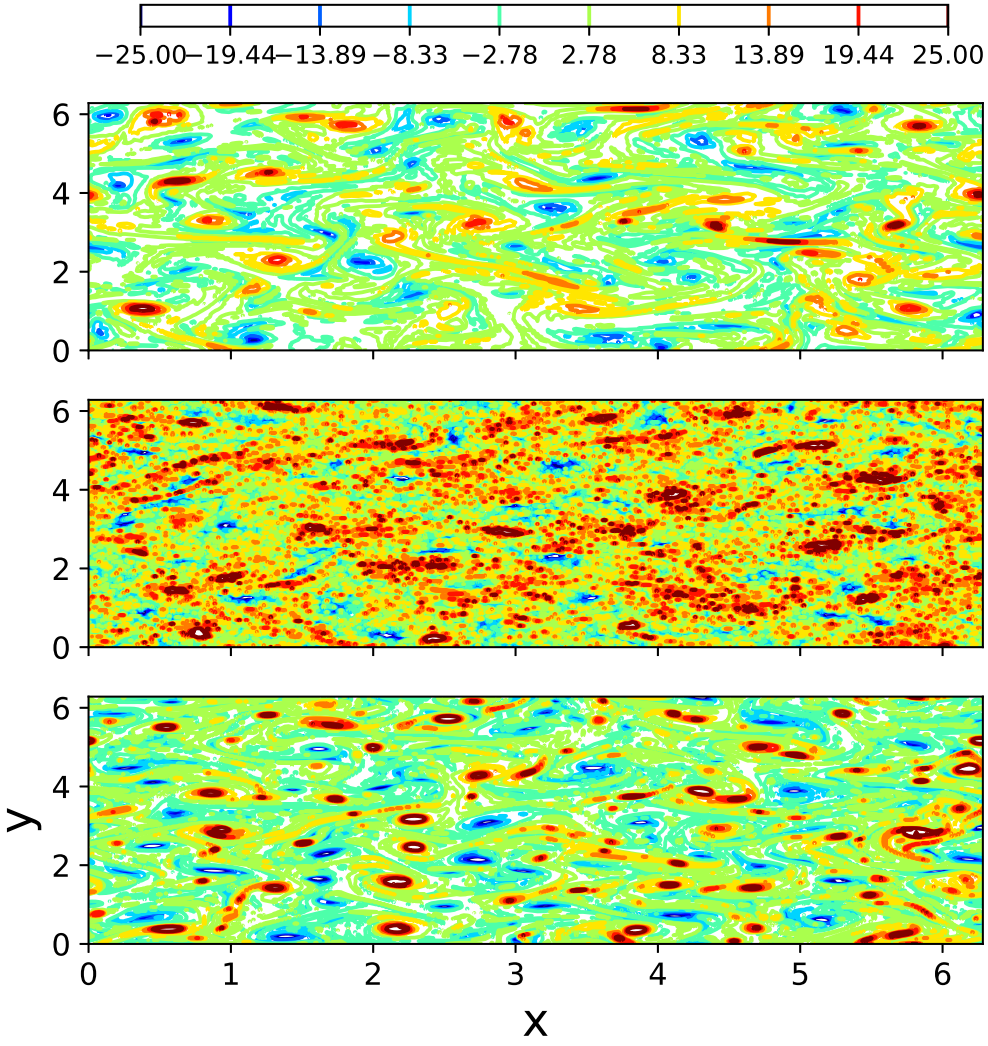


Figure 3.6: Visual quantification of the a-posteriori performance of proposed framework for $Re = 32000$ with stabilized (top), under-resolved (middle) and filtered DNS contours (bottom) for vorticity.

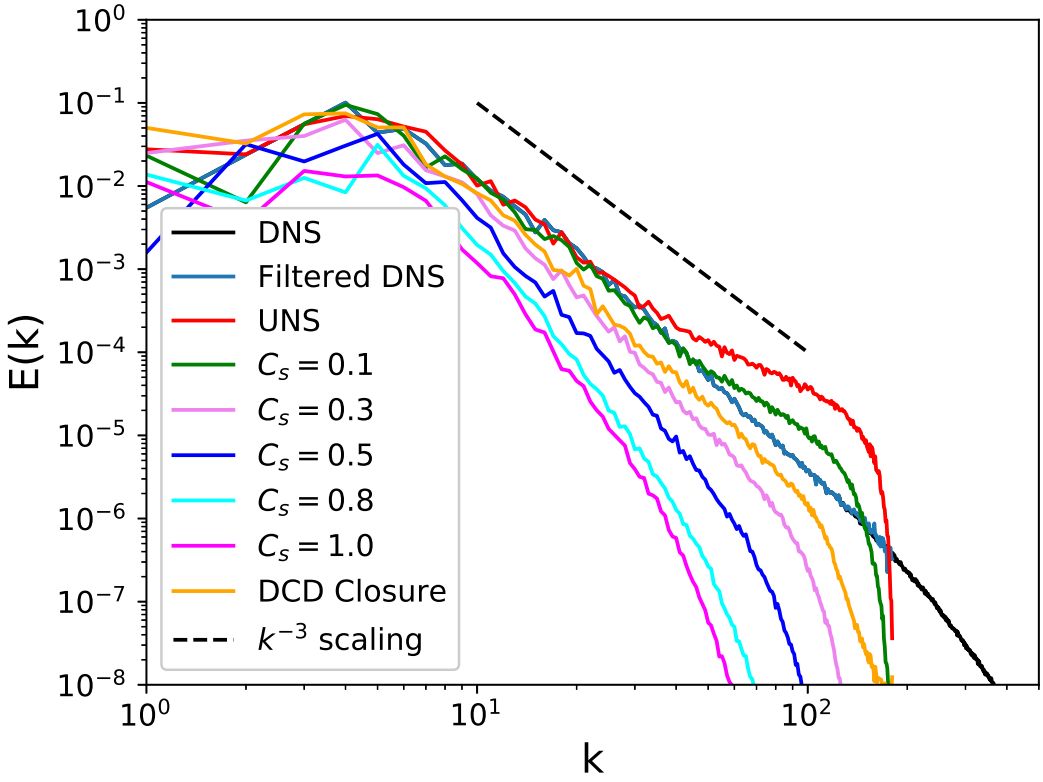


Figure 3.7: Performance comparison of proposed framework with co-efficient dependant Smagorinsky model. One can observe that higher C_s values lead to over-dissipative models.

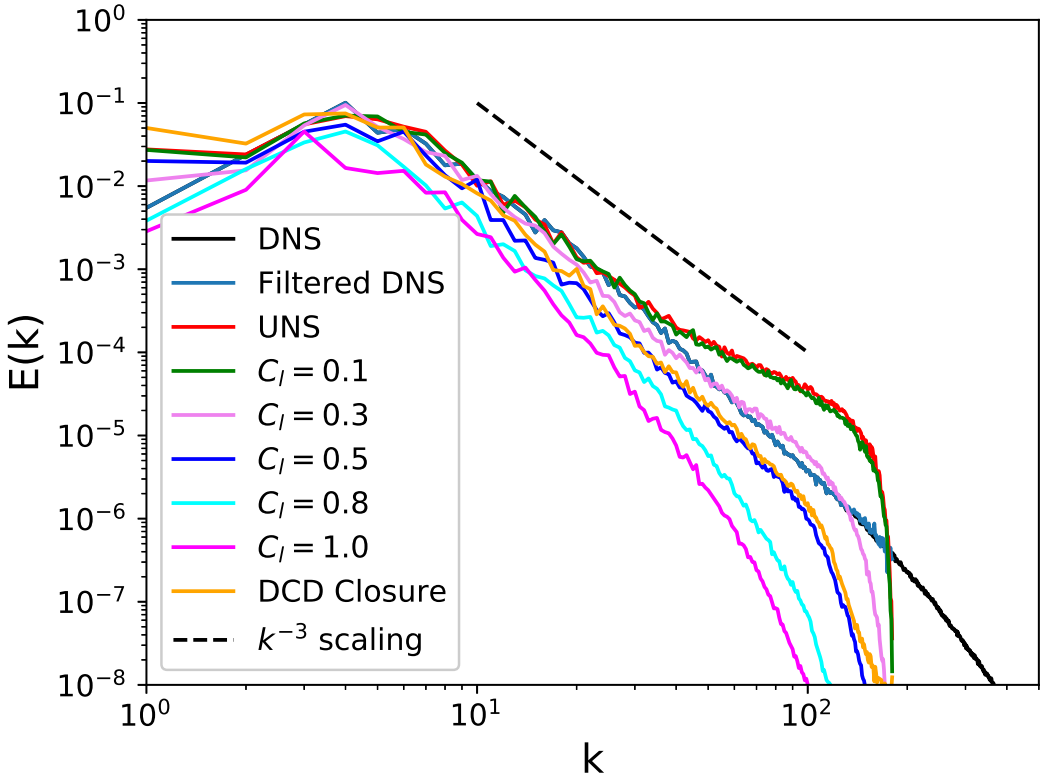


Figure 3.8: Performance comparison of proposed framework with co-efficient dependant Leith model. One can observe that higher C_l values lead to over-dissipative models.

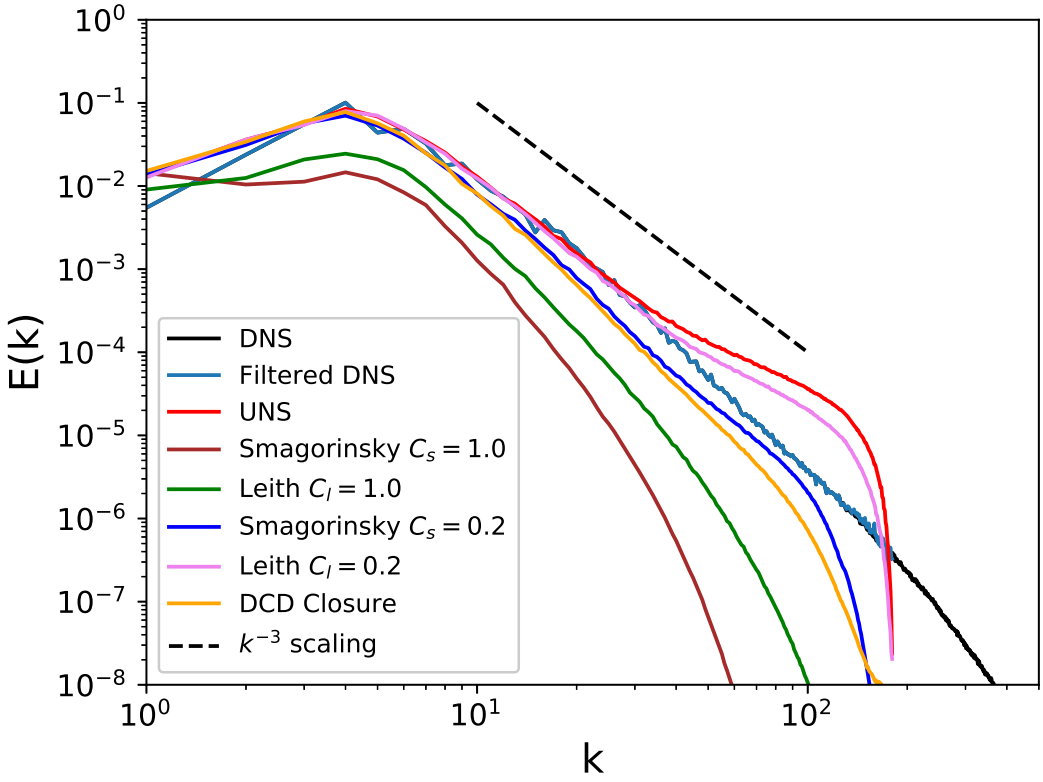


Figure 3.9: Ensemble-averaged a-posteriori performance of proposed framework for $Re = 32000$ in terms of energy spectra. This determines the generalizability of proposed framework.

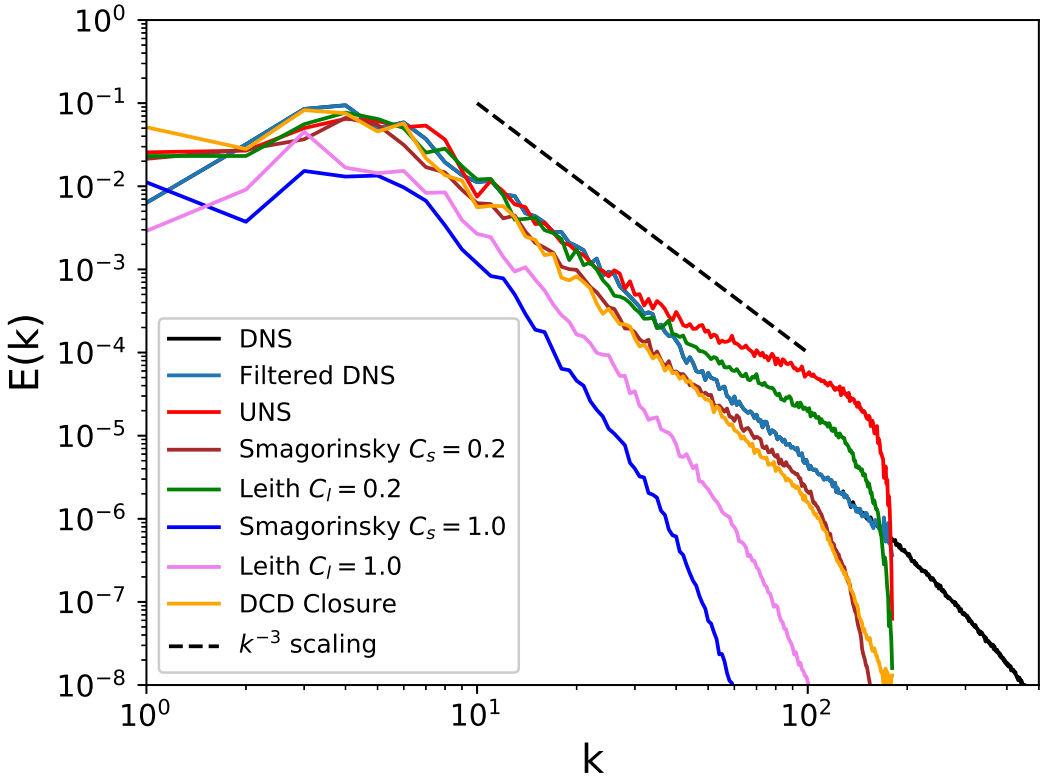


Figure 3.10: The a-posteriori performance of proposed framework for $Re = 64000$ in terms of energy spectra. Training data limited to $Re = 32000$ only.

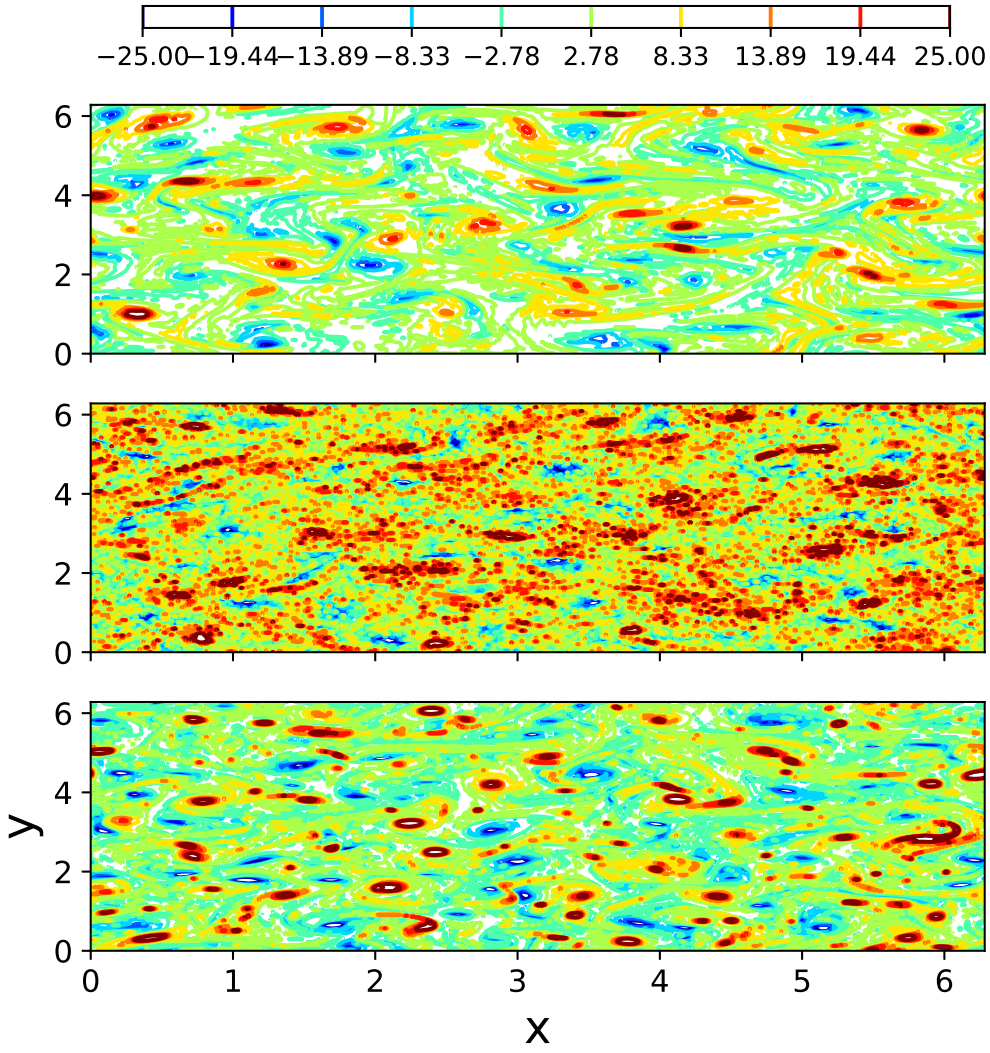


Figure 3.11: Visual quantification of the a-posteriori performance of proposed framework for $Re = 64000$ with stabilized (top), under-resolved (middle) and filtered DNS contours (bottom) for vorticity. Note: Training only with $Re = 32000$ data.

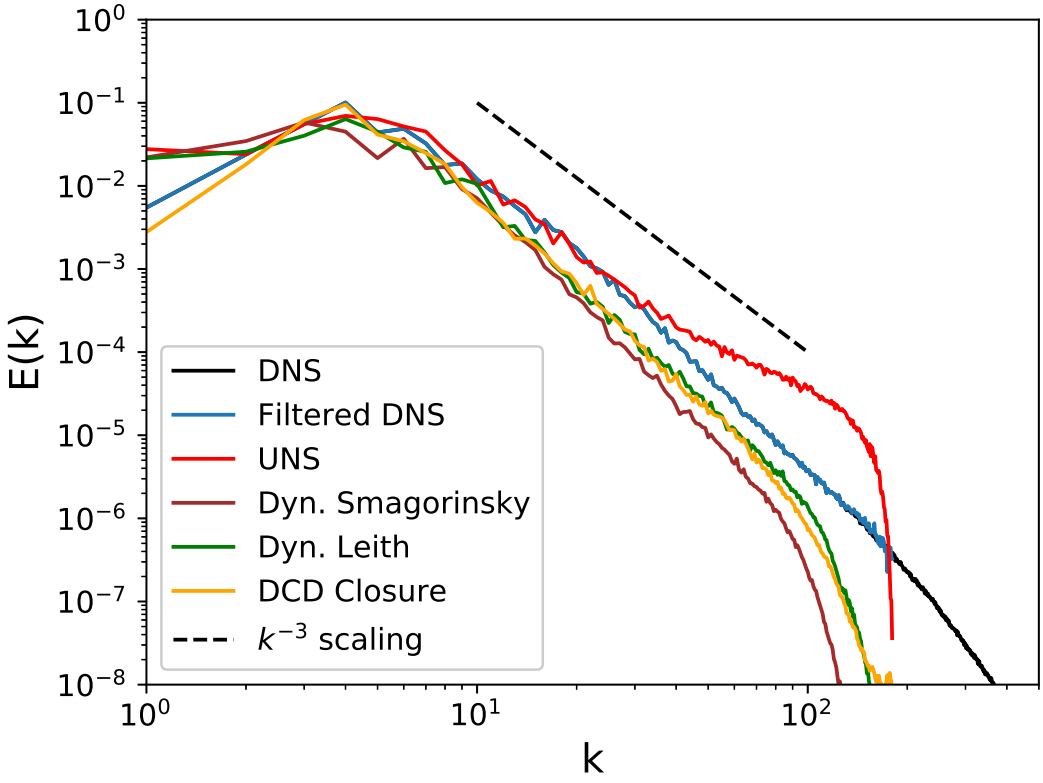


Figure 3.12: A comparison of the proposed framework with the Dynamic Smagorinsky and Dynamic Leith models for $Re = 32000$. One can see an optimal solution being obtained by the data-driven formulation in a similar manner.

a dynamic fashion defining a test filter. We note that even the dynamic formulation requires the specification of an a-priori characteristic filter-width ratio (i.e., a ratio between test and grid filters), κ, which affects a-posteriori results. In this comparison, we have utilized a filter-width ratio of $\kappa = 2$ with the use of an explicit trapezoidal filter. The results of this comparison with our framework are shown for Reynolds numbers of $Re = 32000$ and $Re = 64000$ in Figures 3.12 and 3.13 respectively. One can observe that the performance of the dynamic implementations of our eddy-viscosity hypotheses are recreated in a qualitative fashion. Our model may thus be assumed to be both data-driven and dynamic in nature.

In terms of computational cost, we remark that the proposed framework adds a

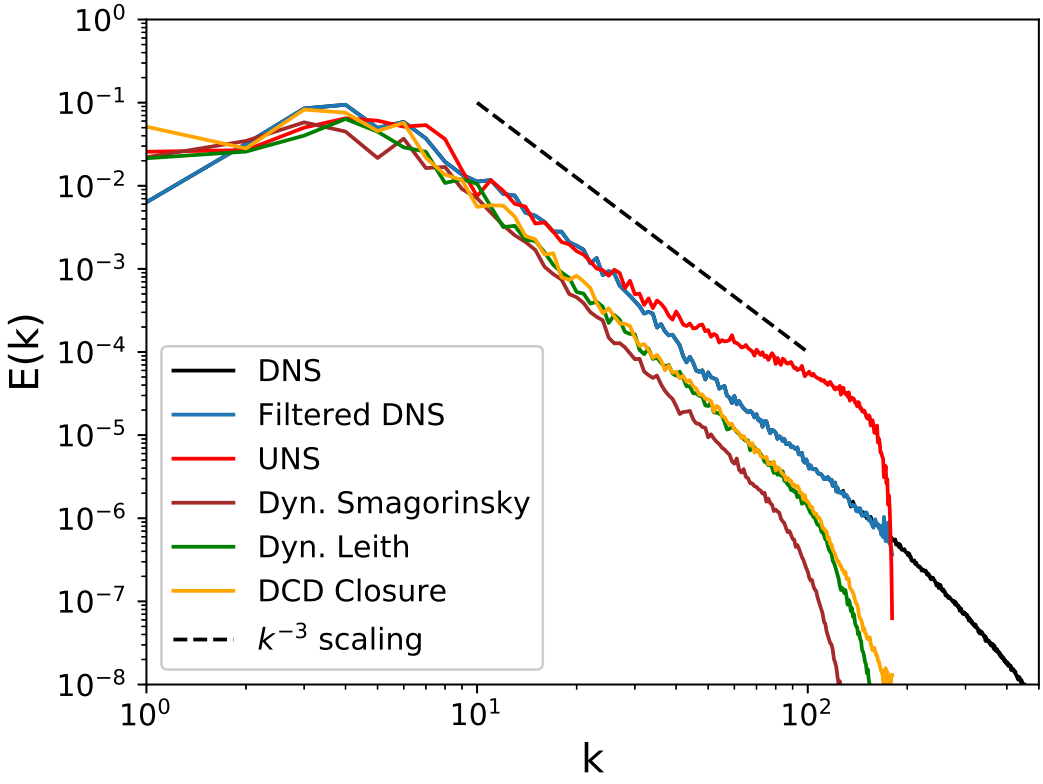


Figure 3.13: A comparison of the proposed framework with the Dynamic Smagorinsky and Dynamic Leith models for $Re = 64000$. One can see an optimal solution being obtained by the data-driven formulation in a similar manner. Training data limited to $Re = 32000$ only.

considerable computational expenditure (a-posteriori simulations led to 4 times the computational cost of the dynamic formulation) in the serial formulation. However, scalable deployments of the proposed framework in distributed environments are a subject of ongoing investigation for reducing this cost. While the data-driven framework promises more accuracy through exposure to multiple sources of turbulence data, its scalable deployment remains an important open question for successful integration into modern computational fluid dynamics solvers.

3.7 Sensitivity study

We investigate the robustness of our framework by ensuring that an optimal number of hidden layers or neurons have been utilized through an a-posteriori sensitivity study where a varying number of layers and neurons are tested for spectral scaling recovery. By keeping the default network architecture as a one layer, 100 neuron network, we investigate the effect of reduction or increase in neurons as well the effect of the number of hidden layers. We note that our studies are performed for $Re = 64000$ as an additional cross-validation.

Figure 3.14 shows the effect of varying network depths, where it can be seen that a one-layered architecture performs sufficiently accurately to be considered optimal for deployment. This hints at a simpler non-linear relationship between the inputs and outputs which has been captured by our framework. Figure 3.15 shows the effect of the number of neurons, where once again, it is observed that reduced model complexity does not impede performance. While this study utilized 100 neurons in the single hidden layer, even 10 would suffice for accurate scaling recovery. These observed behaviors imply that our framework allows for reduced network depths and reduced neurons and their associated computational advantages during training and deployment. However, we must caution the reader that greater amounts of data would necessitate deeper architectures for more generalization. In particular, our ex-

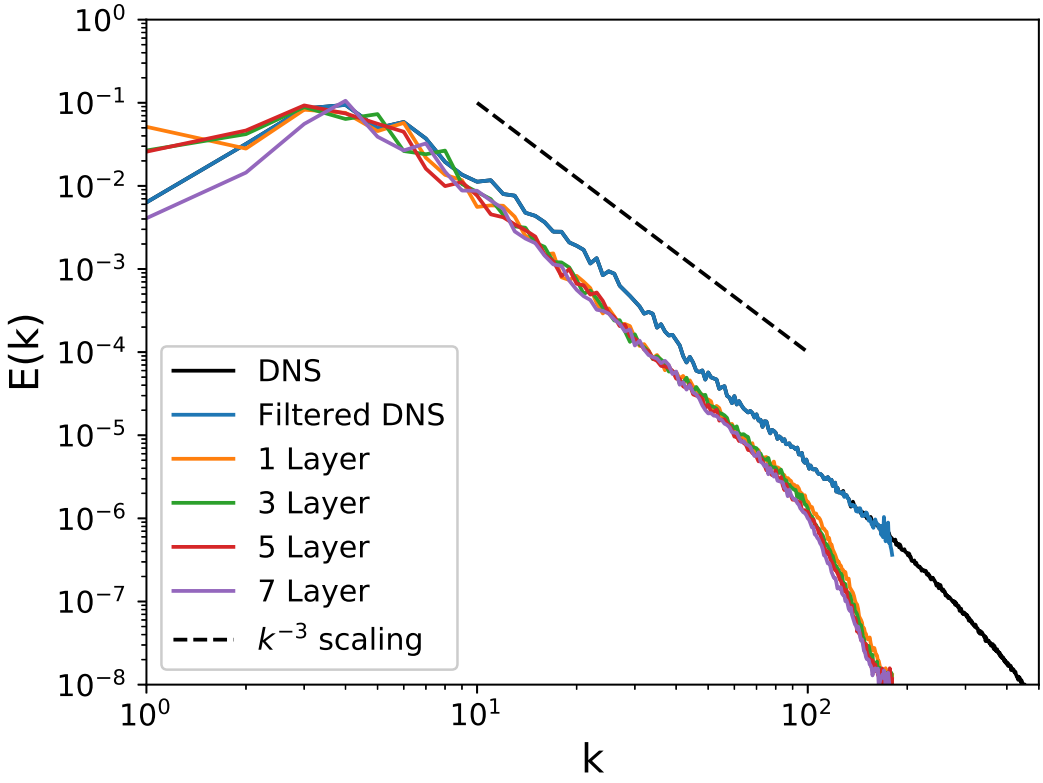


Figure 3.14: Sensitivity study for proposed framework number of layers at $Re = 64000$. Training data limited to $Re = 32000$ only and with 100 neurons in each layer.

pectation is that if multiple flow scenarios were to be learned, simple feed-forward ANNs may prove to be inadequate. In particular, we note that our choice of localized sampling, network architecture and training loss-function are chosen specific to the resolution loss and physics at hand. Greater generalization (through improved diversity of training data) would require revised hyperparameter study.

For our problem of choice, it is evident that a 10 neuron, 1 layer ANN is sufficiently viable for estimating both $\mathbb{M}_1$ and $\mathbb{M}_2$. This lends evidence to the fact that our dual network formulation may also allow for simpler learning algorithms (i.e., for this particular problem). We perform an a-priori sensitivity study for training and test mean-squared-error measures for three other well-known statistical learning algorithms such as a linear regressor (LR), a random-forest regressor (RF) (Liaw and

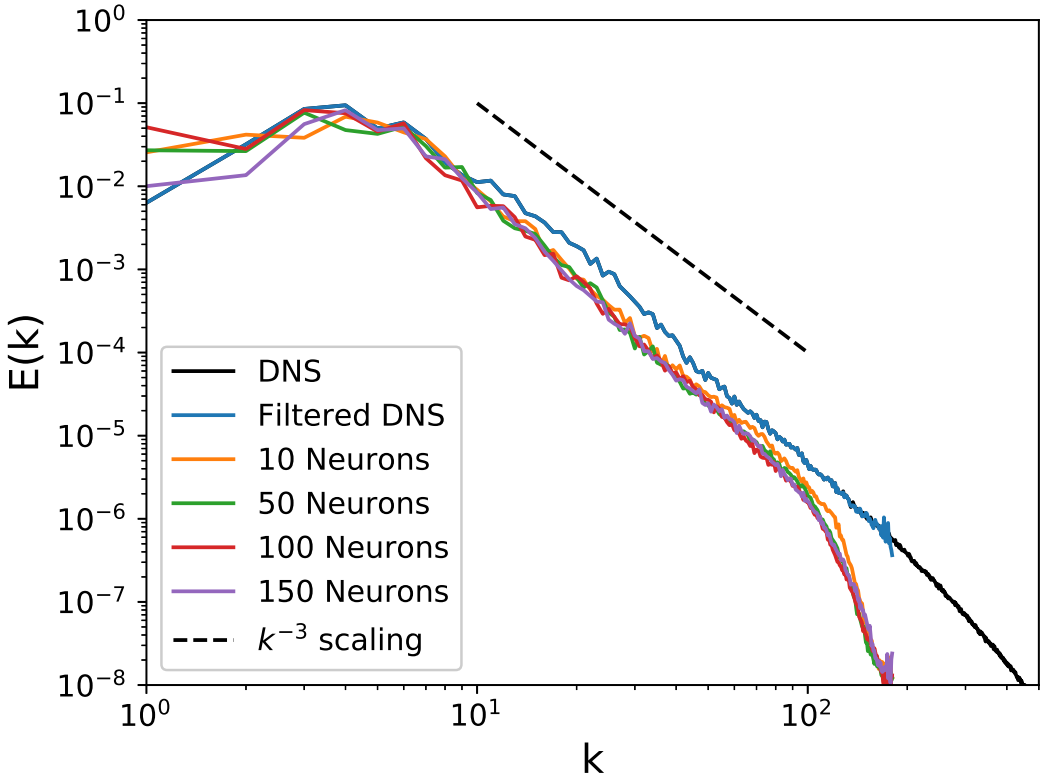


Figure 3.15: Sensitivity study for proposed framework number of layers at $Re = 64000$. Training data limited to $Re = 32000$ only and with 1 hidden layer only.

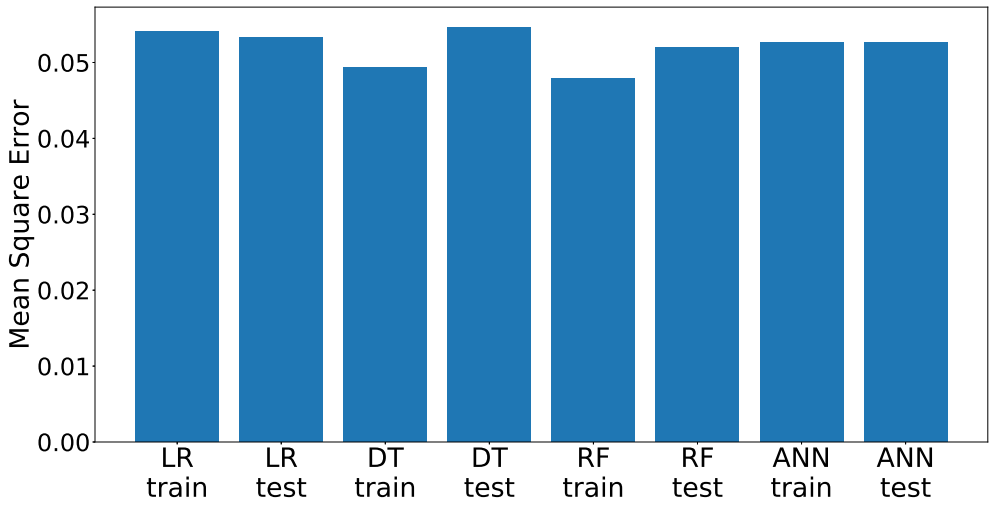


Figure 3.16: Sensitivity study for machine learning algorithm for training and testing mean-squared-errors. These errors are shown for $\mathbb{M}_2$ optimization.

Wiener, 2002) and a decision-tree regressor (DT) (Safavian and Landgrebe, 1991). We utilize the open-source scikit-learn machine learning library in python for standard implementations of these techniques. A quantitative training and testing mean-squared-error performance for these techniques in comparison to the ANN is shown in Figure 3.16 where it is observed that similar performance characteristics are observed despite vastly different learning methodologies for $\mathbb{M}_2$ optimization. It can thus be concluded that the utilization of our dual network framework has led to the simplification of a highly non-linear problem to one that is tractable for linear learning methods.

The linear-regressor is also implemented in an a-posteriori manner as shown in Figures 3.17 and 3.18 for $Re = 32000$ and $Re = 64000$ respectively. The kinetic energy spectra predictions of these linear relationships which estimate the convolutional and deconvolutional relationships are slightly less dissipative in the inertial and grid cut-off length scales for the $Re = 32000$ case. However, very similar performance is obtained for $Re = 64000$. The slightly discrepancy in the performance of the linear

implementations of the convolutional and deconvolutional maps may be attributed to a lower generalizability of the simpler nature of its learning. However, we would like to remark that this has positive implications for the utility of these techniques for the preservation of the solenoidal constraint and frame-invariance in higher-dimensional flows (Stolz and Adams, 1999) on structured grids. We would also like to note that the utilization of the same data-local filter stencil in all locations of the specified mesh ensures Galilean invariance (Razafindralandy et al., 2007). In addition, the use of stencil inputs is philosophically aligned with Moser et al. (2009), where multi-point input data are used for optimal LES formulations. However, further research is necessary for importing concepts related to isotropization of these data-driven filter and inverse kernels for application to general unstructured grids. It is also necessary to explore the possibilities of 'constrained-learning' which may embed the preservation of the solenoidal constraint in higher-dimensions through penalties introduced to the loss-functions (Raissi and Karniadakis, 2018). That is a subject of on-going investigation.

3.8 Modified truncation via mean filtering

The truncation specified in Equation 3.15 and Figure 3.4 leads to an asymmetry in the estimation of the dissipation by finer wavenumbers. To that end, we introduce a modified truncation kernel based on a local-averaging for an added truncation of positive eddy-viscosity predictions to ensure a balance with backscatter. This is introduced through the concept of a locally-averaged eddy-viscosity prediction, for instance, given by

$$\nu^{av}_{i,j} = \frac{1}{9}\left(\nu^{e}_{i,j} + \nu^{e}_{i,j+1} + \nu^{e}_{i,j-1} + \ldots + \nu^{e}_{i-1,j-1}\right), \tag{3.16}$$

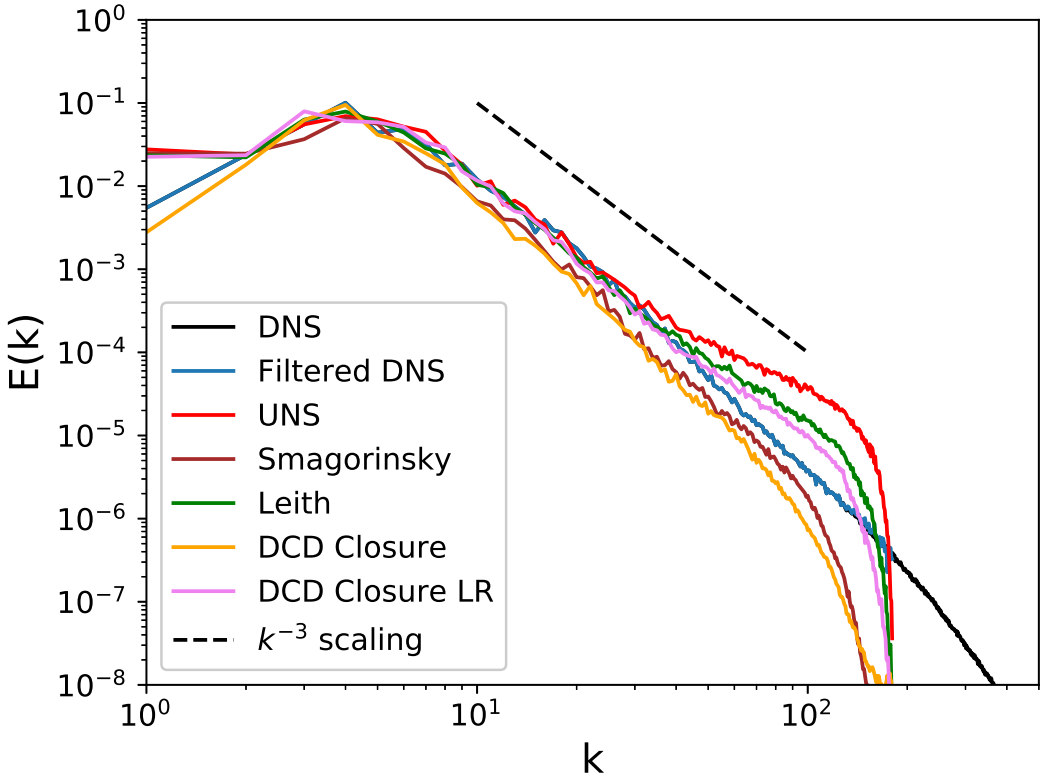


Figure 3.17: The performance of a linear estimator (LR) for convolutional and deconvolutional maps in the proposed framework for $Re = 32000$. A comparison to the default ANN is shown.

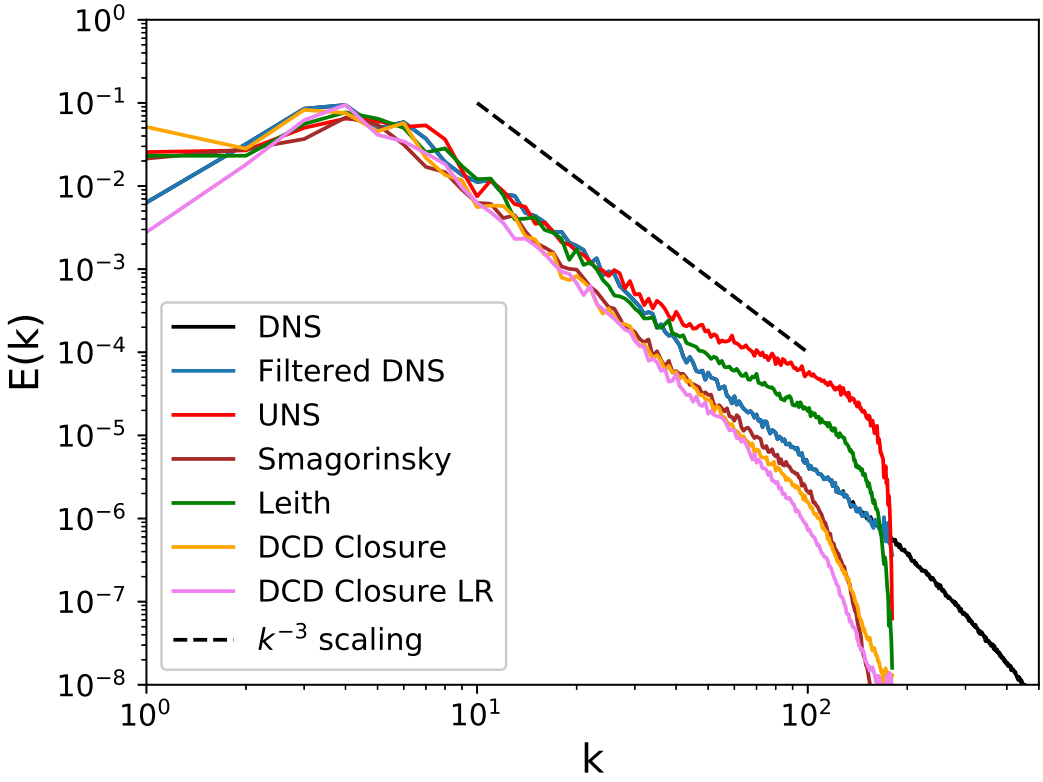


Figure 3.18: The performance of a linear estimator (LR) for convolutional and deconvolutional maps in the proposed framework for $Re = 64000$. A comparison to the default ANN is shown. Training data limited to $Re = 32000$ only.

where

$$\nu^{e}_{i,j} = \frac{\tilde{\Pi}_{i,j}}{\nabla^2 \bar{\omega}_{i,j}}. \tag{3.17}$$

The averaging procedure in Equation 3.16 may also be represented by a mean-filtering-kernel given as

$$\nu^{av} = \frac{\nu^{e}}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}. \tag{3.18}$$

The transfer-function of this kernel may be visualized as shown in Figure 3.19 and this averaging filter has the effect of eliminating localized pointwise values which are unrepresentative of their surroundings.

The quantity $\nu^{av}_{i,j}$ is basically the averaged dissipative (or energy-producing) nature of the local stencil of prediction and the quantity $\nu^{e}_{i,j}$ is the local effective eddy-viscosity prediction by our proposed framework. Our truncation scheme is then expressed as

$$\Pi_{i,j} = \begin{cases} \tilde{\Pi}_{i,j}, & \text{if } \nu^{av}_{i,j} > \nu^{e}_{i,j} \\ 0, & \text{otherwise.} \end{cases} \tag{3.19}$$

The effect of this modified truncation is described in Figure 3.20 where an increased truncation is observed quite clearly. Our model formulation may thus be assumed to preserve the statistical nature of the negative-eddy viscosities in a locally-averaged manner.

A posteriori deployments of this modified truncation scheme are displayed in Figures 3.21 and 3.22 where an improved capture of the inertial range is observed for $Re = 32000$ and $Re = 64000$ respectively. This implies that the statistical fidelity of

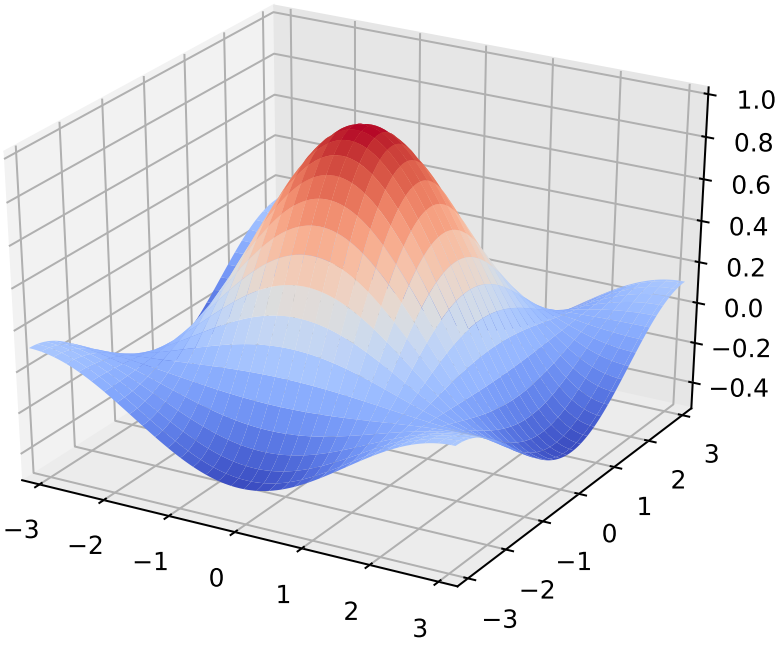


Figure 3.19: Transfer function for truncation kernel to preserve statistical effects of backscatter.

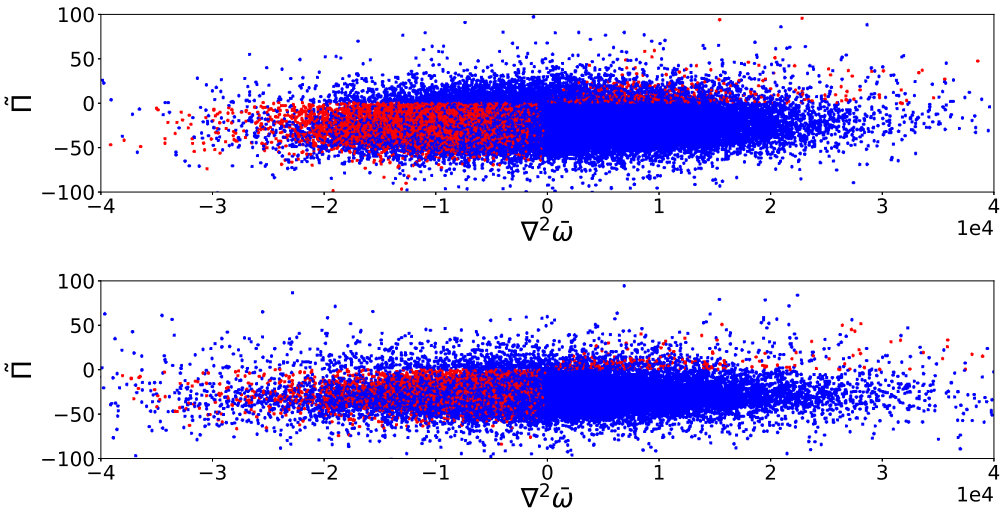


Figure 3.20: A visual assessment of the truncation of our numerical post-processing during deployment given by the BS-1 framework. Blue points indicate truncated deployment for ensuring no negative viscosity and numerical stability. A-priori predictions for $Re = 32000$ (top) and $Re = 64000$ (bottom) shown.

the prediction has been improved by the integration of a local backscatter estimate. The combination of novel truncation strategies may further be studied in the context of this data-driven framework for close agreement with theoretical scaling laws.

3.9 Concluding remarks

In this investigation, we have put forth and analyzed a physics-informed data-driven closure modeling framework for non-linear partial differential equations. Our proposal is to use two single-layer feed-forward artificial neural networks for mapping transformations from grid-resolved variables with missing wavenumber content and subsampled direct numerical simulation data in order to close the two-dimensional Navier-Stokes equations. This investigation continues from the authors' previous work (Maulik and San, 2017*a*), which assessed the deconvolutional ability of neural networks, by employing them for estimating sub-grid relationships from grid-resolved variables.

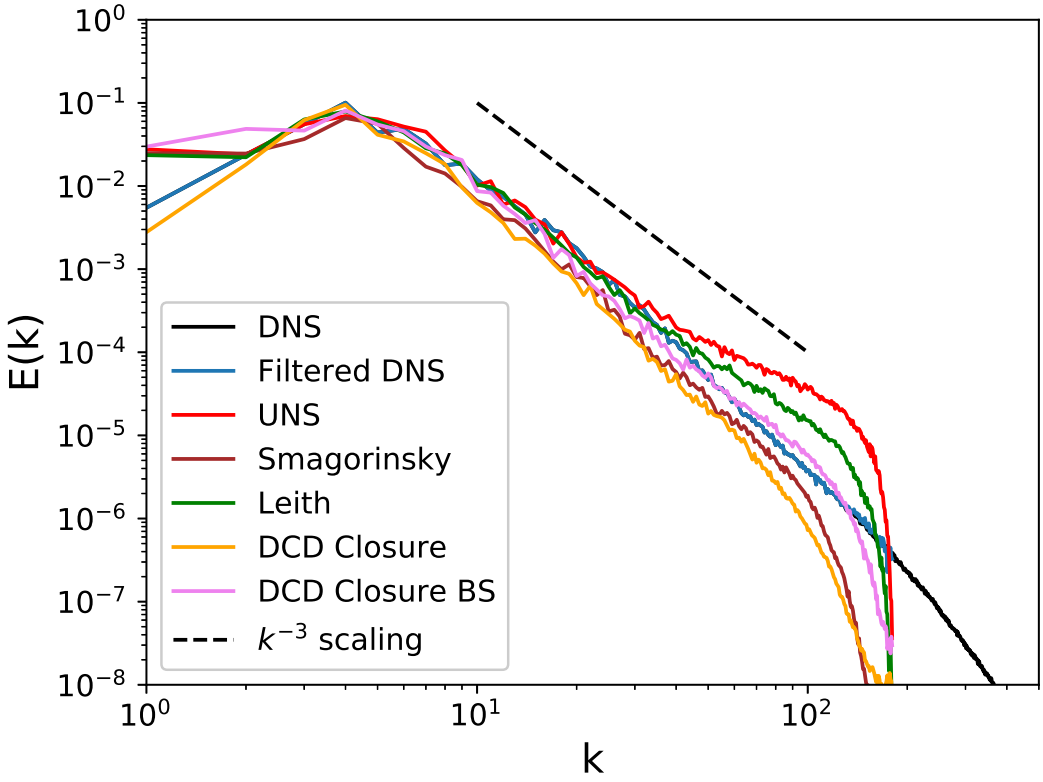


Figure 3.21: A comparison of the choice of a-posteriori truncation utilized in our proposed framework. A statistical preservation of backscatter enforced by our proposed kernel leads to a better agreement with the inertial range statistics for $Re = 32000$.

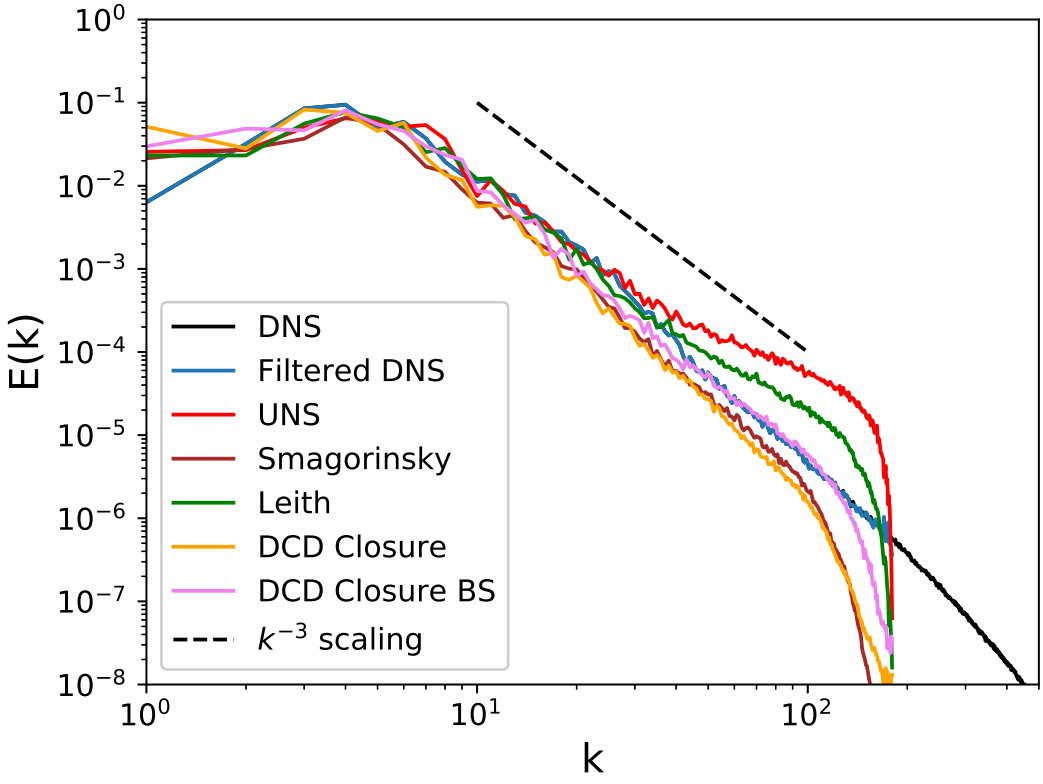


Figure 3.22: A comparison of the choice of a-posteriori truncation utilized in our proposed framework. A statistical preservation of backscatter enforced by our proposed kernel leads to a better agreement with the inertial range statistics for $Re = 64000$. Training data limited to $Re = 32000$ only.

Our framework precludes the utilization of any phenomenological arguments or model form constraints and relies, instead, solely on the approximation of the Fourier cut-off filtering inherent in coarse-graining as well as its approximate inverse. We remark that while there is truly no way to invert a Fourier cut-off filter, a-priori exposure to samples from resolved and filtered fields are used to estimate the information loss and reconstruct it. For the purpose of numerical stability, we also employ two postprocessing strategies with the first ensuring no aggregate negative viscosities in the computational domain and the second preserving backscatter in a statistical sense. This ensures that the stochastic nature of the network predictions do not trigger numerical instability amplification in an explicit flow computation.

Another important feature of this investigation is that, despite its data-driven nature, our offline training phase necessitates no exposure to the true sub-grid stress data and predictions are viable simply through the estimation of the nature of the coarse-graining process in LES. Our sensitivity study reveals the benefits of this approach, where it is seen that increasing network complexity leads to no appreciable improvement in the a-posteriori performance for this current test case. The need for complicated network architectures (and their associated computational and memory burden) is thus minimized due to the physics-informed nature of our formulation.

Comparison with other well-established linear statistical learning methods also show that the novel dual network formulation presented here reduces the complexity of learning considerably. In particular, the performance of a linear map representation of convolution and deconvolution operations ensures a direct enforcement of the solenoidal constraint on the convolved and deconvolved fields for applicability to higher dimensions. *A posteriori* realizations of the linear mappings between grid-resolved and sub-grid space, show an exhibition of the bias-variance trade-off issue where the simpler nature of the linear regressor leads to lower generalization for a different data-set. However, an effective parameter and model-form free closure is

readily obtained in this case as well.

We also note that the data-local nature of our framework with the combination of solely one map (each for convolution and deconvolution) ensures that frame-invariance is respected for the specified mesh. As a future direction, this framework shall be studied with the view of integrating physics-based constraints in the offline training phase. These may be introduced through optimization penalties for continuity enforcement and for isotropization on arbitrary meshes. These are necessary for the generalization of this framework to higher-dimensional flows with arbitrary boundary conditions.

While the results of this study have proven promising for the development of purely data-driven closures for LES, the true test of these ideologies would be to develop generalized closures for a variety of flows. In terms of a long-term goal, the preliminary results displayed here must translate to a situation where a-posteriori closure is determined by a-priori exposure to a variety of flow classes. Additionally, the stencil based formulation for a predictive map leads to a resolution dependence of the trained relationships. This is because our LES to DNS ratio is fixed during the specification of training data. An exposure to different levels of coarse-graining for potential predictions would also increase the generalizability of this framework. With that in mind, we remark that the framework proposed here represents the advantages of implementing a data-driven paradigm from a physics-informed point of view with consequent benefits for framework complexity and ease of deployment.

CHAPTER IV

Sub-grid modelling for two-dimensional turbulence using neural networks

4.1 Abstract

In this investigation, a data-driven turbulence closure framework is introduced and deployed for the sub-grid 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 sub-grid 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 sub-grid 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 sub-grid quantity inference. In addition, it is also observed that some measure of a-posteriori error must be considered during

This chapter is adapted from Maulik et al., J. Fluid Mech., 858, 122-144

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.

4.2 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 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 et al., 2005; Sagaut, 2006). It has generally been observed that the choice of the sub-grid model is physics dependant, i.e., that different flow phenomena require different expressions for sub-grid 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 and Tem-

pleton, 2015; Tracey et al., 2015; Xiao et al., 2016; Singh et al., 2017; Wang, Wu and Xiao, 2017; Wang, Wu, Ling, Iaccarino and Xiao, 2017; Weatheritt and Sandberg, 2017*b*; Schaeffer, 2017; Wu et al., 2018*a*; Raissi and Karniadakis, 2018; Wan et al., 2018; Mohan and Gaitonde, 2018), 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 and Schreck, 1997; Cohen et al., 2003; Mannarino and Mantegazza, 2014; San and Maulik, 2018) as well as for modelling methodologies for coarse-grained meshes such as LES and RANS simulations (Maulik and San, 2017*a*; Wang, Wu and Xiao, 2017; Wu et al., 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 and Moser, 1999; Moser et al., 2009; King et al., 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 sub-grid forcing quantity such as the one utilized in Ling, Kurzawski and Templeton (2016) where a deep ANN has been described for Reynolds stress predictions in an invariant subspace. ANNs have been also implemented in Parish and Duraisamy (2016) to correct errors in RANS

turbulence models after the formulation of a field-inversion step. Gamahara and Hattori (2017) detailed the application of ANNs for identifying quantities of interest for sub-grid modelling in a turbulent channel flow through the measurement of Pearson correlation coefficients. Milano and Koumoutsakos (2002) also implemented these techniques for turbulent channel flow but for the generation of low-order wall models while Sarghini et al. (2003) deployed ANNs for the prediction of the Smagorinsky coefficient (and thus the sub-grid contribution) in a mixed sub-grid model. In Beck et al. (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, sub-grid 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 sub-sampled data sets. The use of such sub-sampled data necessitates a greater emphasis on physics-distillation to prevent extrapolation and over-fitting dur-

ing 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 (Smagorinsky, 1963) and Leith (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 modeled 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 which 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 et al. (2019). 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 sub-grid modelling for the LES of two-dimensional turbulence may be summarized in the following. In terms of the vorticity-streamfunction formulation, our non-dimensional governing equation for incompressible flow may be represented as

$$\frac{\partial \omega}{\partial t} + J(\omega, \psi) = \frac{1}{Re} \nabla^2 \omega, \tag{4.1}$$

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

$$\nabla^2\psi = -\omega. \tag{4.2}$$

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

$$J(\omega,\psi) = \frac{\partial\psi}{\partial y}\frac{\partial\omega}{\partial x} - \frac{\partial\psi}{\partial x}\frac{\partial\omega}{\partial y}. \tag{4.3}$$

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

$$\frac{\partial\bar{\omega}}{\partial t} + J(\bar{\omega},\bar{\psi}) = \frac{1}{Re}\nabla^2\bar{\omega} + \Pi, \tag{4.4}$$

where the overbarred variables are now evolved on a grid with far fewer degrees of freedom. The sub-grid 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

$$\Pi = J(\bar{\omega},\bar{\psi}) - \overline{J(\omega,\psi)}. \tag{4.5}$$

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 non-linear evolution equations introduce unclosed terms that must be modeled to account for local, instantaneous momentum and energy exchange between resolved and unresolved scales. If these inter-eddy interactions are not properly parameterized, then an increase in resolution will not necessarily improve the accuracy of these large scales (Frederiksen and Zidikheri, 2016; Frederiksen et al., 2013). Addi-

tionally, 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 and Greatbatch, 2008; Fox-Kemper et al., 2011; San et al., 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 sub-grid forcing. We note here that the definition of the sub-grid source term given in Equation 4.5 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 et al., 1998; Marshall and Beninati, 2003).

4.3 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 non-linear 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 multi-step 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

element-wise functional modification of the incident vector to generally introduce non-linearity 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 $\mathbf{p}$ resides in a P-dimensional space and our desired output $\mathbf{q}$ resides in a Q-dimensional space, this framework establishes a map $\mathbb{M}$ as follows:

$$\mathbb{M} : \{p_1, p_2, \ldots, p_P\} \in \mathbb{R}^P \rightarrow \{q_1, q_2, \ldots, q_Q\} \in \mathbb{R}^Q. \tag{4.6}$$

A schematic for this map may be observed in Figure 4.1, where input, output and hidden spaces are summarized. In equation form, our default optimal map is given by

$$\begin{aligned} \mathbb{M} : \{&\bar{\omega}_{i,j}, \bar{\omega}_{i,j+1}, \bar{\omega}_{i,j-1}, \ldots, \bar{\omega}_{i-1,j-1}, \\ &\bar{\psi}_{i,j}, \bar{\psi}_{i,j+1}, \bar{\psi}_{i,j-1}, \ldots, \bar{\psi}_{i-1,j-1}, |\bar{S}|_{i,j}, |\nabla\bar{\omega}|_{i,j}\} \in \mathbb{R}^{20} \rightarrow \{\tilde{\Pi}_{i,j}\} \in \mathbb{R}^1. \end{aligned} \tag{4.7}$$

where

$$|\bar{S}| = \sqrt{4\left(\frac{\partial^2\bar{\psi}}{\partial x \partial y}\right)^2 + \left(\frac{\partial^2\bar{\psi}}{\partial x^2} - \frac{\partial^2\bar{\psi}}{\partial y^2}\right)^2}, \quad |\nabla\bar{\omega}| = \sqrt{\left(\frac{\partial\bar{\omega}}{\partial x}\right)^2 + \left(\frac{\partial\bar{\omega}}{\partial y}\right)^2} \tag{4.8}$$

are eddy-viscosity kernel information input to the framework and $\tilde{\Pi}$ is the approximation to the true sub-grid 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 Equation 4.7 is considered 'default' due to the utilization of a 9-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:

$$\Pi = \begin{cases} \tilde{\Pi}, & \text{if } (\nabla^2\bar{\omega})(\tilde{\Pi}) > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{4.9}$$

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 Bousinessq hypothesis for functional sub-grid 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 sub-grid scales). Similarly, the forcing of vorticity at the finer scales may be assumed to be an localized event of high kinetic energy dissipation. In general, Equation 4.9 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 dependency on model-forms or coefficient calculations. In what follows for the rest of this document, our proposed

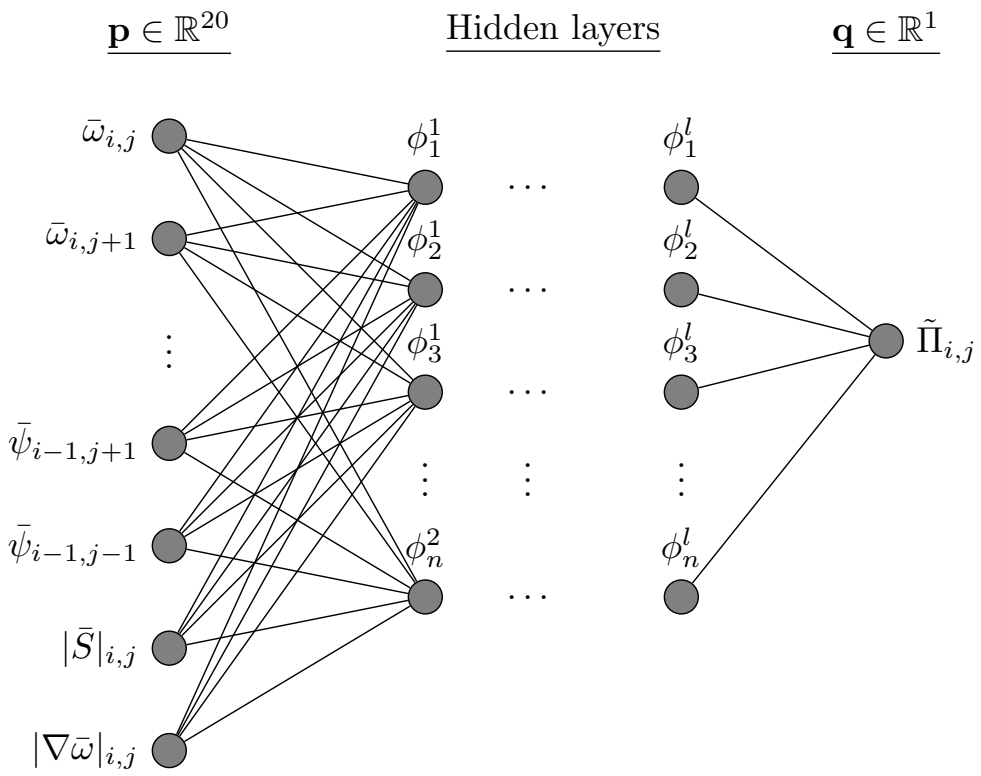


Figure 4.1: Proposed artificial neural network architecture and relation to sampling and prediction space.

framework is denoted ANN-SGS. Details related to hyper-parameter selection and supervised learning of the model are provided in the appendices.

4.4 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 sub-grid 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 behavior for different choices of model coefficients (utilized in effective eddy-viscosity computations using mixing-length based phenomenological arguments). The Smagorinsky or Leith sub-grid scale models may be implemented in the vorticity-streamfunction formulation via the specification of

an effective eddy-viscosity

$$\tilde{\Pi} = \nu_e \nabla^2 \bar{\omega}, \tag{4.10}$$

where the Smagorinsky model utilizes

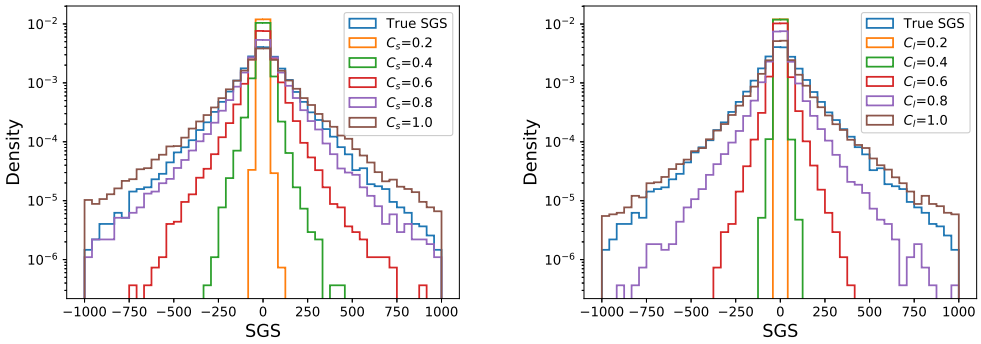
$$\nu_e = (C_s \delta)^2 |\bar{S}|, \tag{4.11}$$

while the Leith hypothesis states

$$\nu_e = (C_l \delta)^3 |\nabla \bar{\omega}|. \tag{4.12}$$

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 4.1, it is apparent that the choice of model-form coefficients C_s and C_l for the Smagorinsky and Leith models dictate 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 readers attention to the fact that ideal reconstructions of the true sub-grid term are with coefficients near the value of 1.0, a value that is rather different to the theoretically accepted values of C_s applicable in three-dimensional turbulence. This dependance 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 behavior.

In contrast, Figure 4.3 shows the performance of the proposed framework in predicting sub-grid contributions purely through the indirect exposure to supervised data in the training process. The figure shows a remarkable ability for Π reconstruction for

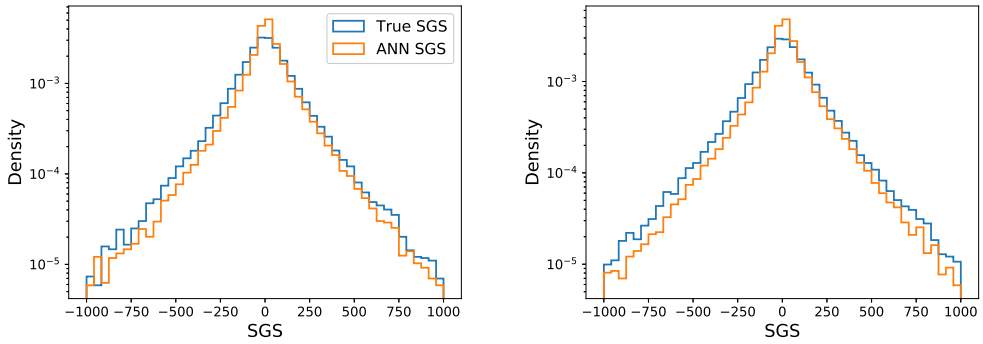


(a) Smagorinsky (b) Leith

Figure 4.2: A-priori performance of Smagorinsky and 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.

both Re values of 32000 and 64000, solely from grid-resolved quantities. Performance similar to ideal model-coefficients mentioned in the previous figure are also observed. The $Re = 64000$ 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 data set with more energetic physics. We may thus conclude that the map has managed to embed a relationship between sharp spectral cutoff filtered quantities and sub-grid source terms.

We also visually quantify the effect of Equation 4.9 (described for the process of numerical realizability) in Figure 4.4 where a hardwired truncation is utilized for precluding violation of viscous stability in the forward simulations of our learning deployment. One can observe that the blue regions of the figure, which are spatial locations of sub-grid forcing ($\tilde{\Pi}$) and Laplacian $\nabla^2\bar{\omega}$ being the 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 sub-grid 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



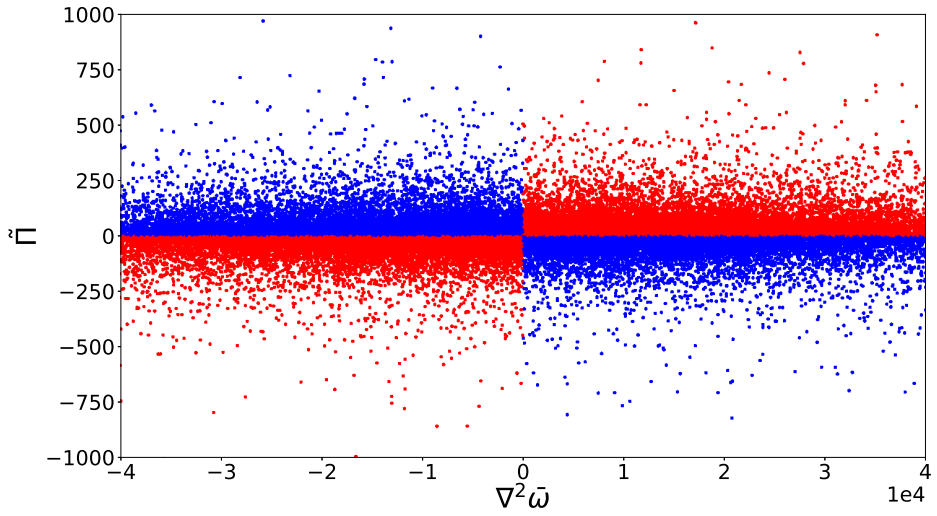
(a) $Re = 32000$ (b) $Re = 64000$

Figure 4.3: A-priori results for the probability density distributions of the true and framework predicted LES source terms for $Re = 32000$ (left) and $Re = 64000$ (right). Note that the training data was generated for $Re = 32000$ only and prediction on $Re = 64000$ represents a stringent validation.

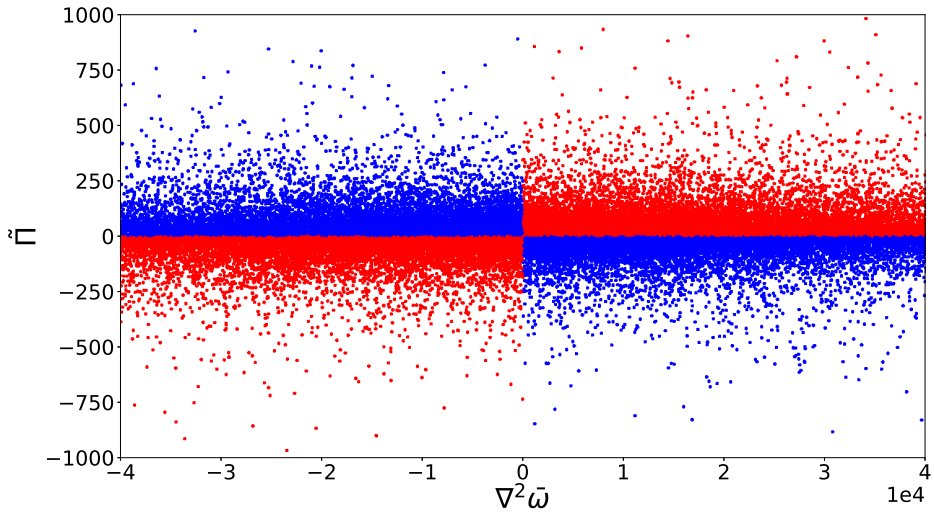
roughly half of the predicted sub-grid terms are truncated matches the observations in Piomelli et al. (1991) where it is observed that forward and backscatter are present in approximately equal amounts when extracted from DNS data. Studies are underway to extend some form of dynamic localization of backscatter to the current formulation along the lines of Ghosal et al. (1995).

4.5 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 and 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 setup 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 are



(a) $Re = 32000$

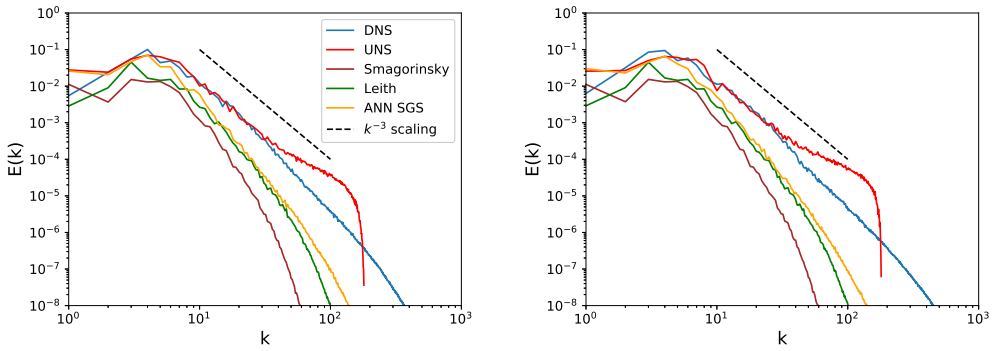


(b) $Re = 64000$

Figure 4.4: An a-priori assessment of the nature of truncation given by Equation 4.9 for $t = 2$ snapshot data at $Re = 32000$ (top) and $Re = 64000$ (bottom). The nature of this truncation is for the preservation of viscous stability in a coarse-grained forward simulation.

through second-order discretizations for all spatial quantities (with a kinetic-energy conserving Arakawa discretization (Arakawa, 1966) for the calculation of the non-linear 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 4.5 displays the statistical fidelity of coarse-grained simulations obtained with the deployment of the proposed framework for $Re = 32000$. 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 cut-off 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 sub-grid 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 $Re = 64000$ 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 (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

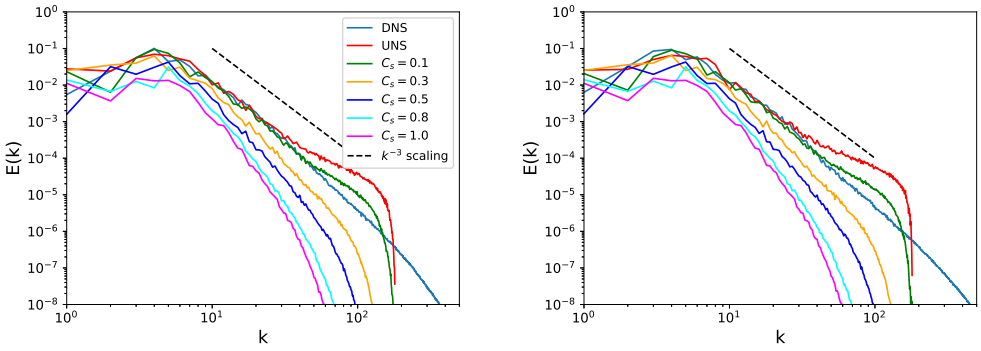


(a) $Re = 32000$ (b) $Re = 64000$

Figure 4.5: A-posteriori results for the spatially-averaged kinetic energy spectra for the proposed framework compared with DNS and UNS solutions. Note that only $Re = 32000$ training data is used for both deployments and network is applied spatially and temporally in a dynamic manner until $t = 4$.

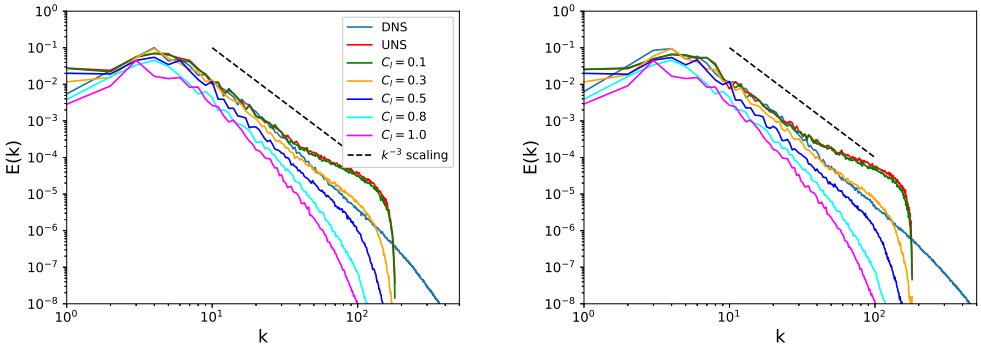
single-valued and varies depending on resolution and flow characteristics (Galperin and Orszag, 1993; Canuto and Cheng, 1997; Vorobev and Zikanov, 2008) with higher values specifically for geophysical flows (Cushman-Roisin and Beckers, 2011). In comparison, the proposed framework has embedded the adaptive nature of dissipation into its map which is a promising outcome. Figures 4.6 and 4.7 show the performance of the Smagorinsky and Leith models, respectively, for a $Re = 32000$ and $Re = 64000$ 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 64000 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 behavior 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



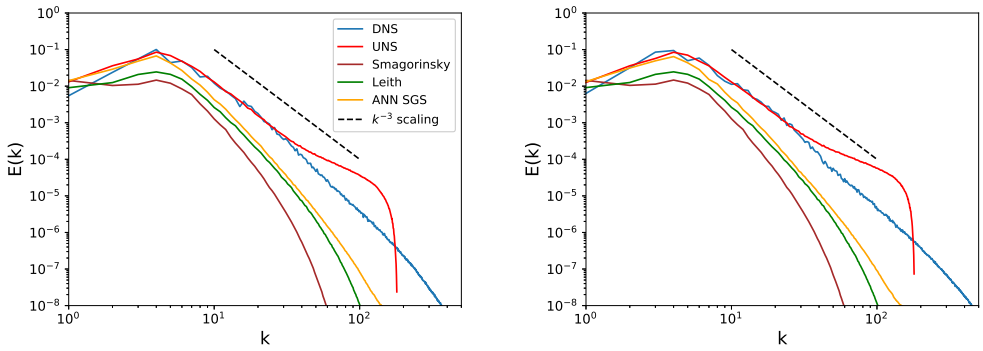
(a) $Re = 32000$ (b) $Re = 64000$

Figure 4.6: 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 dependant on the value of these coefficients.



(a) $Re = 32000$ (b) $Re = 64000$

Figure 4.7: 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 dependant on the value of these coefficients.

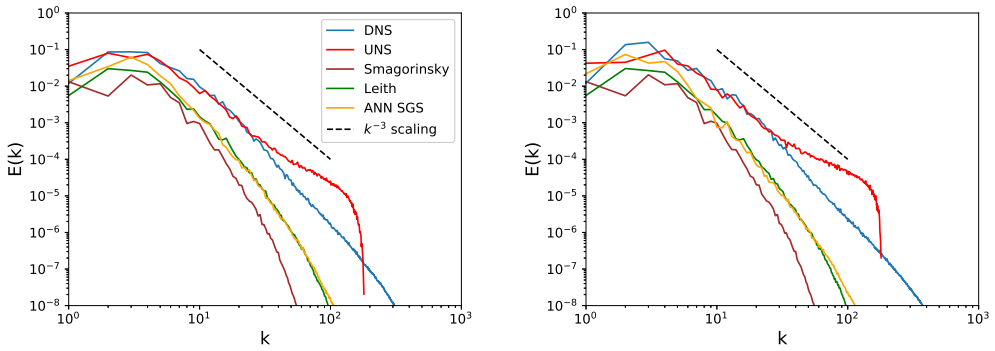


(a) $Re = 32000$ (b) $Re = 64000$

Figure 4.8: A-posteriori results for 24 ensemble-averaged simulations for $Re = 32000$ (left) and $Re = 64000$ (right).

which 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 $Re = 32000$ and at $Re = 64000$, 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 4.8. 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 vorticity fields sharing the same energy spectra. Details related to the generation of these random initial conditions may be found in Maulik and San (2017*c*). In addition, we also display spectra obtained from an a-posteriori deployment of our framework till $t = 6$ for $Re = 32000$ and $Re = 64000$, shown in Figure 4.9, which ensures that the model has learned a sub-grid closure effectively and predicts the vorticity forcing adequately in a temporal region that it has not been exposed to during training.

Figure 4.10 shows a qualitative assessment of the stabilization property of machine learning framework where a significant reduction in noise can be visually ascertained due its deployment. Coherent structures are retained successfully as against UNS



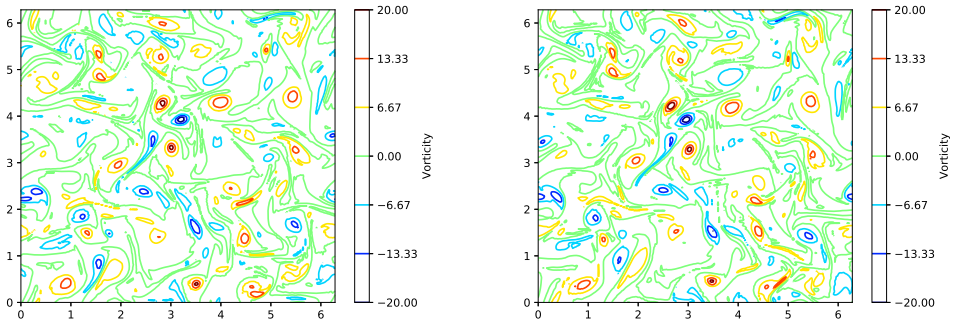
(a) $Re = 32000$ (b) $Re = 64000$

Figure 4.9: The deployment of our framework till $t = 6$ for $Re = 32000$ (left) and $Re = 64000$ (right) showing that a sub-grid 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.

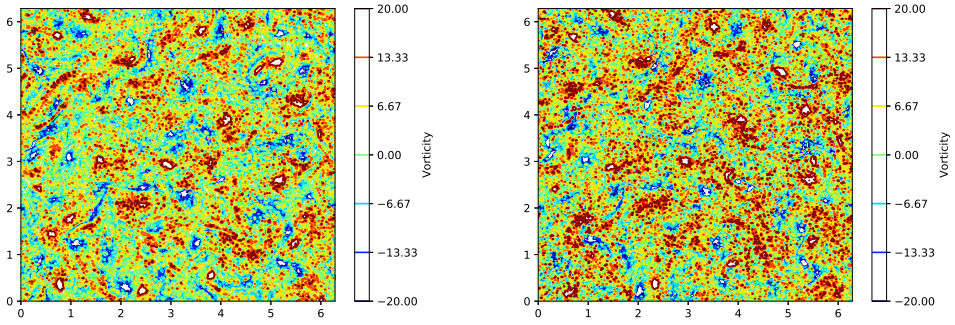
results where high-wavenumber noise is seen to corrupt field realizations heavily. Filtered DNS (FDNS) data obtained by Fourier cut-off filtering of vorticity data obtained from DNS are also shown for the purpose of comparison. As discussed previously, the stabilization behavior is observed for both $Re = 32000$ and $Re = 64000$ data. We may thus conclude that the learned model has established an implicit sub-grid 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 sub-grid models and studies are underway to increase complexity in the mapping as well as input space.

4.6 A-priori and a-posteriori dichotomy

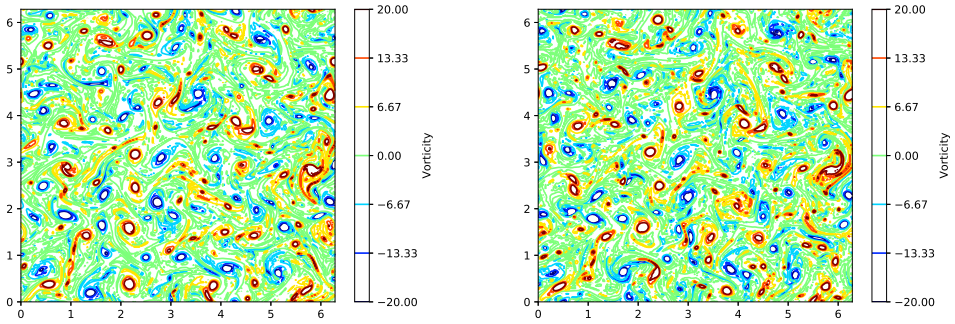
In the previous sections, we have outlined the performance of our proposed framework according to the optimal model architecture chosen by a grid-search for the number of hidden layers as well as the number of hidden-layer neurons. This a-priori hyper-parameter selection is primarily devised on mean-squared-error minimization and is susceptible to providing model architectures which are less resistant to over-fitting and more prone to extrapolation. Our experience shows that an a-posteriori predic-



(a) ANN SGS - $Re = 32000$ (b) ANN SGS - $Re = 64000$



(c) UNS - $Re = 32000$ (d) UNS - $Re = 64000$



(e) FDNS - $Re = 32000$ (f) FDNS - $Re = 64000$

Figure 4.10: A-posteriori results for the proposed framework showing vorticity fields for $Re = 32000$ and $Re = 64000$ data using coarse-grained grids (top). We also provide no-model simulations (middle) and filtered DNS contours (bottom) for the purpose of comparison.

tion (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.

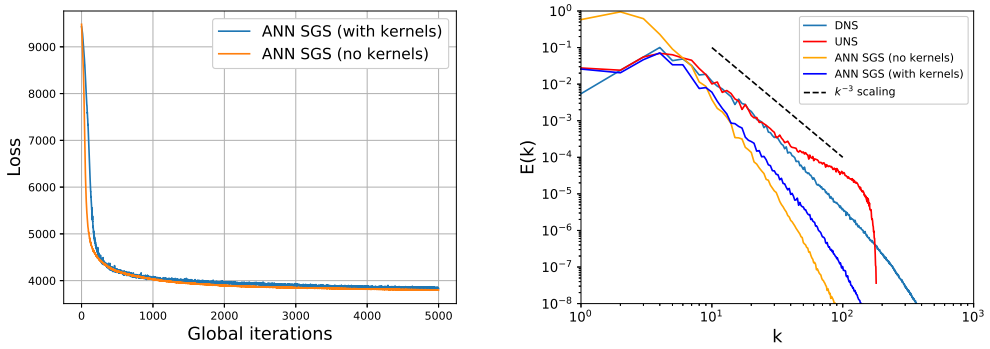
4.6.1 Effect of eddy-viscosity inputs

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 9-point stencils for vorticity and streamfunction as shown in the mathematical expression given by

$$\begin{aligned} \mathbb{M} : \{\bar{\omega}_{i,j}, \bar{\omega}_{i,j+1}, \bar{\omega}_{i,j-1}, \ldots, \bar{\omega}_{i-1,j-1}, \\ \bar{\psi}_{i,j}, \bar{\psi}_{i,j+1}, \bar{\psi}_{i,j-1}, \ldots, \bar{\psi}_{i-1,j-1} \in \mathbb{R}^{18} \rightarrow \{\tilde{\Pi}_{i,j}\} \in \mathbb{R}^{1}. \end{aligned} \tag{4.13}$$

As shown in Figure 4.11, 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 sub-grid quantities from the 18-dimensional input space displayed an unconstrained behavior at the larger scales with the formation of non-physical large scale structures (also shown in Figure 4.8). 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 5-layer, 50 neuron ANN (one that was deemed too complex by our grid-search) with results shown in Figure 4.12. 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 that 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



(a) Learning rate

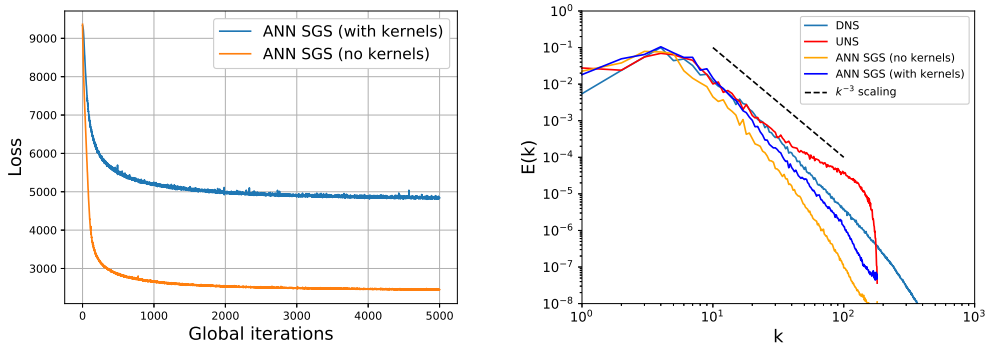
(b) A-posteriori deployment

Figure 4.11: A-priori (left) and a-posteriori (right) effect of the utilization of eddy-viscosity kernel inputs in training and deployment for a two-layer 50 neuron network with a 9-point stencil. The presence of these kernels (intangible in a-priori error minimization) leads to constrained statistical fidelity in a-posteriori deployment at $Re = 32000$.

observed with the deeper network with the use of the kernels.

4.6.2 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 Equation 4.9) and the presence of numerical errors during deployment may often necessitate architectures which 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 4.13. This 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 4.12 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



(a) Learning rate (b) A-posteriori deployment

Figure 4.12: A-priori (left) and a-posteriori (right) effect of the utilization of eddy-vicsosity kernel inputs in training and deployment for a five-layer 50 neuron network with a 9-point stencil. The presence of these kernels leads to higher training errors but viable statistical fidelity in a-posteriori deployment at $Re = 32000$.

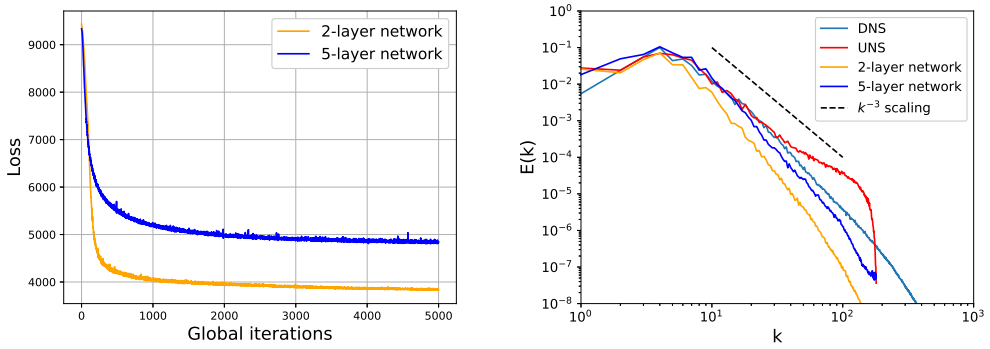
Leith eddy-viscosities in their input space.

4.6.3 Stencil selection

Another comparison is made when the input dimension is substantially reduced by choosing a 5 point stencil (instead of the aforementioned 9 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 stream-function). 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

$$\begin{aligned} \mathbb{M} : \{&\bar{\omega}_{i,j}, \bar{\omega}_{i,j+1}, \bar{\omega}_{i,j-1}, \bar{\omega}_{i+1,j}, \bar{\omega}_{i-1,j} \\ &\bar{\psi}_{i,j}, \bar{\psi}_{i,j+1}, \bar{\psi}_{i,j-1}, \bar{\psi}_{i+1,j}, \bar{\psi}_{i-1,j}, |\bar{S}|_{i,j}, |\nabla\bar{\omega}|_{i,j}\} \in \mathbb{R}^{12} \rightarrow \{\tilde{\Pi}_{i,j}\} \in \mathbb{R}^{1}. \end{aligned} \tag{4.14}$$

Figure 4.14 shows the performance of this setup in training and deployment where it can once again be observed that a-posteriori analysis is imperative for determining a map for the sub-grid terms. While training errors are more or less similar, the reduced stencil fails to capture the non-linear relationship between the resolved and cut-off

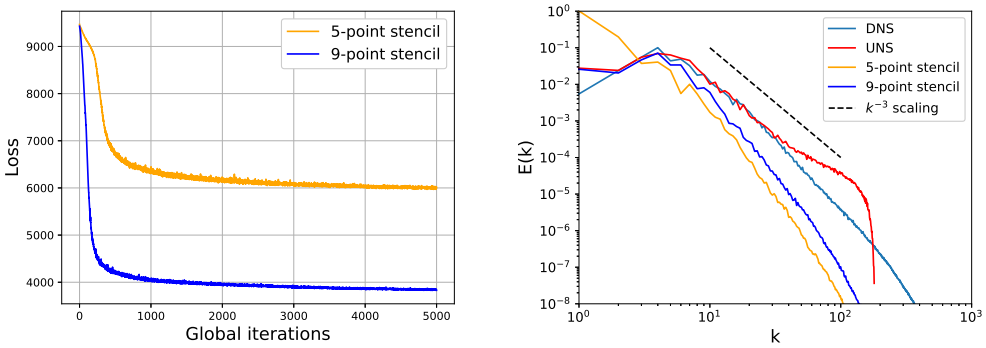


(a) Learning rate (b) A-posteriori deployment

Figure 4.13: A-priori (left) and a-posteriori (right) effect of the number of hidden-layers in the proposed framework. While the two-layered ANN with a 9-point stencil leads to excellent a-priori results, the five-layered network predicts k^{-3} scaling more accurately in deployment for an a-posteriori simulaion at $Re = 32000$.

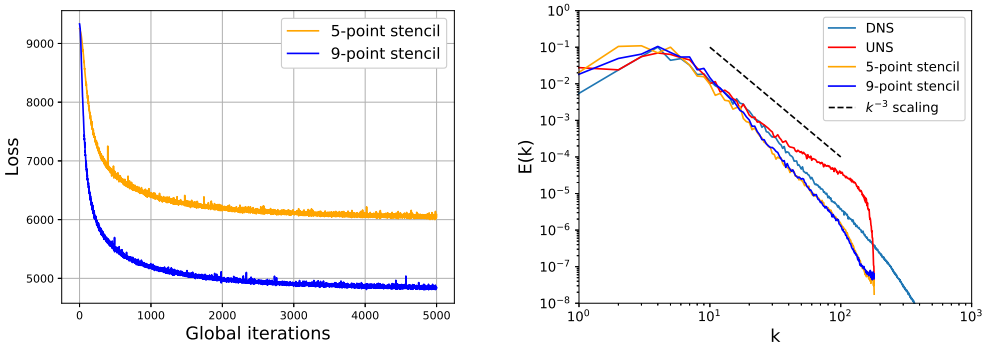
scales with consequent results on the statistical fidelity of the lower wavenumbers. We perform a similar study related to this effect of data-locality on a deeper network given by 5 layers and 50 neurons to verify the effect of the deeper architecture on constrained prediction. The results of this training and deployment are shown in Figure 4.15 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 take-away from this section thus becomes the fact that optimal architectures and maps for sub-grid 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.



(a) Learning rate (b) A-posteriori deployment

Figure 4.14: A-priori (left) and a-posteriori (right) effect of the stencil size in the 2-layer, 50 neuron framework for a $Re = 32000$ simulation. While the 5-point stencil leads to similar a-priori training errors, an a-posteriori deployment at $Re = 32000$ reveals its limitations.



(a) Learning rate (b) A-posteriori deployment

Figure 4.15: A-priori (left) and a-posteriori (right) effect of the stencil size in the 5-layer, 50 neuron framework for a $Re = 32000$ simulation. With deeper architectures, the 5 and 9-point stencils show similar statistical performance

4.7 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 sub-sampled 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 con-

clusion 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 toward 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 sub-grid 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 underway 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 in this document, 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 enhancements 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 phenomenon with complex initial and boundary conditions. The results in this document indicate a promising first step in that direction.

4.8 Appendix

4.8.1 Hyper-parameter optimization

In this appendix, we detail the process of a-priori architecture selection before training and deployment. Our hidden layers have neurons which are activated by the rectified-linear (ReLU) function. The choice of the ReLU activation was made for efficient optimization of the network architecture by bypassing the problems of vanishing gradients inherent in sigmoidal activation functions (Ling, Kurzawski and Templeton, 2016).

For the purpose of optimal network architecture selection, we utilize a grid-search selection coupled with a 3-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 1 to 8) is optimized three times using a 3-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 4.16. We note that the number of neurons in this first grid-search is fixed at 50 although similar trends are recovered with varying specifications between 10 and a 100. Our mean cost index is given by the following expression for each location on the grid

$$\text{Mean cost index} = \frac{1}{K}\sum_{i=1}^{K}\left|\left|\Pi_K^{true} - \tilde{\Pi}_K\right|\right|_2 \tag{4.15}$$

where K refers to the training fold chosen for gradient calculation in the backpropagation 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 4.16 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 multi-dimensional 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 4.17. Details of our network optimization and dataset generation are provided in the next section.

4.8.2 Network training

For the purpose of generating an optimal map discussed in the previous section, we utilize a supervised learning with sets of labeled inputs and outputs obtained from direct numerical simulation data (DNS) for two-dimensional turbulence (San and Staples, 2012; Maulik and San, 2017*c*). Our grid-resolved variables (which we remind the reader, are denoted as overbarred quantities) are generated by a Fourier cut-off 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 on the fact that, while the DNS data generated multiple time snapshots of flow evolution, data was harvested from times $t = 0, 1, 2, 3$ and 4 for the purpose of training and validation. This represents a stringent sub-sampling

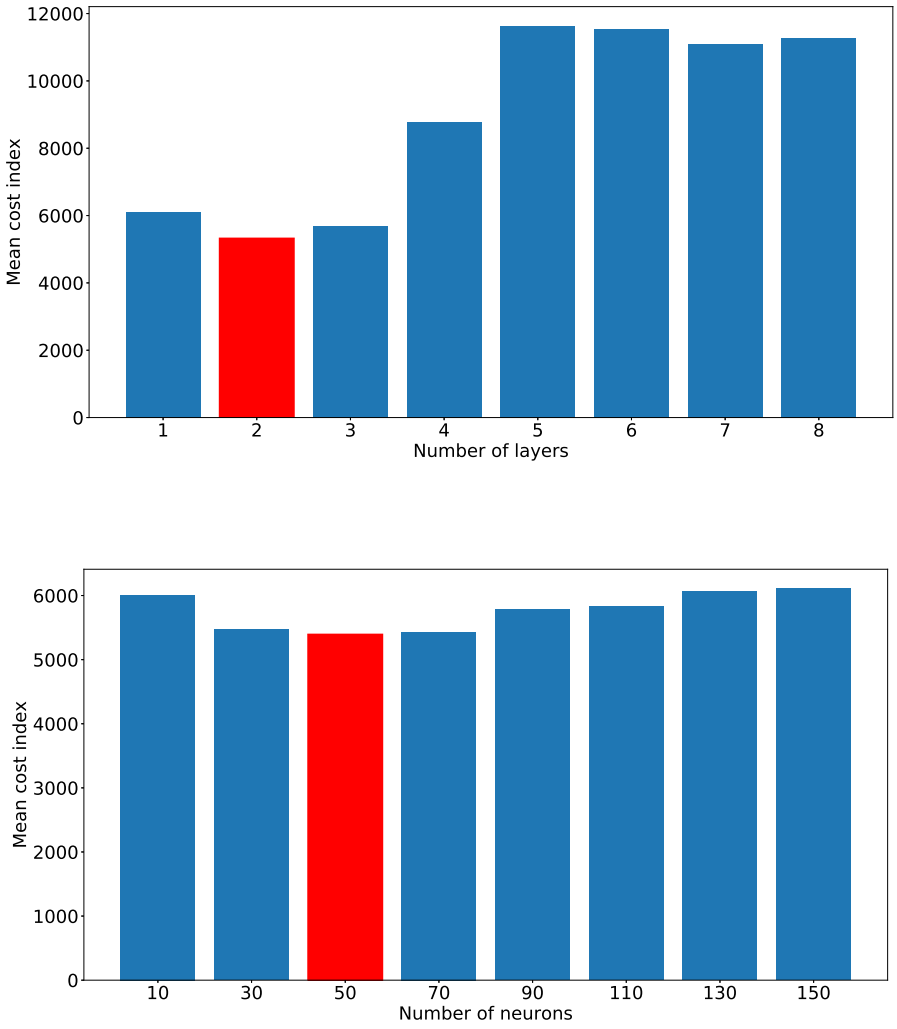


Figure 4.16: Quantification of hyper-parameter optimization shown for number of layers (top) and number of neurons (bottom). An optimal network architecture of two-layers and 50 neurons is chosen for our study.

of the total available data for map optimization. To quantify this sub-sampling, we note that we had potential access to 40000 space-time snapshots of DNS data out of which only 5 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 data sets is given by $Re = 32000$ 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 backpropagation for parameter update. Validation was utilized to record the performance of the trained network on data it was not exposed to during training. Similar behavior 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 and Ba, 2014) implemented in the open-source ANN training platform TensorFlow. Figure 4.17 shows the learning rate of the proposed framework with very similar behavior between training and validation loss implying a successfully optimized map. We remark that while the network may have learned the map from the data it has been provided for training and validation, testing would require an a-posteriori examination as detailed in the following section.

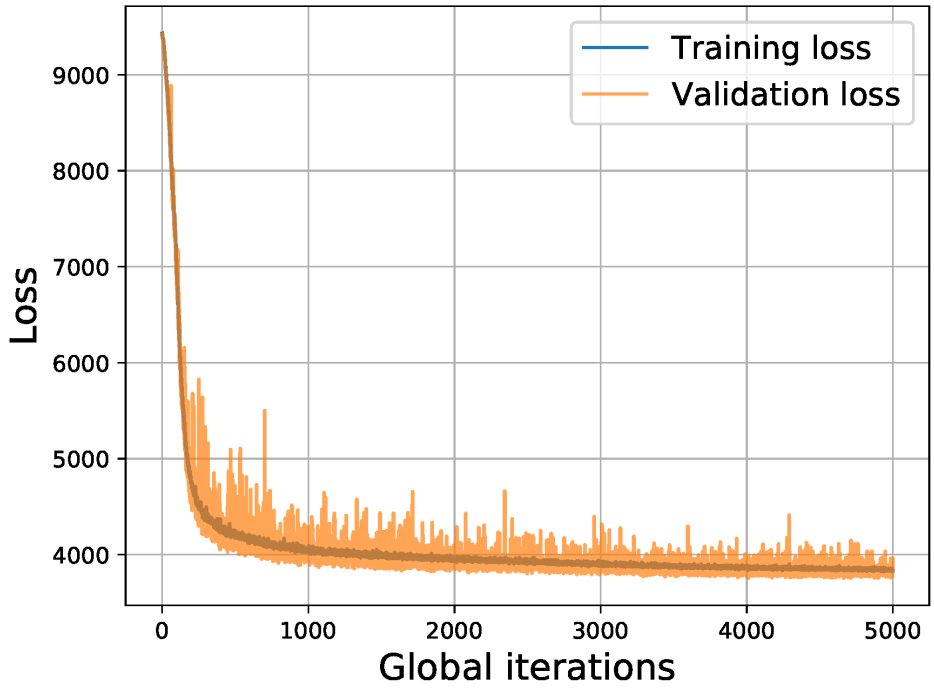


Figure 4.17: Learning rate of the proposed optimal model architecture. Note how training and validation loss are correlated closely for this learning problem.

CHAPTER V

Eddy-viscosity predictions through the machine learning of sub-grid stresses for the large-eddy-simulation of turbulence

5.1 Introduction

In recent times, there has been a widespread interest in the potential of data-driven techniques for potentially improving turbulence modeling in Reynolds-averaged Navier-Stokes and large eddy simulations (LES) (Duraisamy et al., 2019). Most studies are oriented around the specification of a turbulence closure by leveraging information from high-fidelity experimental or computational data sets. Some notable examples of closure modeling using machine learning techniques include (King et al., 2016) where *a priori* optimization was utilized to minimize the L^2-error between true and modeled sub-grid quantities in a least-squares sense using a parameter-free map utilizing a Volterra series. The study in (Vollant et al., 2017) has utilized artificial neural networks (ANN) for computing sub-grid quantities from filtered direct numerical simulation (DNS) data (in a manner similar to that described in the following article). In the study described in (Beck et al., 2018), a variety of neural network architectures such as convolutional and recurrent neural networks are studied for predicting closure terms for decaying homogeneous isotropic turbulence.

In this note, we outline preliminary results for a data-driven turbulence model devised to predict closure terms for the large eddy simulation (LES) of decaying

This chapter is adapted from Maulik et al., under review for publication as a short note at AIAA J.

isotropic turbulence. Our data-driven framework is given by a multi-layered artificial neural network devised to predict outputs according to a localized query of grid-resolved quantities. Our outputs are given by the components of the sub-grid stress tensor from which an isotropic data-driven eddy-viscosity is calculated using a least-squares averaging and truncation for stability (the latter inspired by Beck et al. (2018)). Training data for our ANN is generated by employing a spectrally-sharp filter on Taylor-Green vortex (TGV) DNS data at $Re = 1600$. Our model is assessed by a deployment at Reynolds number (Re) 1600 and 5000, the latter case proving as a test for generalizability. Note, however, that the use of spectrally sharp filtering requires a-posteriori postprocessing which we perform through a least-squares projection and truncation to obtain an averaged eddy-viscosity model.

The proposed formulation is therefore aligned with the philosophy of optimal LES which attempts to predict defects from direct numerical simulation (DNS) data (Moser et al., 2009) and results in a spatio-temporally dynamic closure which predicts the spatially-averaged kinetic energy spectra and energy-dissipation rates with good accuracy.

5.2 Data-driven closure modeling

As mentioned previously, our machine learning philosophy of choice is a deep artificial neural network that estimates a map between our flow resolved variables and the sub-grid tensor. Our framework, also known as a multilayered perceptron, consists of a set of linear and non-linear mathematical transformations of a vector in input space to a result in output space. The purpose of training this network lies in estimating the nature of these transformations in an offline fashion for posterior deployment.

As such these transformations manifest themselves in several stages (or layers) of matrix operations. Each layer is an intermediate vector in this multi-step transformation and is subjected to another local transformation called a biasing and activation

before the next matrix operation is performed. Biases are a constant vector for each substage or layer which are added to the incident vector on its way to a transformed output. In this study, we utilize a rectified linear activation for imparting non-linearity to the transformation between inputs and outputs. Note that each component of an intermediate vector corresponds to a unit cell called a neuron. At the final (or output) layer, a biasing and activation leads to the desired output vector which may serve the purpose of classification or regression. The learning in this study is *supervised* implying labeled benchmark output data for the purpose of characterizing model accuracy. Mathematically, if our input vector $\mathbf{p}$ resides in a P dimensional space and our desired output $\mathbf{q}$ resides in a Q dimensional space, our DNN is used to establish a map $\mathbb{M}$ as follows:

$$\mathbb{M} : \{p_1, p_2, \ldots, p_P\} \in \mathbb{R}^P \rightarrow \{q_1, q_2, \ldots, q_Q\} \in \mathbb{R}^Q. \tag{5.1}$$

This map $\mathbb{M}$ is stored in the form of weight matrices and bias vectors which operate on the input vector. Note that it is generally accepted that multilayered perceptrons may be considered universal function approximators for appropriately chosen activation functions (Hornik et al., 1989). The map for our proposed model may thus be summarized as

$$\begin{aligned} \mathbb{M}_1 : \{&\bar{u}_{i,j,k}, \bar{u}_{i+1,j,k}, \ldots, \bar{u}_{i-1,j-1,k-1}, \bar{v}_{i,j,k}, \bar{v}_{i+1,j,k}, \ldots, \bar{v}_{i-1,j-1,k-1}, \\ &\bar{w}_{i,j,k}, \bar{w}_{i+1,j,k}, \ldots, \bar{w}_{i-1,j-1,k-1}\} \in \mathbb{R}^{81} \rightarrow \{\tau_{11}^{ANN}, \tau_{12}^{ANN}, \ldots, \tau_{33}^{ANN}\} \in \mathbb{R}^6, \end{aligned} \tag{5.2}$$

where $(\overline{u}, \overline{v}, \overline{w})$ are the coarse-grained velocity components in three-dimensions, i, j, k are spatial indices on a discrete grid and the components of $\boldsymbol{\tau}^{ANN}$ are a data-driven approximation to the true sub-grid scale tensor given by

$$\tau_{ij}^{sgs} = \bar{u}_i \bar{u}_j - \overline{u_i u_j}. \tag{5.3}$$

and where $\overline{a}$ is a coarse-grained projection of a field variable a and i, j now correspond to tensor indices. We note that our training and testing data (for map optimization and assessment) is generated by assuming that the coarse-grained projection for any variable may be obtained by a spectral cut-off filter and thus precludes errors due to discretization based filtering. One may also utilize a Gaussian type filtering for generating this data. We note that the proposed mapping corresponds to the utilization of a box stencil (i.e. the 27 nearest neighbor grid points in three dimensions) around the point of prediction with each grid point contributing three quantities of interest given by their flow components for a total of 81 inputs.

In terms of the data utilized for network optimization, we utilize primitive variables from DNS snapshots of the TGV problem at $t = 5$ and $t = 10$ representing a transitional and fully turbulent realization of the unsteady physics of this test-case. This data set was generated from a full-order forward simulation of the TGV problem at $Re = 1600$ and at $N^3 = 512^3$ and the coarse-grained quantities were generated by a degree-of-freedom reduction in the wavenumber domain using a spectrally-sharp cut-off filter. This lets us compute coarse-grid resolved quantities such as $\overline{\mathbf{u}}$ and the sub-grid tensor τ at the reduced resolution of $N^3 = 64^3$. Our snapshots at time $t = 5$ and $t = 10$ display a Kolmogorov length scale of $\eta \approx 0.0619$ and $\eta \approx 0.048$ respectively. A coarse grained resolution of $N^3 = 64^3$ corresponds to a grid size of approximately 0.1 which shows that a sub-grid closure is necessary. The Taylor microscale based Reynolds number for these snapshots correspond to $Re_\lambda \approx 194.93$ and $Re_\lambda \approx 67.40$ respectively as computed from the DNS.

The ANN is trained for optimal predictions by monitoring mean-squared-error values of training and validation data sets for total of 500 epochs (or total sweeps of the data) after which assessments of performance are carried out. Our network architecture is specified by five hidden layers and 100 neurons in each hidden layer. The choice of this architecture was through investigation of optimal combinations to

get Pearson correlation coefficient values of $R^2 > 0.75$ for the training loss (Gamahara and Hattori, 2017).

5.3 Stable projection of data-driven predictions

Conventional turbulence models utilize algebraic or differential equation based specifications for an eddy-viscosity (ν_e) where we shall attempt to approximate these contributions through our trained ANN during a-posteriori simulations. However, these predictions require some postprocessing for stability due to the inherent stochastic nature of the ANN as well as the assumptions on filter specification during a-priori LES data generation from DNS. In addition, a-posteriori deployments must be robust to numerical errors during flow evolution for which it is necessary to project our predictions to a stable space. We project the outputs of our ANN to a stable space for forward deployments via the combination of a truncated least-squares ideology as well as the essence of the eddy-viscosity hypothesis. This is described in the following paragraph. Note that the nearly imcompressible case considered here allows us to fix our turbulent Prandtl number estimation for the sub-grid heat flux. We introduce the following notation to simplify our future derivations with

$$\Pi_{ij} = \frac{\partial \bar{u}_i}{\partial x_j} + \frac{\partial \bar{u}_j}{\partial x_i} - \frac{2}{3}\frac{\partial \bar{u}_k}{\partial x_k}\delta_{ij} \tag{5.4}$$

being linear terms based on the grid-resolved strain rate and where δ_{ij} is the Kronecker delta. Also, the grid-resolved terms utilized to compute Π_{ij} are individially calculated at cell faces of our finite volume formulation. We now introduce a truncated-least-squares procedure to obtain a stable data-driven eddy viscosity which is predicted by our trained network by way of approximations to τ_{mn}. We first formulate a least-

squares error minimization problem given by the following objective function

$$J(\nu_e) = \sum_{i=1}^{3}\sum_{j=1}^{3}(\tau_{ij}^{ANN} - \nu_e \Pi_{ij})^2. \tag{5.5}$$

where there are only six summation terms due to the symmetry of τ_{ij}^{ANN} and Π_{ij}. A least-squares minimization based calculation of eddy-viscosity gives us

$$\nu_e = \sum_{i=1}^{3}\sum_{j=1}^{3}\frac{\tau_{ij}^{ANN}\Pi_{ij}}{\Pi_{ij}\Pi_{ij}}. \tag{5.6}$$

This predicted value of an effective eddy viscosity ν_e is truncated between $[0, \frac{10}{Re}]$ in a manner similar to that proposed in (Beck et al., 2018) for viscous stability of the proposed framework and to ensure consistency with the second-law of thermodynamics. We also note that a purely positive eddy-viscosity ensures time irreversibility which is a desired characteristic of linear eddy-viscosity models Carati et al. (2001). The culmination of this framework to an averaged linear eddy-viscosity model also implies that time, pressure and generalized Galilean invariance is preserved. However, the data-driven calculation of the eddy-viscosity leads to a loss in reflectional and rotational invariance Silvis et al. (2017). This deployment provides stable and accurate predictions for the sub-grid source terms.

5.4 Results

5.4.1 A-priori testing

In this sub-section, we assess the statistical fidelity of a-priori predictions from our trained ANN for snapshots outside the training data set. Statistical assessments are represented in the form of probability density functions (PDFs) of the six true and predicted components (the latter denoted ML) of the sub-grid tensor (i.e., $\boldsymbol{\tau}^{sgs}$ and $\boldsymbol{\tau}^{ANN}$ respectively). Since our tensor considers the deviatoric contribution of

the sub-grid term, all PDFs are centered at a mean value of zero. In general, it is observed that the predictions are sufficiently well correlated with the truth as shown in Fig. 5.1 where the trained map is deployed for TGV DNS data at $Re = 5000$ and $t = 15$ which is utilized as a test data set for out of training range prediction assessment. We remind the reader that the snapshot data used for training was for TGV DNS data at $Re = 1600$ and at $t = 5$ and $t = 10$. For the purpose of comparison we also display the performance of the Dynamic Smagorinsky (DS) closure in this a-priori assessment. In terms of L2-norm comparisons, the machine learning framework provided errors of [10.38, 7.06, 6.82, 10.44, 6.77, 10.10] whereas the dynamic Smagorinsky technique provided [10.77, 7.75, 7.23, 10.77, 7.23, 10.32] for the tensor components $[\tau_{11}, \tau_{12}, \tau_{13}, \tau_{22}, \tau_{23}, \tau_{33}]$.

5.4.2 A-posteriori testing

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 and San, 2017*a*), the results of the following section are promising in that they represent a sub-grid model computation in temporally and spatially dynamic fashion. This test setup 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 addition, we also track the averaged kinetic energy dissipation rate for our forward deployments (a classical measure of the quality of a TGV computation). All our assessments are compared with no-model simulations and the dynamic Smagorinsky model (Germano et al., 1991) for the TGV test-case which is a prototype problem which models the transition to and decay of isotropic, homogeneous and turbulent periodic vortical

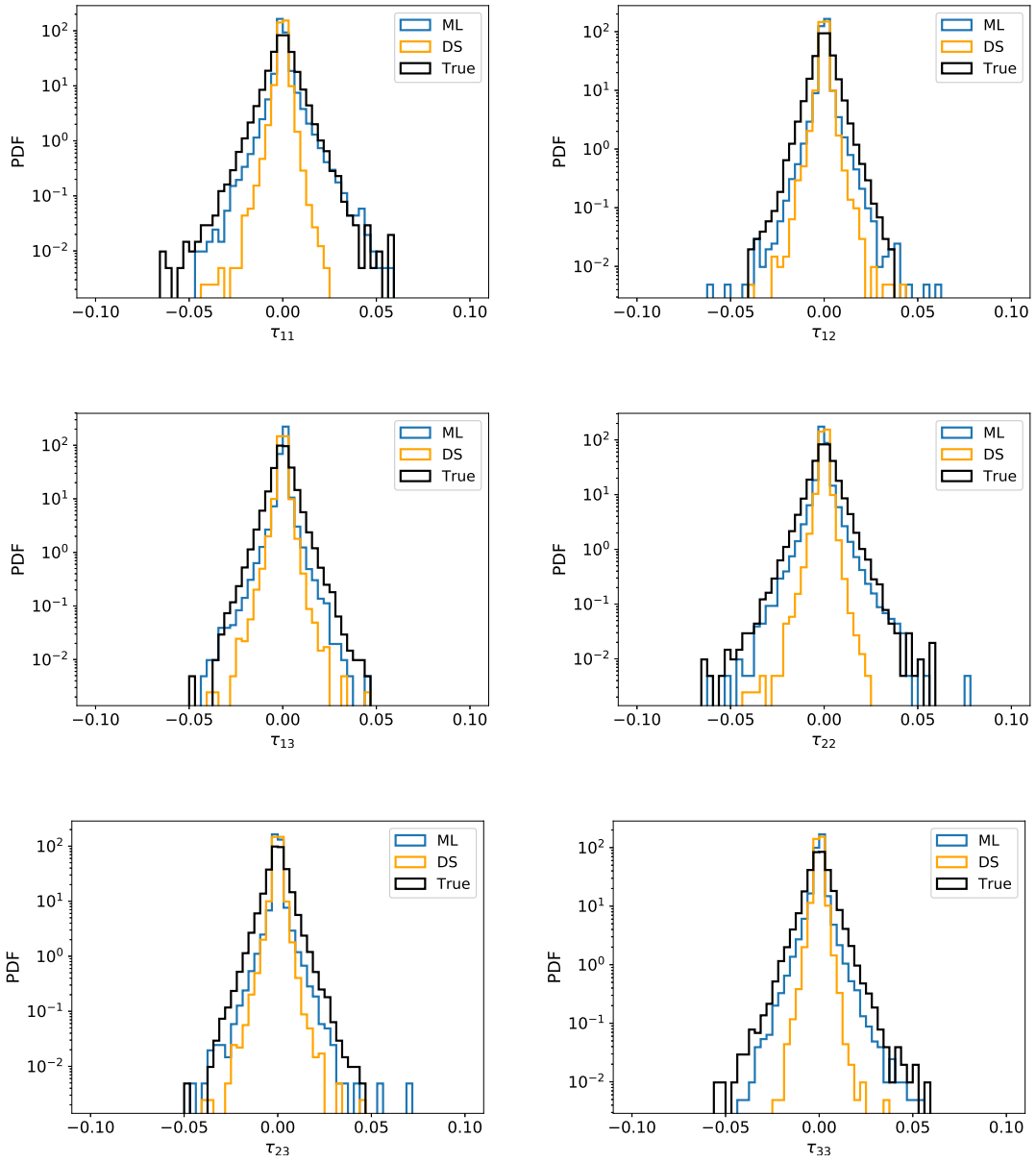


Figure 5.1: A-priori statistical recovery of sub-grid stress distributions at $t = 15$ for $Re = 5000$. Note that this data is 'out-of-range' since our sampling was done for $t = 5$ and $t = 10$ alone at a lower Reynolds number of 1600.

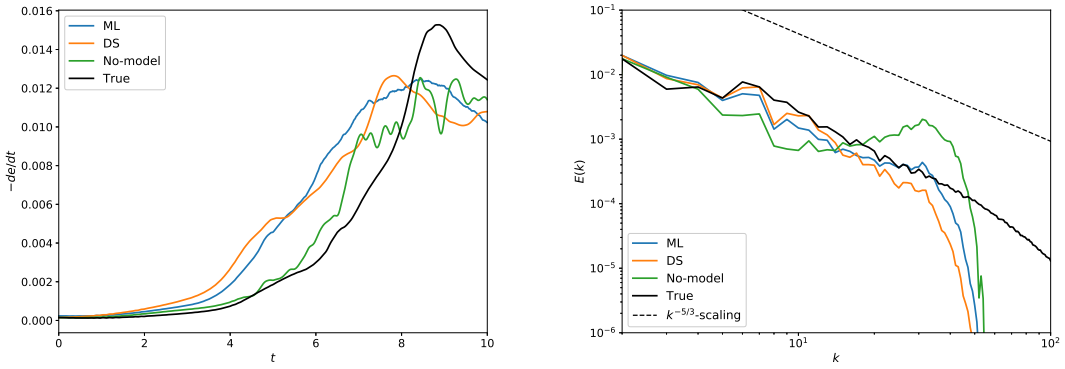


Figure 5.2: A-posteriori testing of our proposed framework (ML) with energy dissipation rate predictions (left) and spatially averaged kinetic energy spectra (right) for $Re = 5000$ at $N^3 = 64^3$ degrees of freedom. Note that the network was optimized for predictions at $Re = 1600$ alone. The dynamic Smagorinsky (DS) model is also shown for comparison.

flows. Details related to the numerics of these assessments may be found in Rahman et al. (2018).

Figure (5.2) shows an a-posteriori deployment of the proposed idea for a $Re = 5000$ simulation of the TGV test case at $N^3 = 64^3$ degrees of freedom. We have also shown results from our high fidelity DNS runs (at $N^3 = 512^3$ degrees of freedom) as well as the performance of the dynamic Smagorinsky model. One can observe that the proposed framework is capable of recovering the energy dissipation rate trends of the dynamic Smagorinsky approach. In addition it is also seen that the proposed model captures the $k^{-5/3}$ scaling expected in the TGV test case adequately in comparison to the no-model simulation which shows some high wavenumber aliasing error. We remind the reader that the proposed approach utilizes no test-filtering strategy for calculation of a dynamic Smagorinsky coefficient and calculates an equation-free dynamic eddy viscosity spatio-temporally. We note that the filter to Kolmogorov length scale ratio for this unsteady problem peaks at approximately 4.0 and that the network is not exposed to $Re = 5000$ data during map optimization.

Figure (5.3) shows the deployment of our trained ANN and least-squares pro-

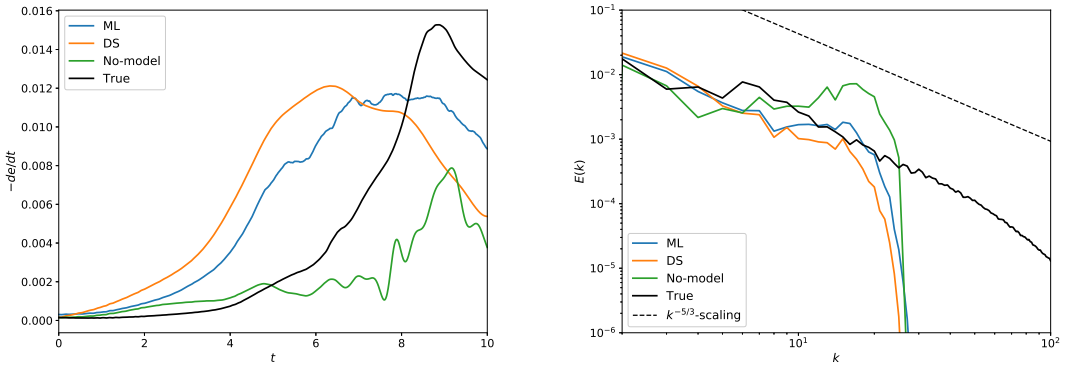


Figure 5.3: A-posteriori testing of our proposed framework with energy dissipation rate predictions (left) and spatially averaged kinetic energy spectra (right) for $Re = 5000$. Note that the network was optimized for predictions at $N^3 = 64^3$ degrees of freedom while it is deployed here for $N^3 = 32^3$. The dynamic Smagorinsky model is also shown for comparison.

jection method for a further coarsened grid of $N^3 = 32^3$ where it is seen that the predictive nature of the framework is superior to the no-model simulation by way of assessing the energy spectra. In addition, the proposed framework also provides more accurate estimates for the energy dissipation rate and the kinetic energy spectra than the DS approach which proves to be slightly more dissipative. However, it is seen to be more stable than the no-model results which show considerable noise accumulation at grid cut-off wavenumbers. The filter to Kolmogorov length scale ratio for this test-case peaks at approximately 8.0.

5.5 Conclusions

The primary focus of this investigation is to develop a data-driven strategy to learn and stabilize the non-linear sub-grid scale source terms required for closing the coarse-grained Navier-Stokes equations to be utilized in an LES. Our framework relies on the use of an ANN to map the highly non-linear relationship between the inputs, given by the resolved flow variables and the outputs given by the sub-grid stresses. Our supervised learning procedure requires labeled data which is generated in an a-priori

fashion through low-pass filtering DNS snapshots with a spectrally sharp cut-off filter to reduce degrees of freedom. Our framework is trained, by way of a mean-squared-error loss function on just two snapshots of the artificially generated coarse-grained data obtained from the full-order simulation fields for the Taylor-Green vortex test case at $t = 5$ and $t = 10$. Our framework is thus able to recover a spatio-temporally dynamic closure for decaying isotropic turbulence.

Our model is tested in both a-priori and a-posteriori assessments for grid resolutions and Re values which it has not been exposed to during learning. This validates the generalizability and robustness of the learning for unsteady deployment. Stability in a-posteriori is ensured through an intermediate truncated least-squares-projection to predict an effective eddy-viscosity. For this purpose we follow the recommendations outlined in (Beck et al., 2018) which ensures stability of the inherently stochastic nature of the predictions. In particular, our a-posteriori deployment is performed for a filter length scale that is up to 8 times that of the Kolmogorov length scale. It is observed that the framework recovers a performance similar to the classical dynamic Smagorinsky turbulence model. This is a promising development, since the proposed framework requires no test-filtering strategy and no functional form for the eddy-viscosity.

Our studies in this article have shown that ANNs may be utilized for parameter free turbulence models through the specification of appropriate constraining strategies in a-posteriori. Although a-priori optimal architecture selection in the form of optimal layers or neurons remains an open question, we are provided with promising results from the point of view of a long term goal for adaptive closures that generalize over various sub-grid physics. The future development of the ideas in this note rely on the inclusion of reflectional and rotational invariance properties and symmetry considerations into sub-grid predictions which may potentially obviate the need for truncation (for instance through tensor-basis networks as shown in Ling, Kurzawski

and Templeton (2016)). Our goal is to utilize the learnings from this preliminary study as a benchmark for further improvement in the accuracy and understanding of machine learning based sub-grid modeling of turbulence.

CHAPTER VI

Sub-grid scale model classification and blending through deep learning

6.1 Abstract

In this article we detail the use of machine learning for spatiotemporally dynamic turbulence model classification and hybridization for the large eddy simulations (LES) of turbulence. Our predictive framework is devised around the determination of local conditional probabilities for turbulence models that have varying underlying hypotheses. As a first deployment of this learning, we classify a point on our computational grid as that which requires the functional hypothesis, the structural hypothesis or no modeling at all. This ensures that the appropriate model is specified from a-priori knowledge and an efficient balance of model characteristics is obtained in a particular flow computation. In addition, we also utilize the conditional probability predictions of the same machine learning to blend turbulence models for another hybrid closure. Our test-case for the demonstration of this concept is given by Kraichnan turbulence which exhibits a strong interplay of enstrophy and energy cascades in the wave number domain. Our results indicate that the proposed methods lead to robust and stable closure and may potentially be used to combine the strengths of various models for complex flow phenomena prediction.

This chapter is adapted from Maulik et al., under review for publication in J. Fluid Mech.

6.2 Introduction

Turbulence is an active area of research due to its significant impact on a diverse set of challenges such as those pertaining to the aerospace and geophysical communities. In recent decades, computational fluid dynamics (CFD) has proven to be useful for low-cost realizations of flow phenomena for critical decision making processes. However, CFD is still fairly limited in terms of accuracy due to the exceptional computational expense involved in high-fidelity simulations of turbulence. 'True' numerical experiments require the use of a direct numerical simulation (DNS) of the Navier-Stokes equations. DNS is only possible if a discretized domain can resolve all possible frequencies in a flow and is therefore out of reach of the vast majority of engineering and geophysical applications for the foreseeable future. Large eddy simulations (LES) (Sagaut, 2006) have proven to be a promising strategy for resolving a greater number of scales in a flow but require the specification of a model which represents the interactions of the higher frequencies with the mean flow. This sub-grid scale (SGS) model, also known as a closure, is usually specified in the form of an algebraic or differential equation and is generally flow category specific (Vreman, 2004). This imposes a caveat on the applicability of a SGS model if no a-priori information of the flow category is known. As such, the basic premise of LES is extendable to many partial differential equation systems with quadratic non-linearities. In this paper, we explore the utilization of machine learning for dynamically inferring regions where a particular turbulence modeling hypothesis is applicable with the goal of improving predictive capabilities of turbulence dynamics for a wide range of problems.

The multi-scale nature of turbulence requires the use multiple modeling approximations for the higher wavenumbers which remain unsupported by computational degrees of freedom (a case for most flows of any practical interest). The procedure of modeling these smaller scales is often denoted *closure* due to insufficient knowledge about higher-order wavenumber interactions with the coarse-grained system (Berselli

et al., 2006) and remains vital for the accurate computation of many applications (Hickel et al., 2014; Yu et al., 2016). From an LES point of view, the closure problem may be considered to be dominated by commutative errors in the calculation of the non-linear term as well as the defects associated with commutative errors stemming from the dynamic term. In this study, we focus on the former.

Explicit LES argues for the utilization of sub-grid models specified as algebraic or differential equations for the unresolved scales. These are built on an intuitive reasoning of how the losses of coarse-graining the Navier-Stokes equations may be incorporated into an LES deployment. Some of the most notable sub-grid closure strategies are those given by the linear eddy-viscosity hypothesis, which models the sub-grid stress tensor through the Bousinessq approximation. Within the context of the Navier-Stokes equations, it is generally accepted that the vorticity dominated smaller scales are dissipative (Kolmogorov, 1941*b*) and therefore, most turbulence models seek to specify a sub-grid dissipation (Frisch, 1995). Many functional sub-grid models can be traced back to Smagorinsky (1963), where an effective eddy-viscosity was determined by an a-priori specified mixing length and a $k^{-5/3}$ scaling recovery for the kinetic energy content in the wavenumber domain. Similar hypotheses have also been used for two-dimensional turbulence (often utilized as a test-bed for geophysical scenarios, for instance see Pearson et al. (2017); Pearson and Fox-Kemper (2018)), for approximating the k^{-3} cascade and generally have their roots in dimensional analysis related to the cascade of enstrophy (Leith, 1968). These models may also be classified as *functional* due to the phenomenological nature of their deployment and comprise the bulk of explicit LES turbulence models used in practical deployments. Explicit LES closures may also be specified through the specification of a low-pass spatial filter to account for the unresolved scales (Bardina et al., 1980; Stolz and Adams, 1999; Layton and Lewandowski, 2003; Mathew et al., 2003; San and Vedula, 2018) where phenomenology is bypassed but ansatz are provided for the bulk dissipative nature

of the smaller scales through the control of a characteristic filter-width. In either scenario, (i.e., whether structural or functional), the choice of the phenomenology (or dissipation control parameter) plays a key role in the successful calculation of accurate a-posteriori statistics.

The past few years have seen a rapid increase in the use of machine learning for various scientificand engineering applications. For turbulence, some widely used strategies for prediction and inference include symbolic regression such as in Weatheritt and Sandberg (2016, 2017*a*,*b*), where functional model-forms for Reynolds-averaged Navier-Stokes (RANS) deployments were generated through evolutionary optimization against high-fidelity data. Other techniques incorporating Bayesian ideologies have also been used, for instance in Xiao et al. (2016) where an iterative ensemble Kalman method was used to assimilate prior data for quantifying model form uncertainty in RANS models. In Wang, Wu and Xiao (2017); Wang, Wu, Ling, Iaccarino and Xiao (2017) and Wu et al. (2018*a*), random-forest regressors were utilized for RANS turbulence-modeling given direct numerical simulation (DNS) data. In Singh and Duraisamy (2016) and Singh et al. (2017), an ANN was utilized to predict a non-dimensional correction factor in the Spalart-Allmaras turbulence model through a field-inversion process using experimental data. Bypassing functional formulations of a turbulence model was also studied from the RANS point of view by Tracey et al. (2015). Ling and Templeton (2015) utilized support vector machines, decision trees and random forest regressors for identifying regions of high RANS uncertainty. A deep-learning framework where Reynolds-stresses would be predicted in an invariant subspace was developed by Ling, Kurzawski and Templeton (2016). Machine learning of invariance properties has also been discussed in the context of turbulence modeling by Ling, Jones and Templeton (2016). The reader is directed to a recent review by Duraisamy et al. (2019), for an excellent review of turbulence modeling using data-driven ideas.

As shown above, the use of data-driven ideologies and in particular artificial neural networks (ANNs) has generated significant interest in the turbulence modeling community for addressing long-standing challenges (also see Sotgiu et al. (2018); Zhu et al. (2019); Zhang et al. (2019) for recent examples). A multilayered ANN may be optimally trained to approximate any non-linear function (Hornik et al., 1989) and the large data sets involved in turbulence research couple with ever-improving computing capabilities has also motivated the study of ANN based learning. Within the context of LES (and associated with the scope of this paper) there are several investigations into sub-grid modeling using data-driven techniques. In an early study of the feasibility of using learning from DNS, Sarghini et al. (2003) deployed ANNs for estimating Smagorinsky model-form coefficients within a mixed sub-grid model for a turbulent channel flow. ANNs were also used for wall-modeling by Milano and Koumoutsakos (2002) where it was used to reconstruct the near wall field and compared to standard proper-orthogonal-decomposition techniques. An alternative to ANNs for sub-grid predictions was proposed by King et al. (2016) where a-priori optimization was utilized to minimize the L^2-error between true and modeled sub-grid quantities using a parameter-free Volterra series. Maulik and San (2017*a*) utilized an extreme-learning-machine (a variant of a single-layered ANN) to obtain maps between low-pass spatially filtered and deconvolved variables in an a-priori sense. This had implications for the use of ANNs for turbulence modeling without model-form specification. A more in-depth investigation was recently undertaken by Fukami et al. (2018) where convolutional ANNs were utilized for reconstructing from downsampled snapshots of turbulence. Maulik et al. (2018) also deployed a data-driven convolutional and deconvolutional operation to obtain closure terms for two-dimensional turbulence. Gamahara and Hattori (2017) utilized ANNs for identifying correlations with grid-resolved quantities for an indirect method of model-form identification in turbulent channel flow. The study by Vollant et al. (2017) utilized ANNs in conjuc-

tion with optimal estimator theory to obtain functional forms for sub-grid stresses. In Beck et al. (2018), a variety of neural network architectures such as convolutional and recurrent neural networks are studied for predicting closure terms for decaying homogeneous isotropic turbulence. A least-squares based truncation is specified for stable deployments of their model-free closures. Model-free turbulence closures are also specified by Maulik et al. (2018, 2019) and Wang et al. (2018), where sub-grid scale stresses are learned directly from DNS data and deployed in a-posteriori assessments. King et al. (2018) studied generative-adversarial networks and the LAT-NET Hennigh (2017) for a-priori recovery of statistics such as the intermittency of turbulent fluctuations and spectral scaling.

While a large majority of the LES-based frameworks presented above utilize a least-squares error minimization technique for constructing maps to sub-grid stresses *directly* for theoretically optimal LES (Langford and Moser, 1999; Moser et al., 2009; LaBryer et al., 2015), this work is novel in that it utilizes sub-grid statistics (pre-computed from DNS data) to train a classifier. Our trained intelligence utilizes the most appropriate turbulence modeling modelling hypothesis (i.e., either structural or functional) from a-priori experience to close the LES governing equations. It is also deployed to blend turbulence models linearly at each point during flow evolution for a novel hybrid closure. In this manner, we are able to co-deploy models having fundamentally different underlying hypothesis for turbulence parameterizations in a stable manner. This is similar to the study employed in Ling and Kurzawski (2017) where machine learning is utilized for adaptive error corrections in RANS deployments. In the rest of this article, we discuss the governing equations of decaying Kraichnan turbulence, introduce our machine learning architecture and its optimization and detail its a-priori and a-posteriori performance through statistical assessments.

6.3 Governing equations

We proceed by outlining our Kraichnan turbulence test-case which (alongwith quasi-geostrophic turbulence) is an important prototype for geophysical flow-phenomenon with high aspect ratios and for which turbulence model research remains highly active (Pearson et al., 2017). The governing equations of motion for Kraichnan turbulence are given by the two-dimensional Navier-Stokes equations in a periodic domain. The non-dimensionalized version of these equations may be expressed in the vorticity (ω) and stream function (ψ) formulation as

$$\begin{gathered} \frac{\partial \omega}{\partial t} + J(\omega, \psi) = \frac{1}{Re} \nabla^2 \omega, \\ x, y \in [0, 2\pi], t \in [0, 4], \end{gathered} \tag{6.1}$$

where we define the Jacobian (or the nonlinear term as)

$$J(\omega, \psi) = \frac{\partial \omega}{\partial x} \frac{\partial \psi}{\partial y} - \frac{\partial \omega}{\partial y} \frac{\partial \psi}{\partial x}, \tag{6.2}$$

and the conservation of mass is enforced by

$$\nabla^2 \psi = -\omega. \tag{6.3}$$

A measure of multi-scale behavior in this system is given by the Reynolds number (Re). A high value of Re combined with a coarse-grid projection of these equations results in insufficient support for the finest structures in the flow evolution, leading to noise accumulation at grid cut-off and potential floating point overflow of the numerical evolution of this problem. A sufficiently coarse-grained representation of the

governing equations introduced previously are given by the LES governing equations

$$\begin{aligned} \frac{\partial \bar{\omega}}{\partial t} + J(\bar{\omega}, \bar{\psi}) &= \frac{1}{Re}\nabla^2 \bar{\omega} + \Pi, \\ \nabla^2 \bar{\psi} &= -\bar{\omega}, \end{aligned} \tag{6.4}$$

where Π may be assumed to be the perfect closure given by

$$\Pi = J(\bar{\omega}, \bar{\psi}) - \overline{J(\omega, \psi)}. \tag{6.5}$$

When adequately simulated, the Kraichnan turbulence test cases results in the classical k^{-3} scaling of the energy spectra (Kraichnan, 1967). In practice, this perfect estimation of loss is never available in a numerical deployment and must be estimated by either an algebraic or differential equation. We focus on two competing ideologies for estimating closure. The first is given by the functional hypothesis and may be expressed as

$$\Pi = \nu_e \nabla^2 \bar{\omega} \tag{6.6}$$

where the Smagorinsky approximation to the eddy-viscosity ν_e is given by

$$\begin{aligned} \nu_e &= (C_s \delta)^2 |\bar{S}|, \\ |\bar{S}| &= \sqrt{4(\frac{\partial^2 \bar{\psi}}{\partial x \partial y})^2 + (\frac{\partial^2 \bar{\psi}}{\partial x^2} - \frac{\partial^2 \bar{\psi}}{\partial y^2})^2}. \end{aligned} \tag{6.7}$$

A successful application of this closure necessitates a dynamic calculation of the Smagorinsky coefficient C_s that requires the specification of a test-filter and a spatial-averaging for stabilized deployment. This approach is the well-known dynamic Smagorinsky (DS) closure (Germano et al., 1991) and its two-dimensional abstraction for Kraichnan turbulence has been presented by San (2014).

A competing ideology is given by the structural (or scale-similarity) hypothesis

which assumes that the LES equations are projections of the Navier-Stokes equations to a smoother space where an inverse-filtering operator may be utilized to recover the finer scales that are lost. Mathematically,

$$\Pi = J(\bar{\omega}, \bar{\psi}) - \widetilde{J(\omega^*, \psi^*)} \tag{6.8}$$

where ω^* and ψ^* are approximately deconvolved variables obtained through an inverse-filtering procedure and a Gaussian-type filter kernel (given by the $G(a) = \tilde{a}$) is an approximation of the projection to the LES space. However, these techniques are limited due to the underlying assumption of isomorphism between the LES and the Navier-Stokes equations (Germano, 2015). In practice, this implies that structural hypotheses are appropriate only if finer structures are sufficiently well-resolved on a particular grid. As such, this diminishes their benefit for practical flows where grid cut-off wave numbers are generally much smaller than the largest wave number in the flow. The breakdown of structural closures manifests itself in the form of stability issues. For this reason, many successful closure deployments utilize linear combinations of structural and functional models (Habisreutinger et al., 2007). In this work, we implement approximate-deconvolution (AD) (Stolz and Adams, 1999) which utilizes an iterative application of the trapezoidal filter kernel for inversion of filtered grid-quantities and utilize three iterative resubstitutions to deconvolve our grid-resolved variables.

6.4 Machine learning

We now turn to procedure of utilizing DNS data for learning when to switch between one of three closure scenarios. Of these three options, two are given by the choice of the functional hypothesis and AD. The third option is that of a no-model scenario where our learning determines that closure-modeling is unnecessary. The third sce-

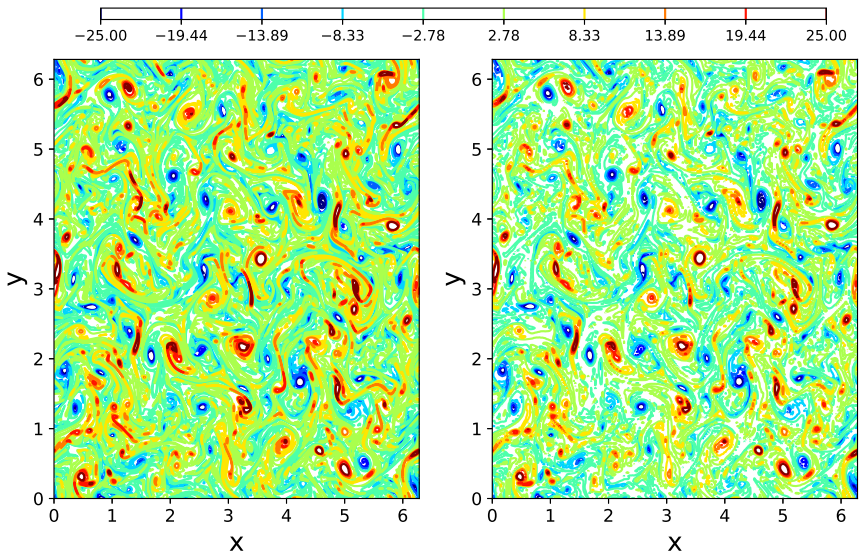


Figure 6.1: Visualization of the effect of Fourier cut-off filtering with DNS (left) and corresponding FDNS (right).

nario is retained since there is a possibility for localized areas in a flow resolution to have adequate grid-support so that the contributions of the sub-grid scale become negligible. This switching between scenarios is spatio-temporally dynamic. Before proceeding, we note that the functional deployment eschews the dynamic procedure and simply sets a $C_s = 1.0$ for the calculation of the eddy-viscosity as utilized in the standard Smagorinsky model (Smagorinsky, 1963) as is common in geophysical scenarios (Cushman-Roisin and Beckers, 2011). This ensures that the proposed framework remains computationally tractable, easily interpretable and deployable. We proceed by outlining our learning strategy through the utilization of DNS data. Five snapshots of DNS data at $Re = 32000$ and at 2048^2 degrees of freedom (from 40000 available snapshots) are utilized to compute the grid-filtered variables (herein denoted by FDNS) at 256^2 degrees of freedom through the application of a spectral cut-off filter. Perfect closure values (Π) are then obtained (the reader is directed to (Maulik et al., 2019) for details related to the calculation of these quantities). Figure 6.1 visually quantifies the effect of the spectral domain filtering where the FDNS of a snapshot of vorticity is shown. We then introduce the a-priori eddy-viscosity given

by

$$\nu_e^a = \frac{\Pi}{\nabla^2 \bar{\omega}} \tag{6.9}$$

where all the terms on the right-hand side of the above equation are available through calculation from the DNS snapshots. The a-priori eddy viscosity is centered at a value of zero (corresponding to a region where closure modeling is unnecessary) and has tails in the negative and positive directions (a hallmark of isotropic turbulence). A core component of the hypothesis in this work stems from the fact that structural hypotheses are not limited to positive eddy-viscosity predictions alone. The reader may note that models utilizing the functional hypothesis always lead to positive eddy-viscosities. The a-priori eddy viscosities calculated from the DNS data are then projected onto a Gaussian distribution where values lying at a distance of 1% of the standard-deviation from the mean (of zero) are labeled as those requiring no closure (due to the low strength of the a-priori eddy-viscosity). Values lying beyond this range are labeled as functional or structural, depending on if they are positive or negative, respectively. This information is encoded in one-hot labeling for a classification deployment and a corresponding schematic for this hypothesis segregation is shown in Figure 6.2. It is observed that the a large portion of the available data lies within the first standard deviation of the mean eddy-viscosity. This leads to the potential of turbulence modeling classification being considered from outlier detection point-of-view. A factor which motivates the choice of the Gaussian distribution is the nature of Kraichnan turbulence (which is isotropic in nature with Gaussian statistics). However, we note that machine learning algorithms are also capable of classifying data belonging to complex distributions and that this hypothesis segregation may be tuned for better accuracy. Also, the choice of the 1% hyper-parameter is also motivated by observing a-posteriori training accuracies where it is noticed that a relatively simple architecture (mentioned next) is efficiently able to discern

the varying hypothesis. Values greater than 1% for model delineation led to reduced learning accuracies indicating that a physical delineation potentially exists in this projection and categorization. Further study for adding complexity to the hypothesis segregation is necessary.

Each label for the a-priori eddy-viscosity is also associated with an input kernel of grid-resolved quantities. This kernel is given by a stencil of 9 inputs of vorticity and stream function each (for a total of 18 input variables). These 9 inputs of each field are given by a query of the field quantity at a point on the coarse grid, the 4 adjacent points in each dimension and the 4 diagonally adjacent points. Each sample of our training data thus consists of 18 inputs of vorticity and stream function and outputs given by one-hot labels for the choice of closure modeling strategy. In this article, we have leveraged the fact that the mean of vorticity and streamfunction are both very close to zero and do not necessitate normalization. In addition, the non-dimensionalized formulation of the governing equations implies that our inputs are all dimensionless. However, we note that for practical deployments of any local-kernel based machine learning queries, grid-resolved quantities must be normalized and non-dimensionalized.

Mathematically, if our input vector $\mathbf{p}$ resides in a P-dimensional space and our desired output $\mathbf{q}$ resides in a Q-dimensional space, this framework establishes a map $\mathbb{M}$ as follows:

$$\mathbb{M} : \{p_1, p_2, \ldots, p_P\} \in \mathbb{R}^P \rightarrow \{q_1, q_2, \ldots, q_Q\} \in \mathbb{R}^Q. \tag{6.10}$$

Accordingly, the framework utilized in this article leads to the following relation:

$$\mathbb{M} : \{\mathbf{p}\} \in \mathbb{R}^{18} \rightarrow \{P(\mathbf{q}|\mathbf{p})\} \in \mathbb{R}^3, \tag{6.11}$$

where our input and output spaces are given by

$$\begin{gathered}\mathbf{p}_{i,j} = \{\bar{\omega}_{i,j}, \bar{\omega}_{i,j+1}, \bar{\omega}_{i,j-1}, \ldots, \bar{\omega}_{i-1,j-1}, \bar{\psi}_{i,j}, \bar{\psi}_{i,j+1}, \bar{\psi}_{i,j-1}, \ldots, \bar{\psi}_{i-1,j-1}\}, \\ \mathbf{q}_{i,j} = \{P(\Pi^k_{i,j}|\mathbf{p}_{i,j})\},\end{gathered} \tag{6.12}$$

where i, j refer to the spatial indices on the coarse-grid (i.e., the point of deployment) and k refers to the choice of closure scenario (i.e., structural, functional or no closure). We note here that the choice of the local stencil for ANN query reflects the discretization of the governing equations (with second-order accurate stencils requiring a ± 1 query) and the use of the trapezoidal filter in AD. Also, note that our choice of input space is given by raw variable queries rather than derivatives (or other such engineered terms). This is motivated by an aversion to specify bias towards any particular quantity that may otherwise by learned implicitly by the network. However, we note that the classification workflow may benefit significantly from the inclusion of a feature engineering step prior to optimization. This is a subject of ongoing investigation.

Our optimal map $\mathbb{M}$ is then trained by the following loss-function

$$E(\mathbf{w}) = -\sum_{n=1}^{N}\sum_{k=1}^{K}\{t_{nk}\log(y_{nk}) + (1 - t_{nk})\log(1 - y_{nk})\}, \tag{6.13}$$

where $\mathbf{w}$ are the tunable weights and biases of the network, N is the total number of samples and $K = 3$ is the total number of closure scenarios. Here, t_{nk} refers to the target (or true) label of class k and sample n and y_{nk} refers to its corresponding prediction. Note that one-hot encoding ensures that t_{nk} values are always binary Bishop (2006). For reference, our architecture is trained using the open-source deep learning software Tensorflow and is optimized with the use of ADAM, a popular gradient-descent based optimizer.

Our learning architecture is given by a 5 hidden-layer deep neural network with 40 neurons each for calculating the conditional probabilities of the three closure scenarios

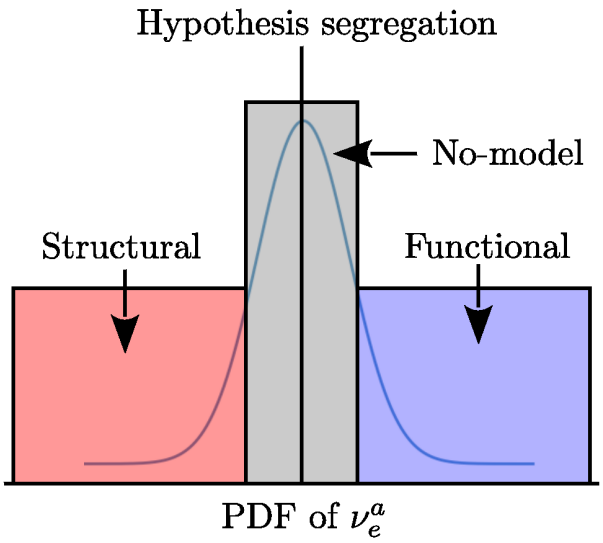


Figure 6.2: Data-segregation for one-hot labeling. The a-priori eddy-viscosities are projected onto a Gaussian distribution where data beyond 1.0% of the standard-deviation is labeled as requiring structural (if negative) or functional (if positive) modeling. The remaining data points are classified as no-model cases.

pointwise in space and time. The hidden layer neurons employ a rectified-linear activation and the output-layer gives us softmax probabilities for the three classes. The scenario with the highest conditional probability is then deployed for model classification i.e.,

$$\Pi_{i,j}^{ML} = \Pi_{i,j}^{k} \text{ s.t. } \underset{k}{\text{argmax}}\, P(\Pi_{i,j}^{k}|\mathbf{p}_{i,j}) \tag{6.14}$$

where $\Pi_{i,j}^{ML}$ refers to the machine learning based turbulence model computation at a point. In the case of model blending, the conditional probabilities for closure scenarios are used to find a linear combination of the standard Smagorinsky and the AD closures. In other words,

$$\Pi_{i,j}^{ML} = P(\Pi_{i,j}^{AD}|\mathbf{p}_{i,j})\Pi_{i,j}^{AD} + P(\Pi_{i,j}^{SM}|\mathbf{p}_{i,j})\Pi_{i,j}^{SM} \tag{6.15}$$

where $\Pi_{i,j}^{AD}$ and $\Pi_{i,j}^{SM}$ are AD and Smagorinsky predictions for the turbulence model

at a point. We note that the same learning framework is deployed in these two conceptually different scenarios.

The framework is trained using the previously mentioned categorical cross-entropy error minimization for the one-hot encoded targets. A three-fold cross-validation is utilized with a grid search for the number of layers (between 1 to 8) and number of neurons (between 10 to 100 at intervals of 10) to arrive at the optimal architecture mentioned previously. This optimal network is then trained for 2000 epochs to arrive at a classification accuracy of 79% for training and approximately 68% accuracy for validation. Convergence in validation loss was observed at around 1500 epochs as shown in Figure 6.3. We note here that our validation data (amounting to one-third of the total training data set) was not exposed to the network during gradient calculation in the back-propagation based training procedure. Effectively, our learning is derived from two-thirds of the total training data while our best-model is chosen from that with the lowest validation loss. This is to ensure that the chances of network extrapolation are minimized. The optimal learning is then deployed into a-posteriori evolution of the Kraichnan turbulence test case where a pointwise closure deployment is performed for a variety of test cases. We also note that our labeled data is pre-processed to ensure that an equal number of samples are available from each classification regime to prevent our learning from prioritizing one outcome over the other two.

6.5 Results

We proceed by examining the performance of our framework for various a-posteriori deployments which act as a rigorous testing of our learning for both classification and blending. We remind the reader that a-posteriori deployments of learning frameworks imply performance assessments in the presence of challenging numerical errors and represent the ultimate test of a data-driven framework. Briefly, the Kraichnan turbu-

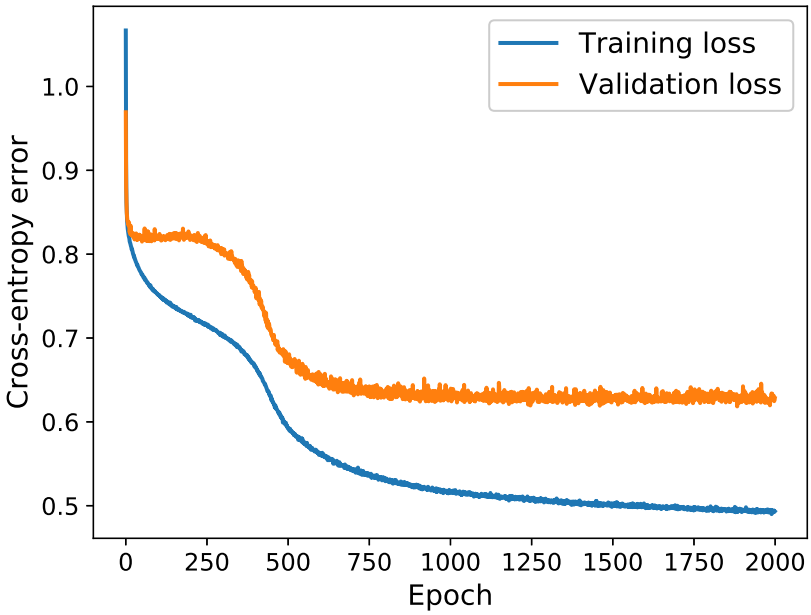


Figure 6.3: Neural-network training and validation loss for the proposed learning framework showing convergence at around 1500 epochs. The best model was chosen according to lowest validation loss for reduced over-fitting in forward deployments.

lence problem is specified by periodic boundary conditions on a rectangular domain and an initial condition which is given by an energy spectrum in wave number space. In this two-dimensional problem very fine scales are developed quickly and this leads to the classical k^{-3} scaling of the kinetic energy spectra which is a characteristic of the cascade of enstrophy in two-dimensional turbulence. The turbulence then decays gradually over time and thus represents an unsteady closure modeling assessment for our proposed framework.

We assess the viability of the proposed framework through energy spectra calculations of various reduced-order deployments as well as vorticity structure functions obtained from the same. Time histories of the turbulent kinetic energy (denoted TKE) and the variance of vorticity (denoted $\sigma^2(\bar{\omega})$) are also plotted for forward deployments. Detailed explanations of the numerical schemes and energy spectra calculations utilized for this problem may be found in Maulik and San (2017*c*). Briefly, all

our spatial numerical schemes are second-order accurate and our time-integration is third-order total-variation diminishing. Our vorticity structure function calculations are given by Grossmann and Mertens (1992):

$$S_\omega = \langle |\bar{\omega}(\mathbf{x}+\mathbf{r}) - \bar{\omega}(\mathbf{x})|^2 \rangle, \tag{6.16}$$

where the angle-brackets indicate ensemble averaging and $\mathbf{x}$ indicates a position on the grid with $\mathbf{r}$ being a certain distance from this location. Our turbulent kinetic energy is given by

$$TKE = \mu(u_f^2 + v_f^2), \tag{6.17}$$

where u_f and v_f are fluctuating quantities given by

$$u_f = \bar{u} - \mu(\bar{u}) \tag{6.18}$$

$$v_f = \bar{v} - \mu(\bar{v}), \tag{6.19}$$

and where $\mu(a)$ implies the spatial mean of a field variable a. We note that the components of velocity u, v are computed by second-order accurate central finite-difference implementations of

$$\bar{u} = \frac{\partial \bar{\psi}}{\partial y}, \quad \bar{v} = -\frac{\partial \bar{\psi}}{\partial x}. \tag{6.20}$$

In a similar manner the variance of vorticity, at each time step, is computed using

$$\sigma^2(\bar{\omega}) = \mu\left((\bar{\omega} - \mu(\bar{\omega}))^2\right). \tag{6.21}$$

In all the following assessments, the proposed framework is denoted as ML (and specified to be deployed as a classifier or a blender) and it is compared to the AD and

DS models. We remind the reader that the framework utilizes the *static* Smagorinsky model (denoted SM) with $C_s = 1.0$ within its formulation but is assessed against the DS approach. The reader may note that the value of $C_s = 1.0$ proves over-dissipative for this particular test case as shown in Maulik et al. (2019).

6.5.1 Model classification

In this section, we deploy our learning framework as a classifier which spatio-temporally switches between three closure modeling hypotheses during flow-evolution. Figure 6.4 shows the performance of our proposed framework for the forward deployment of the Kraichnan turbulence problem in the form of energy spectra predictions at $t = 4$. For comparison, no-model results (denoted UNS), the DS method and AD are also shown along with DNS spectra. One can observe that the classifier balances the dissipative natures of the SM and AD hypothesis to obtain a performance similar to the that of the DS approach. While at the lower wavenumbers, the AD procedure seems to be more accurate in statistical capture, the higher wavenumbers are stabilized adequately by the classifier. We would like to note that that SM hypothesis with $C_s = 1.0$ is highly dissipative and this results shows that the classifier avoids its deployment to a large degree for improved a-posteriori performance. We clarify that for spectral cut-off filtering, FDNS spectra and DNS spectra are identical till the grid cut-off wave number (Maulik et al., 2018).

Figure 6.5 details vorticity structure function assessments in our domain where assessments with FDNS show that the proposed framework is adequately capable of stabilizing turbulence correlations at $t = 4$. We note that the structure functions are predicted more accurately by AD at low values of $\mathbf{r}$ whereas the proposed classifier behaves similar to a DS implementation thereby indicating a dynamic dissipation on the grid. It may be so that the adaptive dissipation prioritizes noise removal and thus introduces errors at low values of $\mathbf{r}$ as seen through stable structure functions

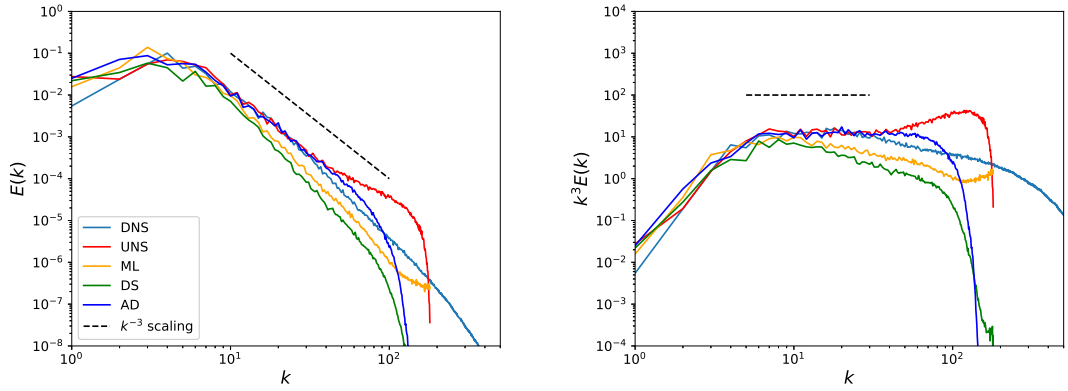


Figure 6.4: *A posteriori* kinetic-energy spectra (left) and compensated kinetic-energy spectra (right) for $Re = 32000$ at $t = 4$ and at $N^2 = 256^2$ degrees of freedom. The proposed framework (deployed as a classifier) balances the dissipative natures of the AD and the DS models.

at saturation (i.e., at higher values of $\mathbf{r}$). A further assessment is deployed in the form of time-histories of TKE and $\sigma^2(\bar{\omega})$ as shown in Figure 6.6. Once again, the classifier is seen to have a varying trend in TKE predictions compared to the AD and DS techniques indicating varying dissipation strengths. The vorticity variance predictions are also seen to be balanced between that of the DS and AD models indicating the balance of dissipative tendencies.

We proceed by performing a thorough validation of our learning framework by assessing its performance for prediction tasks that it has not been exposed to in training. This is established by testing closure efficiency for a Reynolds number of 64000. We remind the reader that map optimization was performed solely for $Re = 32000$ and this represents an additional validation of the learning. Kinetic energy spectra for this experiment are shown in Figure 6.7 where it is observed that the classifier performs in a very similar fashion to the $Re = 32000$ test-case with AD performing more efficiently at the lower wavenumbers of the inertial range but the ML approach stabilizing high-wavenumber noise effectively. This indicates that the learning has generalized, atleast on the current degree of coarse-graining. We also perform additional assessments such as those shown in Figure 6.8 and Figure

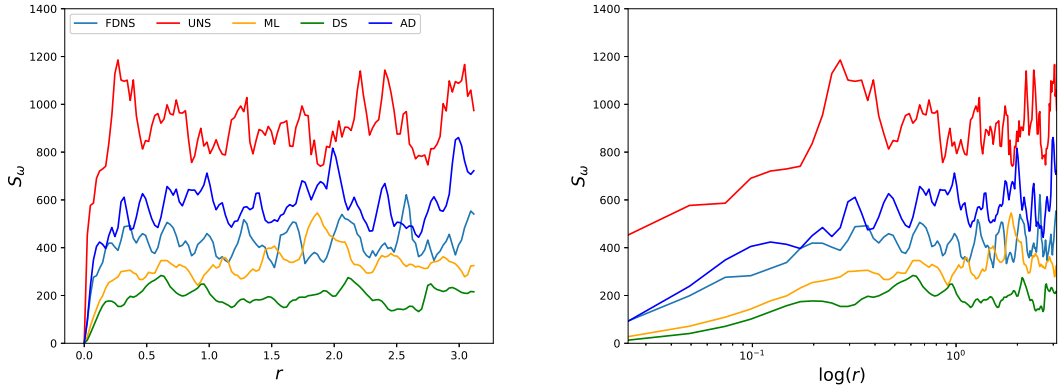


Figure 6.5: *A posteriori* vorticity structure functions plotted against $\mathbf{r}$ (left) and $\log(\mathbf{r})$ (right) for $Re = 32000$ at $t = 4$ and at $N^2 = 256^2$ degrees of freedom. It is observed that AD performs better in the near-region whereas the proposed framework behaves similar to the DS approach.

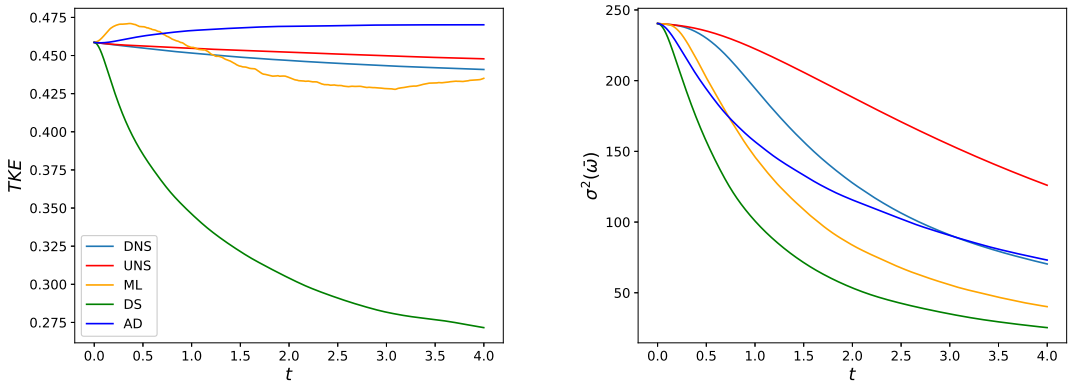


Figure 6.6: Time-histories for turbulent kinetic energy (left) and vorticity variance (right) for $Re = 32000$ at $N^2 = 256^2$ degrees of freedom. The proposed method can be seen to adapt between the behavior of the AD and DS techniques.

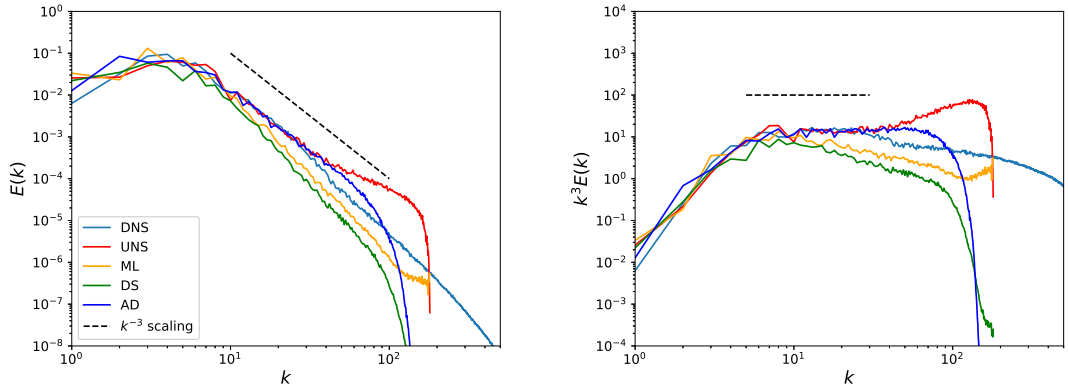


Figure 6.7: *A posteriori* kinetic-energy spectra (left) and compensated kinetic-energy spectra (right) for $Re = 64000$ at $t = 4$ and at $N^2 = 256^2$ degrees of freedom. This assessment displays closure effectiveness for a Reynolds number not utilized in the training data.

6.9. The former shows the vorticity structure function trends for this out-of-training range learning assessment and the latter shows the time-histories of TKE and $\sigma^2(\bar{\omega})$. Very similar trends for both these assessments are obtained when compared to the $Re = 32000$ test-case with time varying trends in TKE and vorticity variance capture.

While the aforementioned test-cases validated the learning of the classifier on different control parameters (and flow evolutions) given by the Reynolds number. We proceed by assessing the performance and stability of the classifier on a reduced degree-of-freedom evolution given by $N^2 = 128^2$. This test was to examine if the classifier could retain a viable learning for deployment on slightly different grid support. Figure 6.10 shows the kinetic energy spectra for a deployment at this reduced degree-of-freedom at a Reynolds number of 32000. It is observed that the proposed classifier is able to avoid inaccuracies related to AD's lack of dissipation. Indeed, it is well-known that AD requires a sufficiently fine resolution in comparison to the eddy-viscosity hypothesis based models for appropriate utilization of their inverse-filtering (Germano, 2015; Guermond et al., 2004). A similar trend may also be observed in Figure 6.11 with the vorticity structure functions where once again the AD technique proves accurate at lower distances in comparison the the DS and the ML methods.

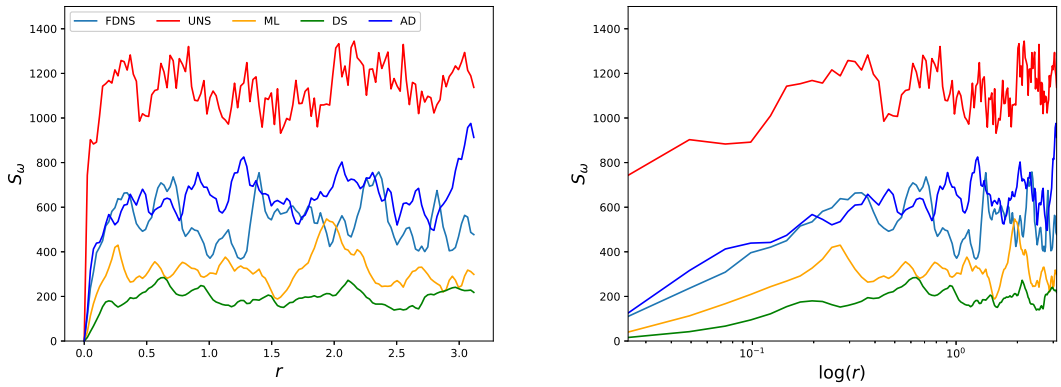


Figure 6.8: *A posteriori* vorticity structure functions plotted against $\mathbf{r}$ (left) and $\log(\mathbf{r})$ (right) for $Re = 64000$ at $t = 4$ and at $N^2 = 256^2$ degrees of freedom. It is observed that solely AD performs better in the near-region whereas the proposed framework behaves similar to the DS approach. The behavior is similar to that observed for within training data regime deployment.

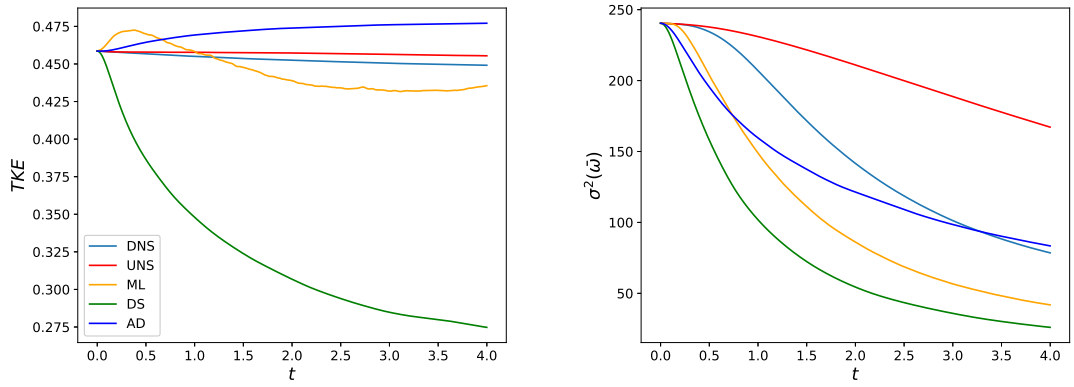


Figure 6.9: Time-histories for turbulent kinetic energy (left) and vorticity variance (right) for $Re = 64000$ at $N^2 = 256^2$ degrees of freedom. The proposed method can be seen to adapt between the behavior of the AD and DS techniques and acts as an additional validation for deployment to different Reynolds numbers.

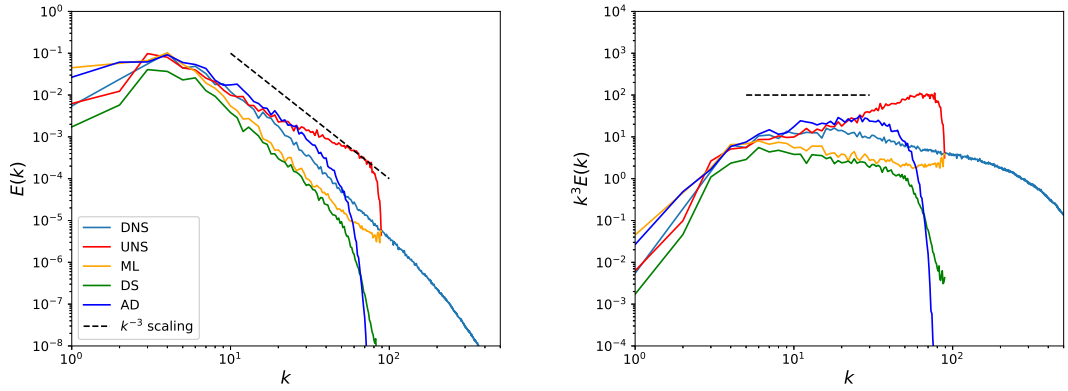


Figure 6.10: *A posteriori* kinetic-energy spectra (left) and compensated kinetic-energy spectra (right) for $Re = 32000$ at $t = 4$ and at $N^2 = 128^2$ degrees of freedom. This assessment displays closure effectiveness for a coarse-grained resolution not utilized in the training data.

The ML classifier however is slightly more accurate than the DS approach. The time-histories for TKE and vorticity variance, shown in Figure 6.12, display a greater amount of variation in the classification framework with TKE values oscillating but remaining close to the DNS results. It must be noted that the no-model and AD hypothesis prevent the classifier from going into a fully SM deployment which is highly dissipative. This explains the similarity with DS results in terms of spectra and vorticity-variance.

In addition to the test-case with a slightly reduced grid-resolution in comparison to training data generation. We also perform a grid-dependence check on the accuracy of our classification framework as shown in Table 6.1. We perform a hypothesis segregation (as introduced previously) to label all points on a coarse-grid with an optimal closure hypothesis and assess if the learning at $N^2 = 256^2$ is able to categorize them appropriately. It can be seen that accuracies around the same resolution as that of the training data are approximately similar to validation accuracy during network optimization. However, on intense coarse-graining, accuracies are seen to drop significantly. However, we note that even at the coarsest resolution of $N^2 = 32^2$, accuracies greater than 33% indicate some form of learning retention.

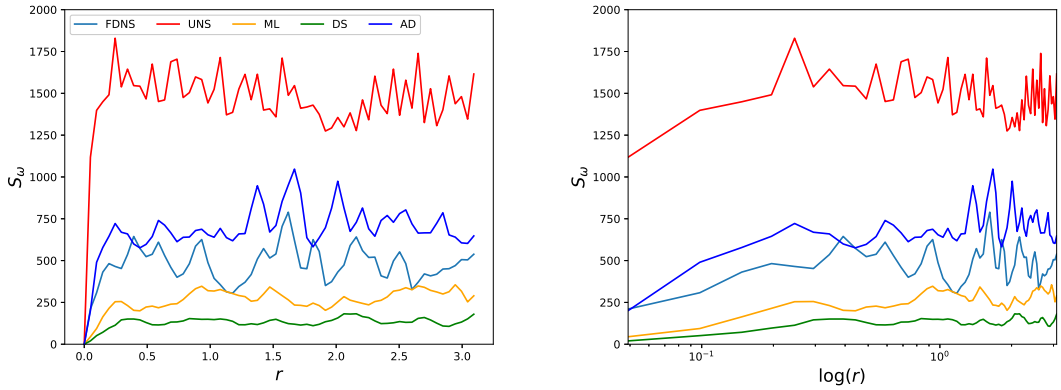


Figure 6.11: *A posteriori* vorticity structure functions plotted against $\mathbf{r}$ (left) and $\log(\mathbf{r})$ (right) for $Re = 32000$ at $t = 4$ and at $N^2 = 128^2$ degrees of freedom. It is observed that solely AD performs better in the near-region whereas the proposed framework behaves similar to the DS approach. The behavior is similar to that observed for within training resolution deployment.

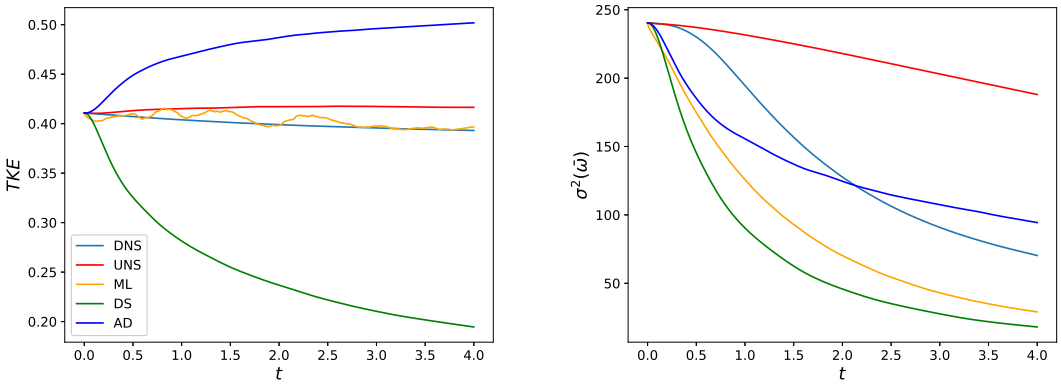


Figure 6.12: Time-histories for turbulent kinetic energy (left) and vorticity variance (right) for $Re = 32000$ at $N^2 = 128^2$ degrees of freedom. The proposed method can be seen to adapt between the behavior of the AD and DS techniques and acts as an additional validation for deployment to similar coarse-grained resolutions.

Time	$N^2 = 512^2$	$N^2 = 256^2$	$N^2 = 128^2$	$N^2 = 64^2$	$N^2 = 32^2$
$t = 1$	65.77	63.04	56.51	52.17	47.65
$t = 2$	60.89	60.47	61.02	55.62	41.99
$t = 3$	68.05	65.08	61.54	53.32	46.29
$t = 4$	63.93	66.04	60.24	48.33	48.54

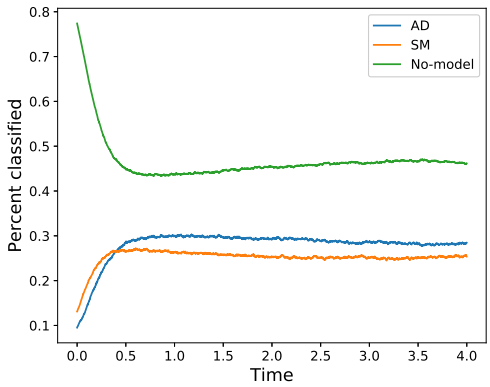
Table 6.1: Classification accuracy percentages for different grid-resolutions in a-priori to illustrate how accurately our base learning can predict correct labels. Accuracies are seen to drop when resolutions are coarsened radically. However, some learning is retained as evidenced by accuracies greater than 33%.

Time	*A priori*			*A posteriori*		
	AD	SM	No-model	AD	SM	No-model
$t = 1$	22.43	21.69	55.87	29.94	26.34	43.72
$t = 2$	22.31	21.08	56.60	29.17	25.37	45.45
$t = 3$	21.37	20.84	57.78	28.68	25.07	46.25
$t = 4$	19.49	22.56	57.94	28.45	25.38	46.17

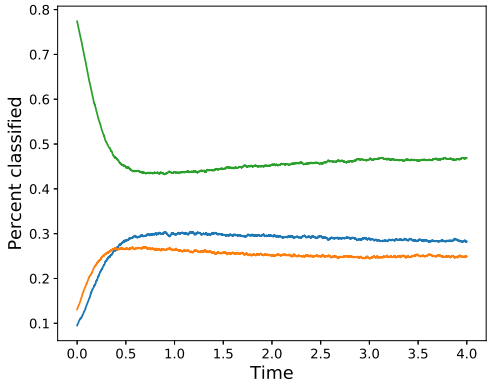
Table 6.2: Classification percentages in a-priori and a-posteriori. One can see deviation from trends due to numerical error accumulation (and greater utilization of closure classifications for subsequent stabilization).

We also determine the effect of network deployment in the presence of numerical errors as shown in table 6.2 where it can be seen that a significant difference in hypothesis choices are observed. In particular, the a-posteriori deployment of the classifier is seen to utilize a greater proportion of the turbulence closure hypotheses, in comparison to the no-model ones. This may be considered as proof of the classifier detecting greater stabilization requirements due to numerical error build-up. It is observed that the AD approach shows a greater increase in deployment than SM. This may be to offset the rather large inaccuracies of the lower wavenumbers in the exceptionally dissipative SM approach. Understanding the nature of classifier adaptation in the presence of numerical errors is an interesting subject of future research that may aid in improved decision making frameworks. We complement the data in table 6.2 by outlining the classification percentages of different hypotheses plotted against time for our three a-posteriori deployments in Figure 6.13. One may notice that the deployment of the framework at the coarser resolution of $N^2 = 128^2$ requires a higher degree of SM and AD classifications for successful stabilization. All experiments are seen to show a gradual increase in closure requirement as scale-separation grows with subsequent saturation in the gradual decay.

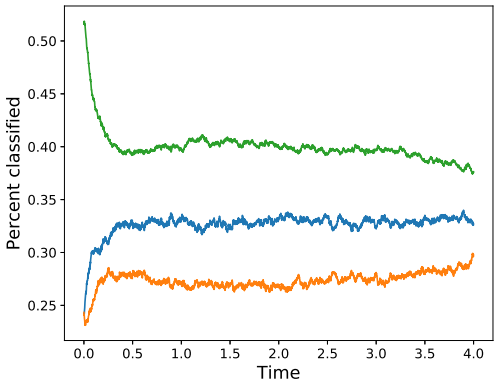
As a final qualitative analysis of our classifier, we plot *a posterior* contours from forward deployments at $N^2 = 256^2$ and $Re = 32000$. In Figure 6.14, vorticity contours from the ML, DS, AD and UNS simulations are shown to assess the stabilization effect



(a) $Re = 32000$, $N^2 = 256^2$



(b) $Re = 64000$, $N^2 = 256^2$



(c) $Re = 32000$, $N^2 = 128^2$

Figure 6.13: The a-posteriori classification percentages of the various modeling hypotheses for our three forward deployments. In all deployments it is observed that the utilization of AD and SM increases as the scale-separation grows and saturates for the slow decay. Noticeably, the deployment at $N^2 = 128^2$ necessitates a higher proportion of AD and SM classifications for improved stabilization.

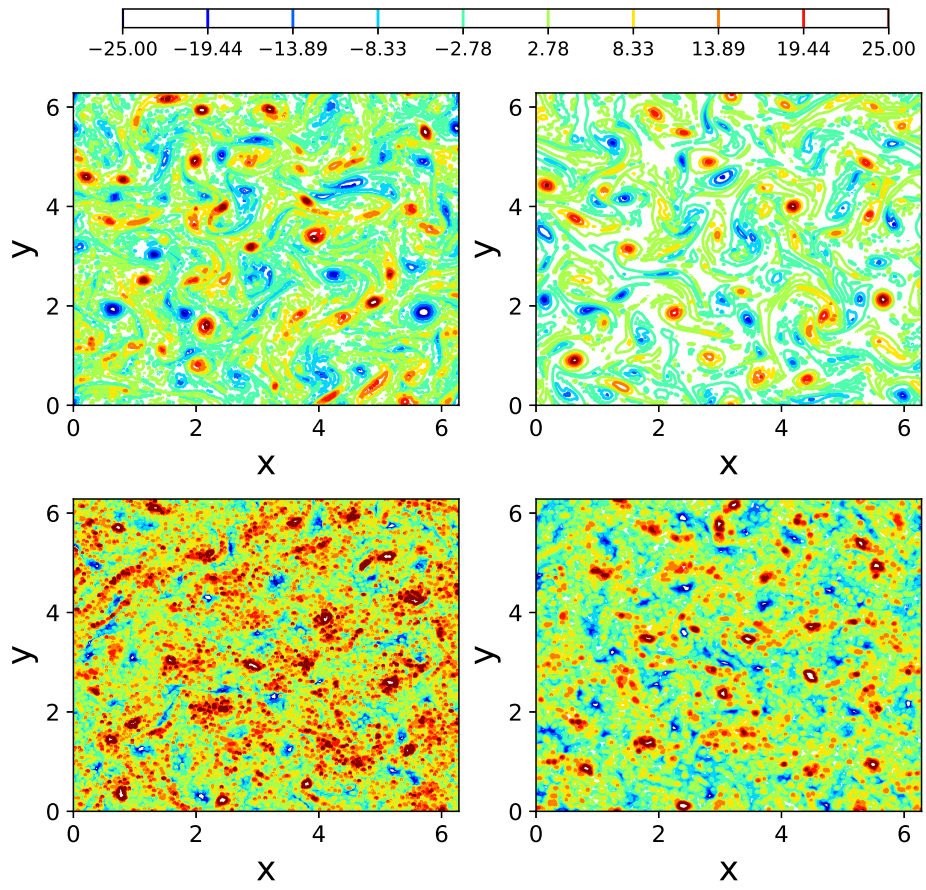


Figure 6.14: *A posteriori* contour results for $Re = 32000$ with the proposed classification framework shown top-left, DS shown top-right, UNS shown bottom left and AD shown bottom right. These may be compared against FDNS contours qualitatively (in Figure 6.1).

of the different frameworks. The classifier can be seen to stabilize high-wavenumber noise adequately, in a manner similar to DS as previous statistics have reflected. The AD approach may be observed to be contaminated with noise that may potentially be harmful for long-time integration.

6.5.2 Model blending

In this section, we deploy our learning in a different manner by utilizing their outputs (i.e., the conditional probabilities of each hypothesis) as a pre-multiplier of the prediction of each modeling hypothesis. We utilize this formulation instead of the

direct prediction of sub-grid contribution coefficients by observing that a greater degree of stability is imparted to the flow-evolution. Indeed, direct regression with sub-grid quantities has been seen to require a-posteriori post-processing for stability due to energy accummulation in the super-grid (when negative eddy-viscosities are predicted effectively)(Maulik et al., 2018, 2019). We recognize (as a limitation), that the utilization of a conditional-probability outputs to linearly combine turbulence modeling predictions from different hypotheses digresses from the core idea of a categorical cross-entropy error minimization. However, as results shall show, the proposed method acts as an effective instrument for blending models in a-posteriori with the requirement of any truncation for stability. We would also like to emphasize here that the same learning is applicable for both classification and blending. We perform a similar set of assessments as outlined in the section 6.5.1.

Figure 6.15 shows the performance of the blending formulation for a Reynolds number of 32000 and at $t = 4$ with $N^2 = 256^2$ degrees-of-freedom with kinetic energy spectra. It is observed that the proposed procedure recovers a dissipative behavior that is very similar to the DS approach. It does this by balancing the coefficients of the AD and SM predictions which adapt to the dynamic dissipation requirement of the flow. Overall, it is observed that the framework behaves in a similar manner to the classifier presented previously with dissipation preventing the accummulation of high-wavenumber errors but causing a mismatch in the inertial range spectra capture. However, the dissipation is dynamic and it prevents the overwhelming damping of the SM deployment by balancing with the AD predictions adaptively. This is reflected in Figure 6.16 as well where the vorticity structure functions once again show that the AD method is more accurate at lower values of $\mathbf{r}$ but the blending allows for a prediction akin to the DS technique. Figure 6.17 shows the time-histories of the TKE and the $\sigma^2(\bar{\omega})$ for the proposed framework compared to DS, AD and UNS. The vorticity-variance shows a trend close to the DS approach as expected but the TKE

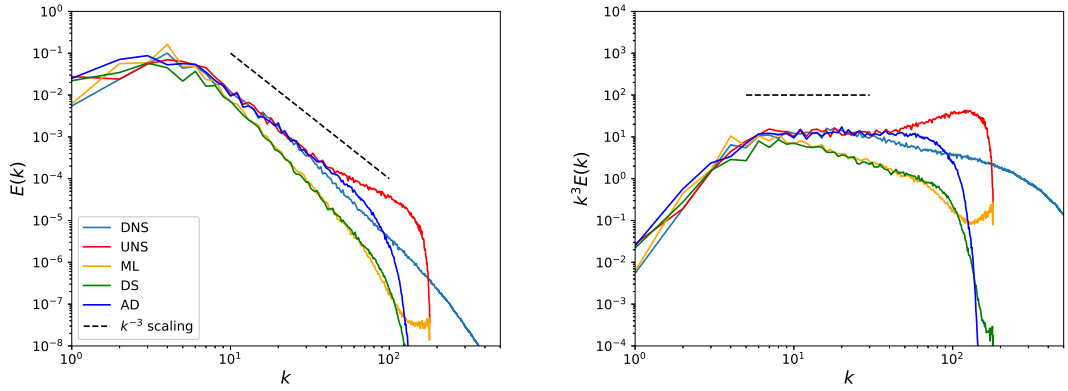


Figure 6.15: *A posteriori* kinetic-energy spectra (left) and compensated kinetic-energy spectra (right) for $Re = 32000$ at $t = 4$ and at $N^2 = 256^2$ degrees of freedom. The proposed framework (deployed as a model blending mechanism) behaves similar to the DS approach at the inertial wavenumbers. We remind the reader that the blending is dynamic between AD and SM.

trends are once again not uniform.

In a fashion similar to that employed in section 6.5.1, we deploy assessments of the blending method to out-of-training predictions for validation. We start with an *a poseirori* deployment at $Re = 64000$ and $N^2 = 256^2$ degrees-of-freedom and observe that the learning is sufficiently generalizable. This is observed from Figure 6.18 where kinetic energy spectra show an aligned prediction to the previous test-case. Vorticity structure functions and time-histories, shown in Figures 6.19 and 6.20 respectively show a similar behavior to the one observed for $Re = 32000$. This implies that the learning, whether utilized as a classifier or a blending mechanism, is generalizable. We also deploy the blending framework at a different degree-of-freedom ($N^2 = 128^2$) to assess it is stable to a slightly different grid support and trends similar to the classifier are observed wherein the framework focuses on dissipation to stabilize the higher wavenumbers in contrast with AD. This is observed in Figure 6.21 for kinetic energy spectra and Fig 6.22 for the vorticity structure functions. Figure 6.23 shows time-series quantities for this test-case with both TKE and vorticity-variance trends resembling the DS method closely. This also echoes with the performance of the

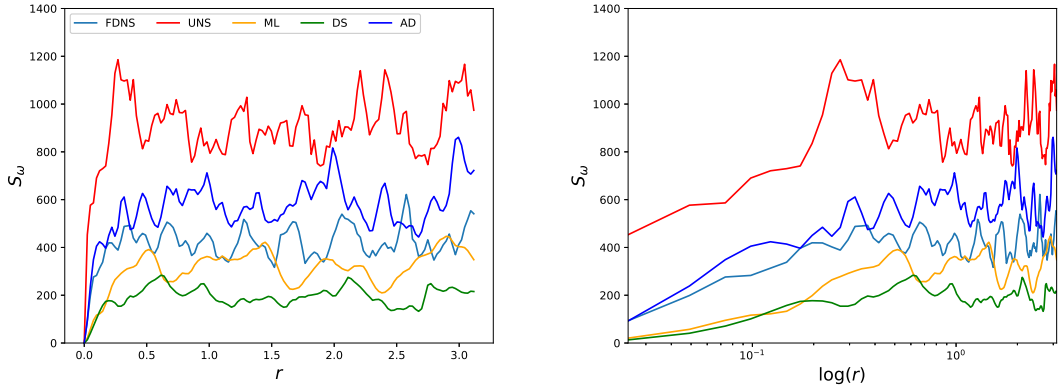


Figure 6.16: *A posteriori* vorticity structure functions plotted against $\mathbf{r}$ (left) and $\log(\mathbf{r})$ (right) for $Re = 32000$ at $t = 4$ and at $N^2 = 256^2$ degrees of freedom. It is observed that solely AD performs better in the near-region whereas the proposed blending (once again) behaves similar to the DS approach. We remind the reader that the blending is dynamic between AD and SM.

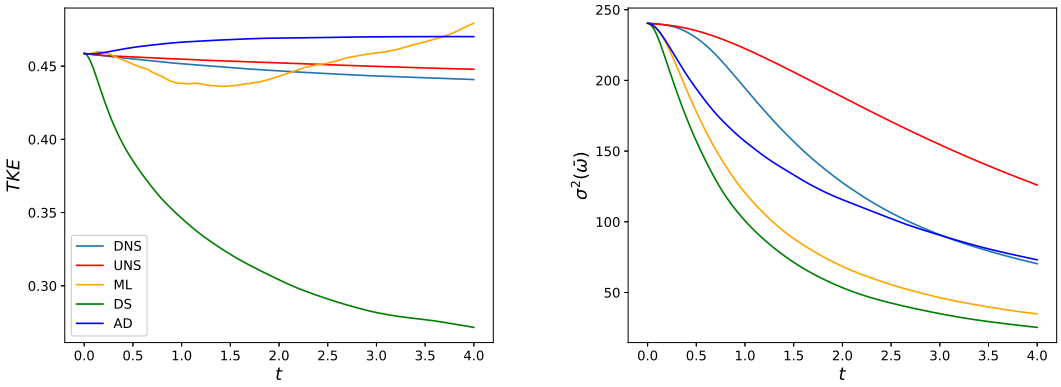


Figure 6.17: Time-histories for turbulent kinetic energy (left) and vorticity variance (right) for $Re = 32000$ at $N^2 = 256^2$ degrees of freedom. The proposed blending technique shows a varying TKE capture behavior due to its adaptive dissipation. Note that the blending is dynamic between AD and SM.

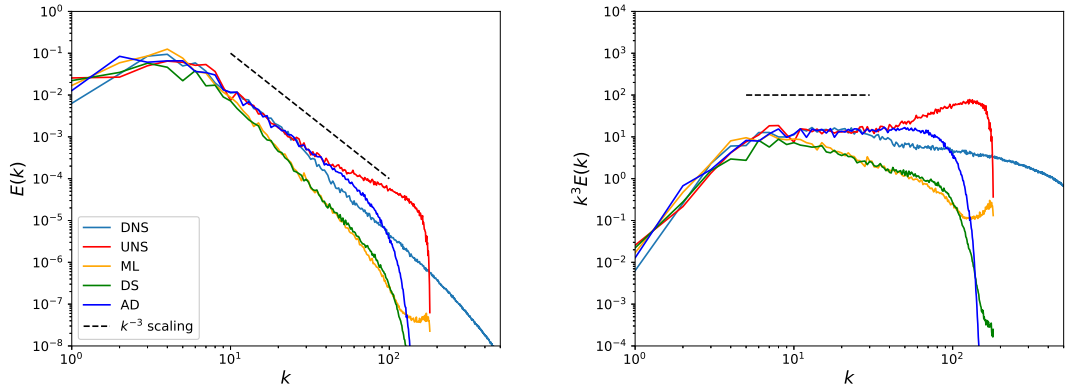


Figure 6.18: *A posteriori* kinetic-energy spectra (left) and compensated kinetic-energy spectra (right) for $Re = 64000$ at $t = 4$ and at $N^2 = 256^2$ degrees of freedom. The proposed framework (deployed as a model blending mechanism) behaves similar to the DS approach at the inertial wavenumbers. Note that the blending is dynamic between AD and SM and training is performed using $Re = 32000$ data alone.

classifier where a coarser-grid resolution led to a performance that was observed to be biased towards the eddy-viscosity hypothesis. However, further studies are necessary to quantify how the model orients itself to compensate for loss of grid-resolution or anisotropies in the flow configuration in a-posteriori deployment.

To conclude this section we show qualitative results from the vorticty contours at the final time of the numerical experiments for our proposed framework and their benchmark counterparts in Figure 6.24. This examination gives us an intuition of the stabilization effect of the proposed framework and it is seen that the predictions are very closely aligned with the DS results. This, once again, validates our dynamic dissipation hypothesis.

6.6 Conclusions & significance

In this article we have proposed a novel data-driven strategy to dynamically assess the utility of a turbulence modeling hypothesis in an LES framework. This strategy is built on the hypothesis that DNS data may be utilized to assess areas where structural or functional models may be more appropriate in an LES deployment. Our

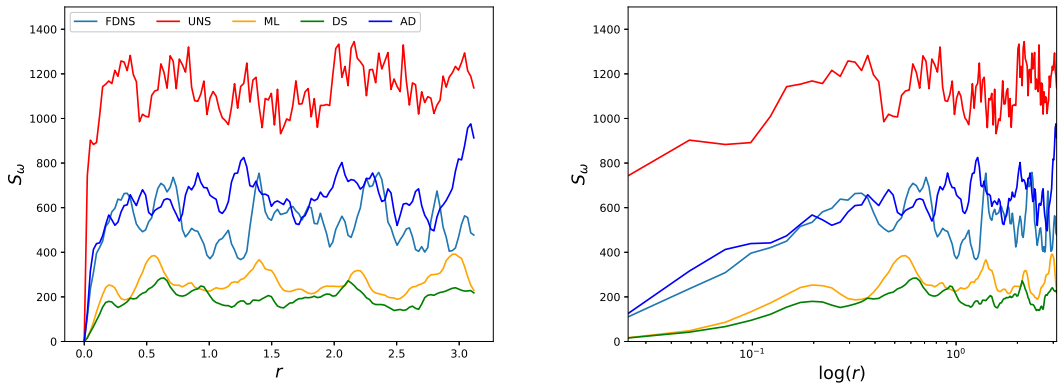


Figure 6.19: *A posteriori* vorticity structure functions plotted against $\mathbf{r}$ (left) and $\log(\mathbf{r})$ (right) for $Re = 64000$ at $t = 4$ and at $N^2 = 256^2$ degrees of freedom. It is observed that solely AD performs better in the near-region whereas the proposed blending (once again) behaves similar to the DS approach. We remind the reader that the blending is dynamic between AD and SM and training is performed using $Re = 32000$ data alone.

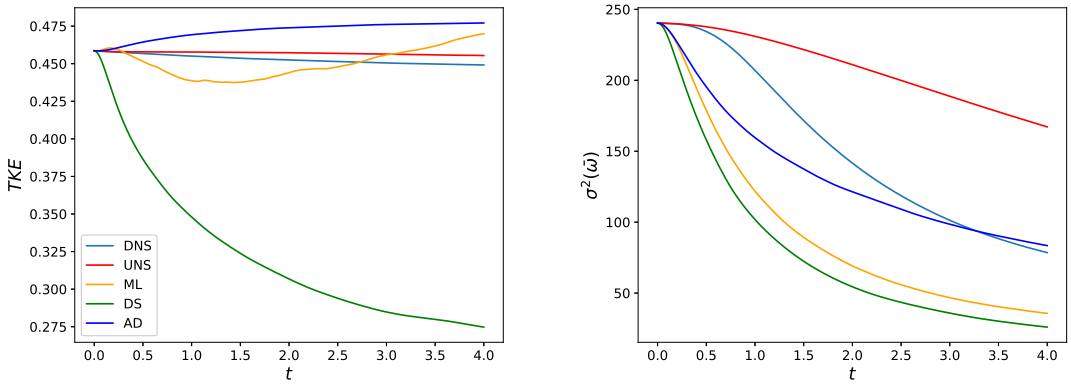


Figure 6.20: Time-histories for turbulent kinetic energy (left) and vorticity variance (right) for $Re = 64000$ at $N^2 = 256^2$ degrees of freedom. We remind the reader that the blending is dynamic between AD and SM and training is performed using $Re = 32000$ data alone.

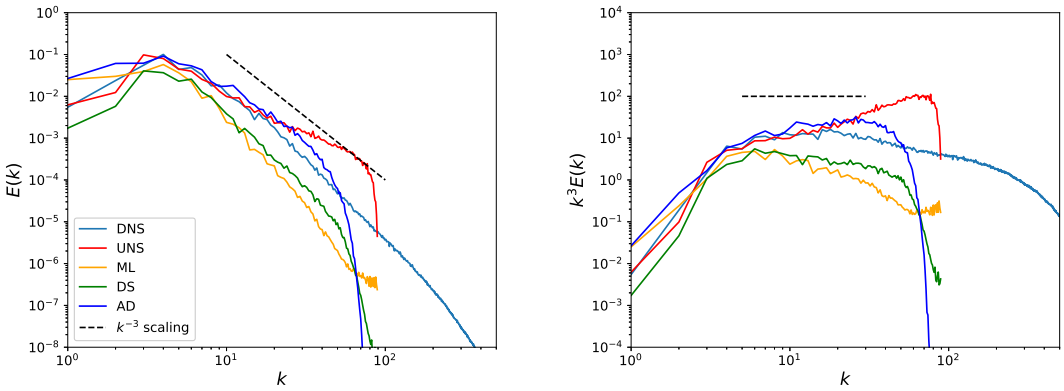


Figure 6.21: *A posteriori* kinetic-energy spectra (left) and compensated kinetic-energy spectra (right) for $Re = 32000$ at $t = 4$ and at $N^2 = 128^2$ degrees of freedom. The proposed framework (deployed as a model blending mechanism) behaves similar to the DS approach at the inertial wavenumbers. We remind the reader that the blending is dynamic between AD and SM and training is performed using $Re = 32000$ and $N^2 = 256^2$ data alone.

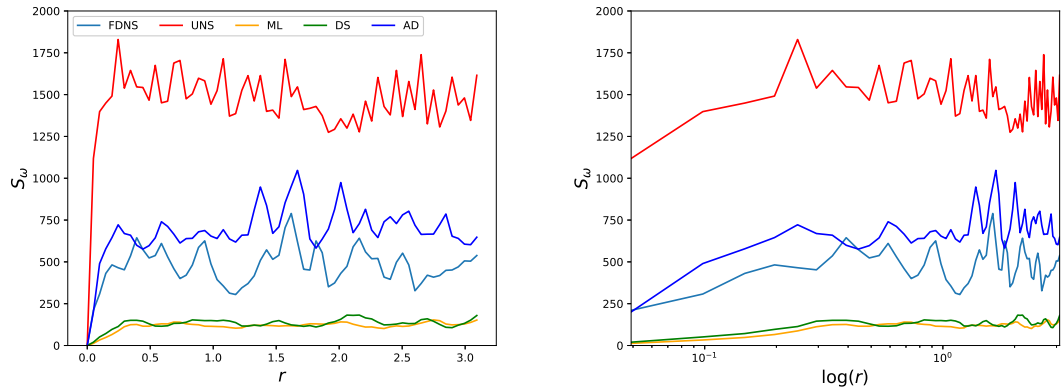


Figure 6.22: *A posteriori* vorticity structure functions plotted against $\mathbf{r}$ (left) and $\log(\mathbf{r})$ (right) for $Re = 32000$ at $t = 4$ and at $N^2 = 128^2$ degrees of freedom. It is observed that solely AD performs better in the near-region whereas the proposed blending (once again) behaves similar to the DS approach. We remind the reader that the blending is dynamic between AD and SM and training is performed using $Re = 32000$ and $N^2 = 256^2$ data alone.

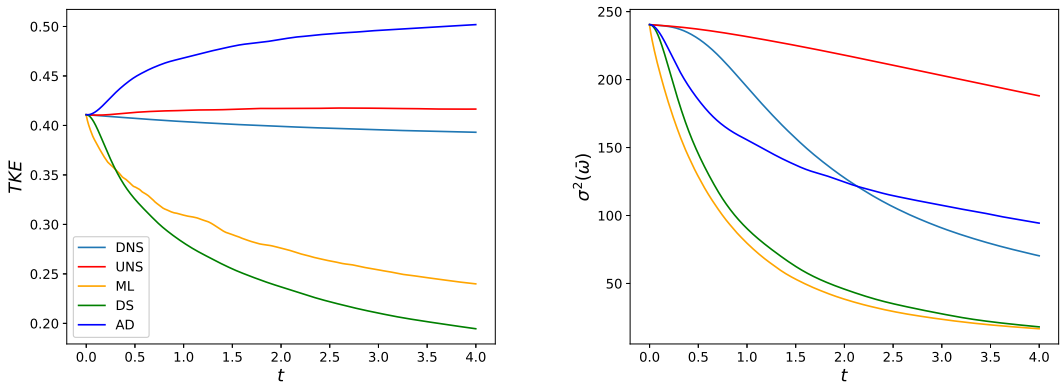


Figure 6.23: Time-histories for turbulent kinetic energy (left) and vorticity variance (right) for $Re = 32000$ at $N^2 = 128^2$ degrees of freedom. The proposed blending technique behaves more dissipatively due to the reduced grid-support. We remind the reader that the blending is dynamic between AD and SM and training is performed using $Re = 32000$ and $N^2 = 256^2$ data alone.

hypothesis segregation and subsequent training culiminates in a learning that may deployed as a classifier of turbulence models at each point on the LES grid as well as a blending technique for balancing turbulence models with different dissipative strengths. When deployed as a classifier, our proposed framework may also predict a 'no-model' situation wherein no sub-grid source-term is deployed. When deployed as a blending mechanism, the learning linearly combines the AD and static Smagorinsky hypothesis premultiplied by their respective conditional probabilities to obtain another hybrid dissipation mechanism. Both frameworks utilize the same learning and are assessed through similar experiments in a-posteriori.

We have rigorously assessed the deployment of our machine learning strategy through the utilization of a Kraichnan turbulence test case. Our assessments are made for Reynolds number values both within and outside that utilized in training to ensure that a generalizable turbulence closure has been developed. In addition, we have also assessed if the proposed closure can be deployed on a coarser grid than one it was trained for. The dissipative and scale-content capture of the proposed framework is compared to the AD and DS techniques through the use of kinetic-energy spectra,

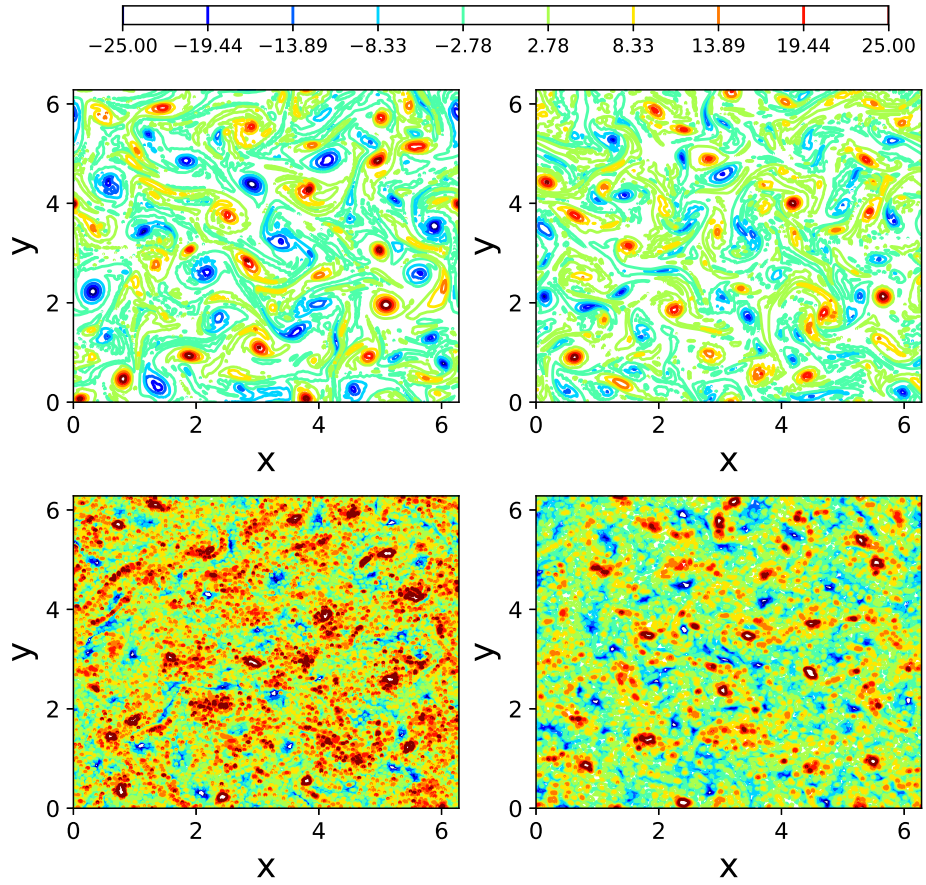


Figure 6.24: *A posteriori* contour results for $Re = 32000$ with the proposed blending framework shown top-left, DS shown top-right, UNS shown bottom left and AD shown bottom right. These may be compared against FDNS contours qualitatively (in Figure 6.1).

vorticity structure functions and time-histories of TKE and vorticity-variance showing a dynamic dissipation akin to the DS. In particular, the statistical fidelity of the data-driven frameworks is seen to be inferior to the AD technique, which provides better estimates of the kinetic energy spectra at lower wavenumbers and also provides most accurate estimates of the vorticity structure function. However, the focus on high-wavenumber noise attentuation leads to no grid cut-off error accummulation and the statistical results of the ML models are very similar to DS in all assessments. Also, it is observed that the data-driven closure (whether deployed as a classifier or a blending instrument) adequately captures the k^{-3} scaling expected for the kinetic energy spectra for the Kraichnan turbulence case and attempts to strike an optimal balance between the dissipative functional kernel and the noise-prone structural kernel. This behavior is interesting as the model classifies solely between AD and the static Smagorinsky hypothesis indicating the extreme dissipation of the latter at $C_s = 1.0$ is effectively alleviated by the spatiotemporal blending. Our closure, thus, attempts to blend the strengths of both modeling strategies to overcome their individual weakness while attempting to preserve trends from DNS.

In terms of future opportunities for this idea, the data-driven element of closure identification lends to the potential development of closures that may discern the physical characteristics of different flow scenarios. However, some challenges associated with progress in this research include considerations of invariance properties, which we have identified as a next step. While the computational costs of the proposed framework have not been studied in detail, an efficient deployment of the proposed framework would need graphical processing unit integration of any practical CFD simulation. The latter would lead to efficient learning queries since all the spatial domain information would be available to the common memory. Another future direction identified in this research is the exposure of different two-dimensional turbulence physics to the classification framework to identify if closure choices can ralso be

influenced by the training data regime. Success in that regard would allow for 'train and forget' closures in problems that have unsteady physics that span fundamentally different turbulence modeling requirements.

CHAPTER VII

Connecting implicit and explicit large eddy simulations of two-dimensional turbulence through machine learning

7.1 Abstract

In this article, we utilize machine learning to dynamically determine if a point on the computational grid requires implicit numerical dissipation for large eddy simulation (LES). The decision making process is learnt through a-priori training on quantities derived from direct numerical simulation (DNS) data. In particular, we compute eddy-viscosities obtained through the coarse graining of DNS quantities and utilize their distribution to categorize areas that require dissipation. If our learning determines that closure is necessary, an upwinded scheme is utilized for computing the non-linear Jacobian. In contrast, if it is determined that closure is unnecessary, a symmetric and second-order accurate energy and enstrophy preserving Arakawa scheme is utilized instead. This results in a closure framework that precludes the specification of any model-form for the small scale contributions of turbulence but deploys an appropriate numerical dissipation from explicit closure driven hypotheses. This methodology is deployed for the Kraichnan turbulence test-case and assessed through various statistical quantities such as angle-averaged kinetic energy spectra and vorticity structure functions. Our framework thus establishes a direct link between the use of explicit LES ideologies for closure and numerical scheme-based modeling of

This chapter is adapted from Maulik et al., under review for publication in Phys. Fluids

turbulence leading to improved statistical fidelity of a-posteriori simulations.

7.2 Introduction

Over the past decade, advances in data collection and increasing access to computational resources have led to a revolution in the use of data-driven techniques for the solution of complex inverse problems. One such problem is that of turbulence, the multiscale nature of which causes extreme computational demands for most practical systems. As a result, turbulence requires the use multiple modeling approximations for the higher wavenumbers which remain unsupported by computational degrees of freedom. One such modeling approach is that of large eddy simulation (LES) (Sagaut, 2006), which attempts to simulate the evolution of the smaller wavenumbers while the unresolved frequencies are modeled by an algebraic or differential equation. As such, the basic premise of LES is extendable to many partial differential equation systems with quadratic non-linearities. The procedure of modeling these smaller scales is often denoted *closure* due to insufficient knowledge about higher-order wavenumber interactions with the coarse-grained system (Berselli et al., 2006) and remains vital for the accurate computation of many applications (Hickel et al., 2014; Yu et al., 2016; Zhou et al., 2018). From an LES point of view, the closure problem may be considered to be dominated by commutative errors in the calculation of the non-linear term as well as the defects associated with commutative errors stemming from the dynamic term. In this study, we focus on the former.

There are two main schools of thought when it comes to the LES of the Navier-Stokes equations. The first of these promotes the use of explicit closures. Explicit LES argues for the utilization of closures in the form of sub-grid models specified as algebraic or differential equations for the unresolved scales. These are built on intuitive reasoning of how the losses of coarse graining the Navier-Stokes equations may be incorporated into an LES deployment. Some of the most notable sub-grid closure

strategies are those given by the eddy-viscosity hypothesis. Within the context of the Navier-Stokes equations, it is generally accepted that the finer scales are dissipative at the Kolmogorov length scales (Kolmogorov, 1941*b*) and therefore, most turbulence models seek to specify a sub-grid dissipation (Frisch, 1995). Most sub-grid models can be traced back to the seminal work of Smagorinsky (Smagorinsky, 1963), where a model was proposed based on the concepts of an effective eddy-viscosity determined by an a-priori specified mixing length and a $k^{-5/3}$ scaling recovery for the kinetic energy content in the wavenumber domain. Similar hypotheses have also been used for two-dimensional turbulence (Leith, 1968) (often utilized as a test-bed for geophysical scenarios, for instance see Pearson and Fox-Kemper (2018) and Pearson et al. (2017)), for approximating the k^{-3} cascade in two-dimensional turbulence and generally have their roots in dimensional analysis related to the cascade of enstrophy. These models may also be classified as *functional* due to the phenomenological nature of their deployment and represent the bulk of explicit LES turbulence models used in practical deployments. Explicit LES closures may also be specified through the specification of a low-pass spatial filter to account for the unresolved scales (Bardina et al., 1980; Stolz and Adams, 1999; Layton and Lewandowski, 2003; Mathew et al., 2003) where phenomenology is bypassed but ansatz are provided for the bulk dissipative nature of the smaller scales through the control of a characteristic filter-width. In either scenario, (i.e., whether structural or functional), the choice of the phenomenology (or dissipation control parameter) plays a key role in the successful calculation of accurate a-posteriori statistics. In contrast, the implicit LES (or ILES) approach utilizes numerical dissipation to model the unresolved scales in a turbulent flow (Grinstein et al., 2007; El Rafei et al., 2017; Margolin, 2018). In essence, the predominantly dissipative effects of the smallest scales are replicated through an artificial numerical dissipation via a biased discretization used in the calculation of the non-linear advective term (Thornber et al., 2007; DeBonis, 2013). The ILES approach is popular due

to reduced algorithmic complexity and represents a union of turbulence modeling and shock capturing mechanisms but is often criticized due to the difficulties involved in quantifying the correct amount of dissipation in a turbulent flow evolution. This results in ILES methods often proving robust and stable but overly dissipative. In this work, we propose a machine learning algorithm to enable selective dissipation within an ILES deployment through the use of explicit LES concepts during the training of the learning framework.

The past few years have seen a rapid increase in the use of data-driven techniques for the spatio-temporal modeling of dynamical systems (Schmidt and Lipson, 2009; Bright et al., 2013; Xiao et al., 2015; Ma et al., 2015; Gautier et al., 2015; Brunton et al., 2016; Schaeffer, 2017; Raissi et al., 2017; Mohan and Gaitonde, 2018; Raissi and Karniadakis, 2018; Rudy et al., 2018; San and Maulik, 2018; Wan et al., 2018; Kim et al., 2018; Muravleva et al., 2018; Jin et al., 2018). When it comes to turbulence, some widely used strategies for inference include symbolic regression (Weatheritt and Sandberg, 2016, 2017*a*,*b*), where functional model-forms for RANS deployments were generated through evolutionary optimization against high-fidelity data. Other techniques incorporating Bayesian ideologies have also been used, for instance by Xiao et al. (2016) where an iterative ensemble Kalman method was used to assimilate prior data for quantifying model form uncertainty in RANS models. In Wang, Wu and Xiao (2017); Wang, Wu, Ling, Iaccarino and Xiao (2017) and Wu et al. (2018*a*), random-forest regressors were utilized for RANS turbulence-modeling given DNS data. In Singh and Duraisamy (2016) and Singh et al. (2017), an ANN was utilized to predict a non-dimensional correction factor in the Spalart-Allmaras turbulence model through a field-inversion process using experimental data. Bypassing functional formulations of a turbulence model (a focus of this study) was also studied from the RANS point of view by Tracey et al. (2015). Ling and Templeton (2015) utilized support vector machines, decision trees and random forest regressors for identifying regions of high

RANS uncertainty. A deep-learning framework where Reynolds-stresses would be predicted in an invariant subspace was developed by Ling, Kurzawski and Templeton (2016). The reader is directed to a recent review by Duraisamy et al. (2019), for an excellent review of turbulence modeling using data-driven ideas.

As shown above, the use of data-driven ideologies and in particular artificial neural networks (ANNs) has generated significant interest in the turbulence modeling community for addressing long-standing challenges. One motivation for the popularity of ANNs is that a multilayered ANN may be optimally trained to universally approximate any non-linear function (Hornik et al., 1989). In addition, the deployment of ANNs is amenable to integration within existing computational frameworks. Greater accessibility to data and ever-improving computing capabilities has also motivated the development of advanced ANN architectures for large-scale learning of complicated physical phenomena such as turbulence. Within the context of LES (and associated with the scope of this paper) there are several investigations into sub-grid modeling using data-driven techniques. In one of the first studies of the feasibility of using learning from DNS based high-fidelity data, Sarghini et al. (2003) utilized ANNs for estimating Smagorinsky model-form coefficients within a mixed sub-grid model for a turbulent channel flow. ANNs were also used for wall-modeling by Milano and Koumotsakos (Milano and Koumoutsakos, 2002) where it was used to reconstruct the near wall field and compared to standard proper-orthogonal-decomposition techniques. An alternative to ANNs for sub-grid predictions was proposed by King et al. (2016) where a-priori optimization was utilized to minimize the L^2-error between true and modeled sub-grid quantities in a least-squares sense using a parameter-free Volterra series. Maulik and San (2017*a*) utilized an extreme-learning-machine (a variant of a single-layered ANN) to obtain maps between low-pass spatially filtered and deconvolved variables in an a-priori sense. This had implications for the use of ANNs for turbulence modeling without model-form specification. A more in-depth

investigation was recently undertaken by Fukami et al. (2018) where convolutional ANNs were utilized for reconstructing from downsampled snapshots of turbulence. Maulik et al. (2018) also deployed a data-driven convolutional and deconvolutional operation to obtain closure terms for two-dimensional turbulence. Gamahara and Hattori (2017), utilized ANNs for identifying correlations with grid-resolved quantities for an indirect method of model-form identification in turbulent channel flow. The study by Vollant et al. (2017) utilized ANNs in conjuction with optimal estimator theory to obtain functional forms for sub-grid stresses. In Beck et al. (2018), a variety of neural network architectures such as convolutional and recurrent neural networks are studied for predicting closure terms for decaying homogeneous isotropic turbulence. A least-squares based truncation is specified for stable deployments of their model-free closures. Model-free turbulence closures are also specified by Maulik et al. (2018, 2019) and Wang et al. (2018), where sub-grid scale stresses are learned directly from DNS data and deployed in a-posteriori assessments. King et al. (2018) studied generative-adversarial networks and the LAT-NET (Hennigh, 2017) for a-priori recovery of statistics such as the intermittency of turbulent fluctuations and spectral scaling. A detailed discussion of the potential benefits and challenges of deep learning for turbulence (and fluid dynamics in general) may be found in the article by Kutz (Kutz, 2017).

While a large majority of the LES-based frameworks presented above utilize a least-squares error minimization technique for constructing maps to sub-grid stresses *directly* for theoretically optimal LES (Langford and Moser, 1999; Moser et al., 2009; LaBryer et al., 2015), this work is novel in that it utilizes sub-grid statistics (pre-computed from DNS data) to train a classifier. This classifier determines whether a location requires dissipation or not through a-priori experience in the learning phase. Once classified, the non-linear term at this particular point is evaluated using one of two schemes. If it is determined that the point requires no sub-grid closure, a

symmetric and second-order accurate, energy and enstrophy conserving Arakawa-scheme (Arakawa and Lamb, 1981) is utilized for the non-linear term computation. If dissipation is necessary, an upwinding scheme is utilized instead. Therefore this study may be interpreted as a machine learning framework for devising hybrid schemes for non-linear term computation with a view to reconstructing turbulence statistics in a superior fashion. We note that the classification framework devised in this study is also deployed in an aligned work to switch between explicit LES hypotheses spatio-temporally thus proving that high-fidelity DNS statistics may be qualitatively utilized to inform modeling strategies through conditional probability predictions. The article shall describe how the proposed framework is effective in moderating the larger dissipation of an upwinded-scheme through assessments on the Kraichnan turbulence test-case.

7.3 Turbulence modeling equations

The governing equations for two-dimensional turbulence are given by the Navier-Stokes equations in the vorticity-stream function formulation. In this formulation, our non-dimensional governing equation for incompressible flow may be represented as

$$\frac{\partial \omega}{\partial t} + J(\omega, \psi) = \frac{1}{Re} \nabla^2 \omega, \tag{7.1}$$

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

$$\nabla^2 \psi = -\omega. \tag{7.2}$$

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

$$J(\omega, \psi) = \frac{\partial \psi}{\partial y}\frac{\partial \omega}{\partial x} - \frac{\partial \psi}{\partial x}\frac{\partial \omega}{\partial y}. \tag{7.3}$$

The stream function and the two-dimensional velocity components are related as

$$u = \frac{\partial \psi}{\partial y}, \quad v = -\frac{\partial \psi}{\partial x}. \tag{7.4}$$

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

$$\frac{\partial \bar{\omega}}{\partial t} + J(\bar{\omega}, \bar{\psi}) = \frac{1}{Re}\nabla^2 \bar{\omega}, \tag{7.5}$$

where the overbarred variables are now evolved on a grid with far fewer degrees of freedom. Due to the reduction in supported frequencies, the non-linear Jacobian fails to capture inter-eddy interactions at different wavenumbers. If it is assumed that the finer scales of vorticity are generally dissipative in nature for two-dimensional turbulence (based on Kraichnan's cascade of enstrophy (Kraichnan, 1967)), dissipative models may be embedded into the coarse-grained evolution of the vorticity evolution equation to recover some portion of the effect of the finer scales Explicit LES closures embed dissipation into the vorticity evolution in the form of eddy-viscosity phenomenology or through structural arguments of scale-similarity. However ILES manipulates the computation of the non-linear Jacobian term to add numerical dissipation to mimic that of the unresolved frequencies. The latter framework, while numerically robust, suffers from difficulties associated with *directed* dissipation where it is often very easy to be over-dissipative in regions where sub-grid dissipation may not be as pronounced. In this article, we introduce a hybrid ILES framework that

focuses upwinding at areas where high probability of sub-grid dissipation necessity is detected.

7.4 Non-linear Jacobian computation

The study utilizes two types of non-linear term computation schemes. Our first choice is symmetric, second-order accurate and conserves energy and enstrophy to minimize numerical dissipation. This is given by the well-known second-order Arakawa scheme (Arakawa and Lamb, 1981) as detailed below. The non-linear term in Equation 7.5 may be numerically calculated on a coarse grid using

$$J^A(\bar{\omega},\bar{\psi}) = \frac{1}{3}\left(J_1(\bar{\omega},\bar{\psi}) + J_2(\bar{\omega},\bar{\psi}) + J_3(\bar{\omega},\bar{\psi})\right) \tag{7.6}$$

where $J^A(\bar{\omega},\bar{\psi})$ will henceforth refer to the Arakawa discretization. The individual terms on the right hand side of the above equation are given as

$$\begin{aligned} J_1(\bar{\omega},\bar{\psi}) = \frac{1}{4\Delta x\Delta y}\big[&(\bar{\omega}_{i+1,j}-\bar{\omega}_{i-1,j})(\bar{\psi}_{i,j+1}-\bar{\psi}_{i,j-1}) \\ &-(\bar{\omega}_{i,j+1}-\bar{\omega}_{i,j-1})(\bar{\psi}_{i+1,j}-\bar{\psi}_{i-1,j})\big], \end{aligned} \tag{7.7}$$

$$\begin{aligned} J_2(\bar{\omega},\bar{\psi}) = \frac{1}{4\Delta x\Delta y}\big[&\bar{\omega}_{i+1,j}(\bar{\psi}_{i+1,j+1}-\bar{\psi}_{i+1,j-1}) \\ &-\bar{\omega}_{i-1,j}(\bar{\psi}_{i-1,j+1}-\bar{\psi}_{i-1,j-1}) - \bar{\omega}_{i,j+1}(\bar{\psi}_{i+1,j+1}-\bar{\psi}_{i-1,j+1}) \\ &+\bar{\omega}_{i,j-1}(\bar{\psi}_{i+1,j-1}-\bar{\psi}_{i-1,j-1})\big], \end{aligned} \tag{7.8}$$

$$\begin{aligned} J_3(\bar{\omega},\bar{\psi}) = \frac{1}{4\Delta x\Delta y}\big[&\bar{\omega}_{i+1,j+1}(\bar{\psi}_{i,j+1}-\bar{\psi}_{i+1,j}) \\ &-\bar{\omega}_{i-1,j-1}(\bar{\psi}_{i-1,j}-\bar{\psi}_{i,j-1}) - \bar{\omega}_{i-1,j+1}(\bar{\psi}_{i,j+1}-\bar{\psi}_{i-1,j}) \\ &+\bar{\omega}_{i+1,j-1}(\bar{\psi}_{i+1,j}-\bar{\psi}_{i,j-1})\big]. \end{aligned} \tag{7.9}$$

The aforementioned scheme is utilized when our proposed classifier recognizes that no dissipation is necessary.

A numerically dissipative computation of the non-linear term allows for that stabilization of noise accumulation at the grid cut-off wavenumbers. Although there are many different methodologies for upwind based dissipation with varying degrees of complexity, in this article, we utilize a conventional upwind-biased scheme as detailed in the following. Our ILES Jacobian is computed as

$$\begin{aligned} J^I(\bar{\omega},\bar{\psi}) =& \bar{u}_{i,j}\frac{\bar{\omega}_{i+1,j}-\bar{\omega}_{i-1,j}}{2\Delta x}+\frac{1}{2}(\bar{u}^+\bar{\omega}_x^- + u^-\bar{\omega}_x^+) \\ &+\bar{v}_{i,j}\frac{\bar{\omega}_{i,j+1}-\bar{\omega}_{i,j-1}}{2\Delta y}+\frac{1}{2}(\bar{v}^+\bar{\omega}_y^- + \bar{v}^-\bar{\omega}_y^+), \end{aligned} \tag{7.10}$$

where

$$\bar{u}^- = \min(\bar{u}_{i,j},0), \quad \bar{u}^+ = \max(\bar{u}_{i,j},0), \tag{7.11}$$

$$\bar{v}^- = \min(\bar{v}_{i,j},0), \quad \bar{v}^+ = \max(\bar{v}_{i,j},0). \tag{7.12}$$

In addition,

$$\begin{aligned} \bar{\omega}_x^- &= \frac{\bar{\omega}_{i-2,j}-3\bar{\omega}_{i-1,j}+3\bar{\omega}_{i,j}-\bar{\omega}_{i+1,j}}{3\Delta x}, \\ \bar{\omega}_x^+ &= \frac{\bar{\omega}_{i-1,j}-3\bar{\omega}_{i,j}+3\bar{\omega}_{i+1,j}-\bar{\omega}_{i+2,j}}{3\Delta x}, \\ \bar{\omega}_y^- &= \frac{\bar{\omega}_{i,j-2}-3\bar{\omega}_{i,j-1}+3\bar{\omega}_{i,j}-\bar{\omega}_{i,j+1}}{3\Delta y}, \\ \bar{\omega}_y^+ &= \frac{\bar{\omega}_{i,j-1}-3\bar{\omega}_{i,j}+3\bar{\omega}_{i,j+1}-\bar{\omega}_{i,j+2}}{3\Delta y}. \end{aligned} \tag{7.13}$$

Note that velocity components are recovered using

$$\begin{aligned} \bar{u}_{i,j} &= \frac{\bar{\psi}_{i,j+1}-\bar{\psi}_{i,j-1}}{2\Delta y} \\ \bar{v}_{i,j} &= -\frac{\bar{\psi}_{i+1,j}-\bar{\psi}_{i-1,j}}{2\Delta x}, \end{aligned} \tag{7.14}$$

where the second-order accurate reconstruction of the velocity leads to overall second-order accuracy for non-linear Jacobian reconstruction using the upwinded procedure outlined above. We also note that our Poisson equation given by Equation 7.2 is solved using a spectrally-accurate scheme.

With the choice of one of the two aforementioned schemes, a point in space-time may or may not have an artificial dissipation imparted to it numerically. However, we mention the caveat that switching between these two schemes would mean that the kinetic energy and enstrophy preserving property of the Arakawa scheme is lost.

7.5 Machine learning for scheme selection

We now discuss the procedure of utilizing DNS data for learning to classify one of the two dissipation scenarios. Of these two options, one is given by the choice of the Arakawa scheme and the other by our upwinded computation of the Jacobian (i.e., when the classification framework has determined that the point does not require sub-grid dissipation or vice-versa respectively). This switching between scenarios is spatio-temporally dynamic. We proceed by outlining our training strategy through the utilization of DNS data. Five equidistant snapshots of DNS data at $Re = 32000$ (i.e., at $t = 0, 1, 2, 3, 4$) and at $N^2 = 2048^2$ degrees of freedom (from 40000 available snapshots) are utilized to compute the grid-filtered variables (denoted FDNS) (at $N^2 = 256^2$ degrees of freedom) through the application of a spectral cut-off filter. Perfect closure values

$$\Pi = J(\bar{\omega}, \bar{\psi}) - \overline{J(\omega, \psi)} \tag{7.15}$$

are then obtained (the reader is directed to Maulik et al. (2019) for details related to the calculation of these quantities). Note here, that the Kraichnan turbulence problem is transient with the evolution of vorticity represented in Figure 7.1 representing

different closure needs over time evolution.

We proceed by introducing the a-priori eddy-viscosity given by

$$\nu_e^a = \frac{\Pi}{\nabla^2 \bar{\omega}} \tag{7.16}$$

where the right-hand side of the above equation may be calculated from DNS snapshots. The a-priori eddy-viscosity is centered at zero (corresponding to where closure modeling is unnecessary) and spreads out in the negative and positive directions (a hallmark of isotropic turbulence). We segregate this a-priori estimate of sub-grid effects into three categories as follows. The a-priori eddy-viscosities calculated from the DNS data are compared with a Gaussian distribution where values lying less than a distance of 1% of the standard-deviation from the mean (which is zero) are labeled as those requiring no dissipation (due to the low strength of the a-priori eddy-viscosity). For posterity, we label these points as $k = 1$. Positive values lying beyond this range are labeled as those requiring sub-grid dissipation and are labeled $k = 2$. Negative values less than 1% of the standard-deviation are also considered to require no dissipation and are labeled $k = 3$. This three-category segregation stems from a learning hypothesis that seeks to identify regions in a flow evolution that require structural, functional or no-closure modeling hypothesis. We link labels of negative or nearly-zero eddy-viscosities to the Arakawa classification and positive eddy-viscosities to the upwinded classification. The positive eddy-viscosity prediction would indicate that the sub-grid term at a point is predominantly dissipative in nature at which point the numerical dissipation of the upwinded scheme would be utilized. We note here that the concept of an a-priori eddy-viscosity lies firmly within the explicit LES hypothesis. The classifier is therefore instrumental in moderating ILES deployments through a decision making process that recognizes the dissipative (or forcing) nature of the sub-grid quantities.

We note that the choice of 1% as the decision parameter for switching between

hypothesis is motivated by a sensitivity study that showed the highest classification accuracy for the ANN framework. Larger choices of this hyper-parameter would result in a classifier that would be prone to classify most points in the 'no-model' zone. However, we clarify that the choice of this value is also correlated with the architecture of the ANN. A potential extension of the proposed hypothesis is to combine architecture search algorithms with varying value of the decision hyper-parameters for larger classification accuracies. In addition, the three-category framework is derived from an aligned study (the reader may refer to the previous chapter) where sub-grid models are determined according to negative, positive and nearly-zero a-priori eddy-viscosities and utilizes the same learning. This enables use to determine a unified framework for switching between turbulence model hypotheses as well as numerical dissipation scenarios. However, we would like to emphasize that, for the purpose of switching between the Arakawa and upwinded Jacobian computation, a simple two-class framework would also suffice.

A one-hot labeling of our eddy-viscosity classes is utilized for a classification deployment and a schematic for this hypothesis segregation and labeling is shown in Figure 7.2. The labels indicate the conditional probability of a point belonging to each possible class. As such, the training labels are given by a value of 1 for the particular class that a point belongs to and zeros for other choices. This is because there is no ambiguity in the class a training sample belongs to. Each label for the a-priori eddy-viscosity is also associated with a corresponding input kernel of grid-resolved quantities. This kernel is given by a local stencil of vorticity and stream function. There are 9 inputs each for vorticity and stream function given by a query of the field quantity at a point on the coarse grid, 4 adjacent points in each dimension (x, y) and the 4 diagonally adjacent points. Each sample of our training data thus consists of 18 inputs of vorticity and stream function and outputs given by one-hot labels for the choice of closure modeling strategy. We then utilize an ANN to establish a

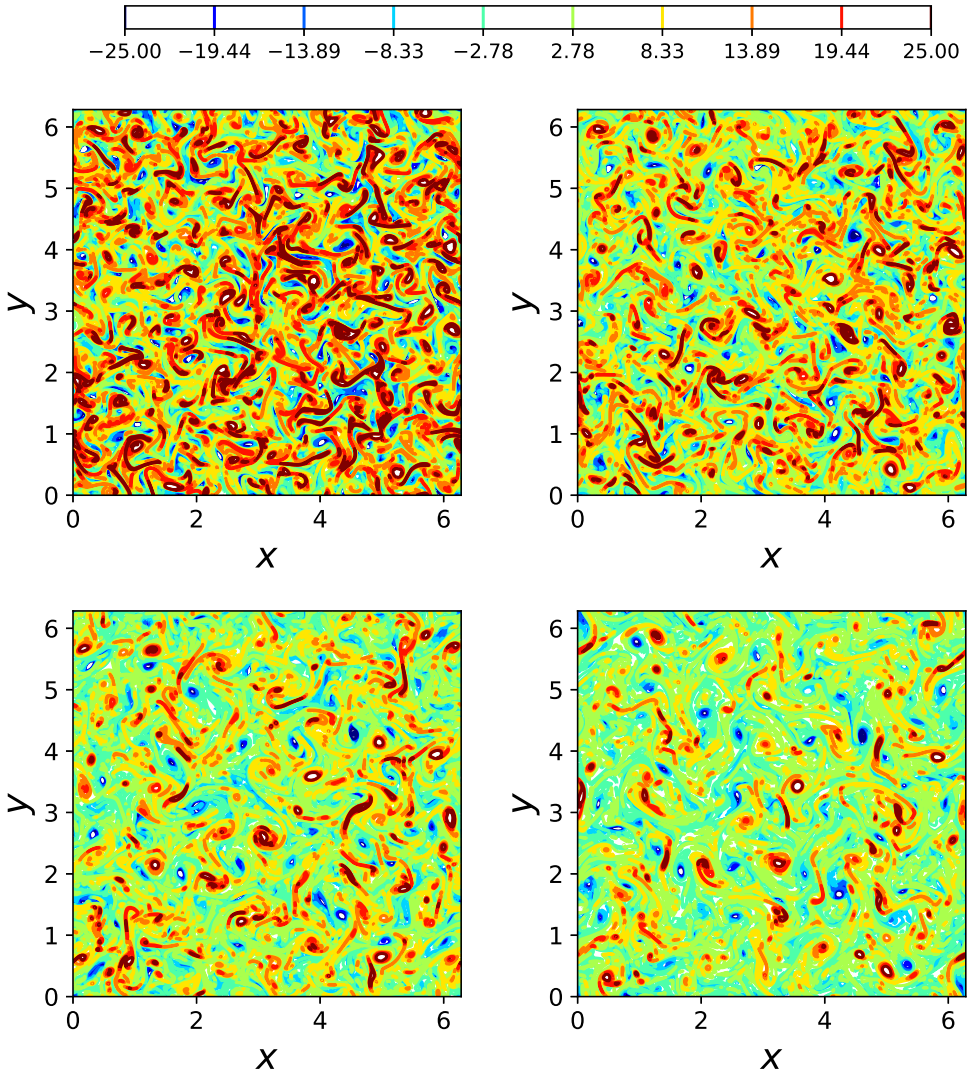


Figure 7.1: Time evolution of the Kraichnan turbulence case with DNS ($N^2 = 2048^2$) contours for vorticity of $t = 1$ (top-left), $t = 2$ (top-right), $t = 3$ (bottom-left), $t = 4$ (bottom-right). One can discern the dissipation of vorticity as the system evolves.

relationship between these inputs and outputs. Mathematically, if our input vector $\mathcal{P}$ resides in a P-dimensional space and our desired output $\mathcal{Q}$ resides in a Q-dimensional space, this framework establishes a map $\mathbb{M}$ as follows:

$$\mathbb{M} : \{\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_P\} \in \mathbb{R}^P \rightarrow \{\mathcal{Q}_1, \mathcal{Q}_2, \ldots, \mathcal{Q}_Q\} \in \mathbb{R}^Q. \tag{7.17}$$

Accordingly, the framework utilized in this article leads to the following relation:

$$\mathbb{M} : \{\mathbf{p}\} \in \mathbb{R}^{18} \rightarrow \{P(\mathbf{q}|\mathbf{p})\} \in \mathbb{R}^3, \tag{7.18}$$

where

$$\begin{aligned} \mathbf{p}_{i,j} = \{&\bar{\omega}_{i,j}, \bar{\omega}_{i,j+1}, \bar{\omega}_{i,j-1}, \ldots, \bar{\omega}_{i-1,j-1}, \\ &\bar{\psi}_{i,j}, \bar{\psi}_{i,j+1}, \bar{\psi}_{i,j-1}, \ldots, \bar{\psi}_{i-1,j-1}\} \end{aligned} \tag{7.19}$$

is our input vector for each query of the machine learning framework and where

$$P(\mathbf{q}|\mathbf{p})_{i,j} = \{P(J^k(\bar{\omega}, \bar{\psi})_{i,j}|\mathbf{p}_{i,j})\}, \tag{7.20}$$

is the conditional probability of a Jacobian computation (given by a connection to the explicit closure hypothesis). Note that i, j refer to the spatial indices on the coarse-grid (i.e., the point of deployment). The indices $k = 1$ and $k = 3$ refer to the Arakawa non-linear Jacobian computation and $k = 2$ refers to the upwinded computation instead (see Figure 7.2). Our optimal map $\mathbb{M}$ is then trained by minimizing the categorical cross-entropy loss-function

$$E(\mathbf{w}) = -\sum_{n=1}^{N}\sum_{k=1}^{K}\{t_{nk}\log(y_{nk}) + (1 - t_{nk})\log(1 - y_{nk})\}, \tag{7.21}$$

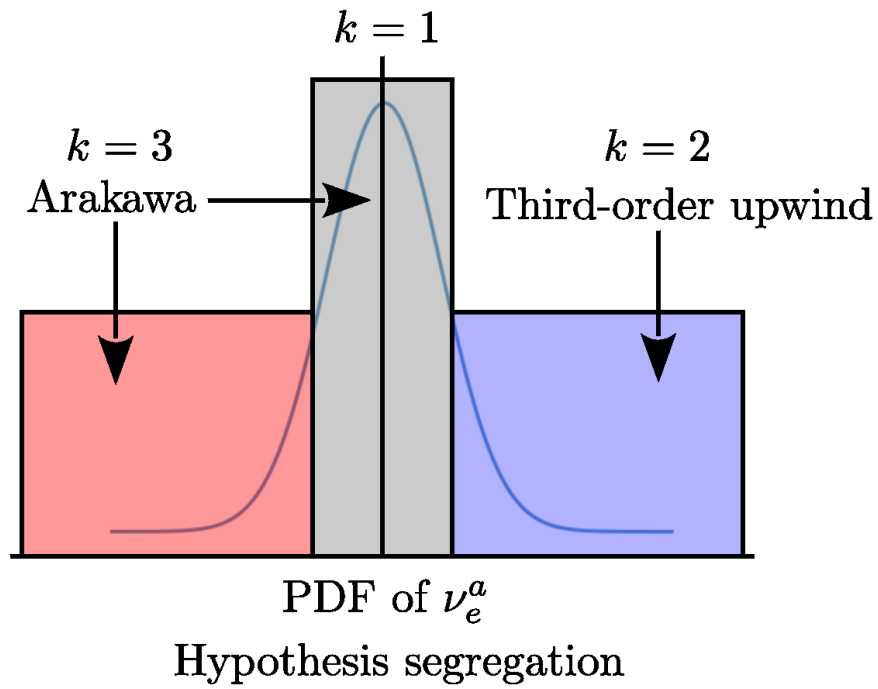
where $\mathbf{w}$ are the variable weight and bias parameters of the network, N refers to the

total number of samples and $K = 3$ is the total number of classification scenarios (i.e., negative, positive or nearly-zero a-priori eddy-viscosities). Here, t_{nk} refers to the true label of class k and sample n and y_{nk} refers to a corresponding prediction of the learning framework. One-hot encoding ensures that t_{nk} values are always binary (Bishop, 2006) and the outputs of the ANN may be interpreted as conditional-probabilities. Our optimal architecture is given by five 40-neuron hidden layers (obtained via grid-search hyper-parameter tuning). All hidden layers utilize ReLU units to impart non-linearity to the layer-wise transformations. For reference, our architecture is trained using the open-source deep learning software Tensorflow and is optimized with the use of ADAM, a popular gradient-descent based optimizer (Kingma and Ba, 2014). Figure 7.3 shows the progress to convergence for our framework with our optimally trained network displaying approximately 79% accuracy in classifying points to their correct labels. To summarize this section, we train a deep ANN to estimate probabilities of negative, positive or nearly-zero eddy-viscosities which are utilized to decide the choice of the Jacobian computation. We clarify that the decision to deploy a particular hypothesis is obtained by utilization of the classification scenario which has the highest conditional probability.

7.6 Results

7.6.1 *A posteriori* deployment

In this section, we detail the results from an a-posteriori deployment of the classification framework (denoted ML henceforth) for the Kraichnan test-case. In the LES evolution of the problem, a considerably coarser grid is used (at $N^2 = 256^2$). We remark that the forward deployment of our framework needs to overcome the challenge of numerical errors and is a robust test of the generalizability and robustness of our learning. Our LES results are assessed using angle-averaged kinetic energy spectra and through structure functions of vorticity. In addition, qualitative com-



One-hot labels

$$k = 1 \rightarrow [1 \quad 0 \quad 0]$$
$$k = 2 \rightarrow [0 \quad 1 \quad 0]$$
$$k = 3 \rightarrow [0 \quad 0 \quad 1]$$

Figure 7.2: Hypothesis segregation and one-hot labeling for our proposed framework. The learning predicts conditional probabilities for the three segregated a-priori eddy-viscosity classes which are utilized for Jacobian calculation decisions spatio-temporally.

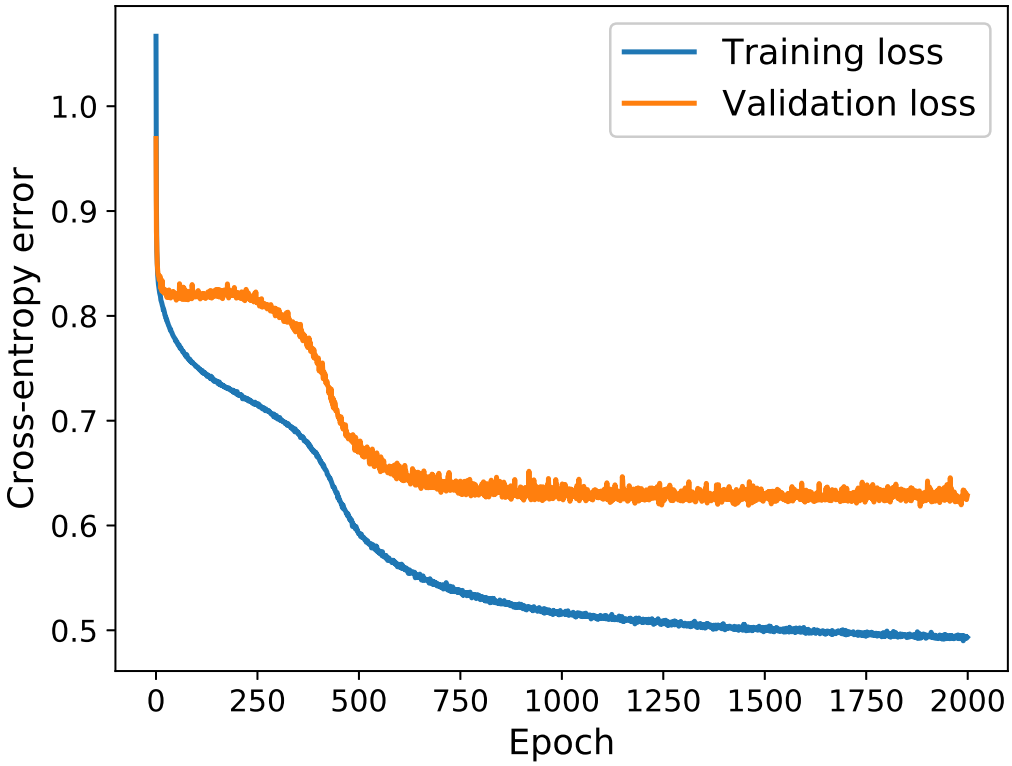


Figure 7.3: Learning rate and convergence of our classification framework training. 2000 epochs were sufficient for converged validation loss.

parisons are also provided through visual examinations of the vorticity contours. We remark that the LES deployment is performed from $t = 0$ to $t = 4$ which spans the training regime data obtained from DNS. In what follows we note that DNS refers to a high-fidelity evolution of the governing equations (i.e., at $N^2 = 2048^2$ degrees of freedom), UNS refers to results obtained using the Arakawa scheme alone and ILES refers to a simulation where the non-linear Jacobian at all points in space and time are upwinded. Figure 7.4 shows the a-posteriori performance of the proposed framework at $Re = 32000$ in terms of energy spectra predictions. The reader may find an exact definition of the kinetic-energy spectra in Maulik and San (2017*c*). We note that the training data was obtained for the same Reynolds number as well. The prediction of the proposed framework is seen to agree remarkably well with DNS. It is apparent that the switching of schemes using the classifier has obtained an optimal balance between both techniques.

Vorticity contours for LES resolution assessments are shown in Figure 7.5, where it is apparent that the proposed framework optimally balances the energy-conserving and dissipative natures of the Arakawa and upwinded schemes respectively. This is verified by qualitative examination with FDNS contours obtained by spectrally filtering the DNS snapshot for $Re = 32000$ at $t = 4$.

A second statistically significant quantity of interest studied in this investigation is the vorticity structure function (Grossmann and Mertens, 1992) given by

$$S_\omega^x = \langle |\omega(x + r, y) - \omega(x, y)|^2 \rangle \tag{7.22}$$

$$S_\omega^y = \langle |\omega(x, y + r) - \omega(x, y)|^2 \rangle, \tag{7.23}$$

where the angle-brackets indicate ensemble averaging and x, y indicate a position on the grid with r being a certain distance from this location. Figures 7.6 and 7.7 show the structure functions obtained from a-posteriori deployments of the UNS, ILES

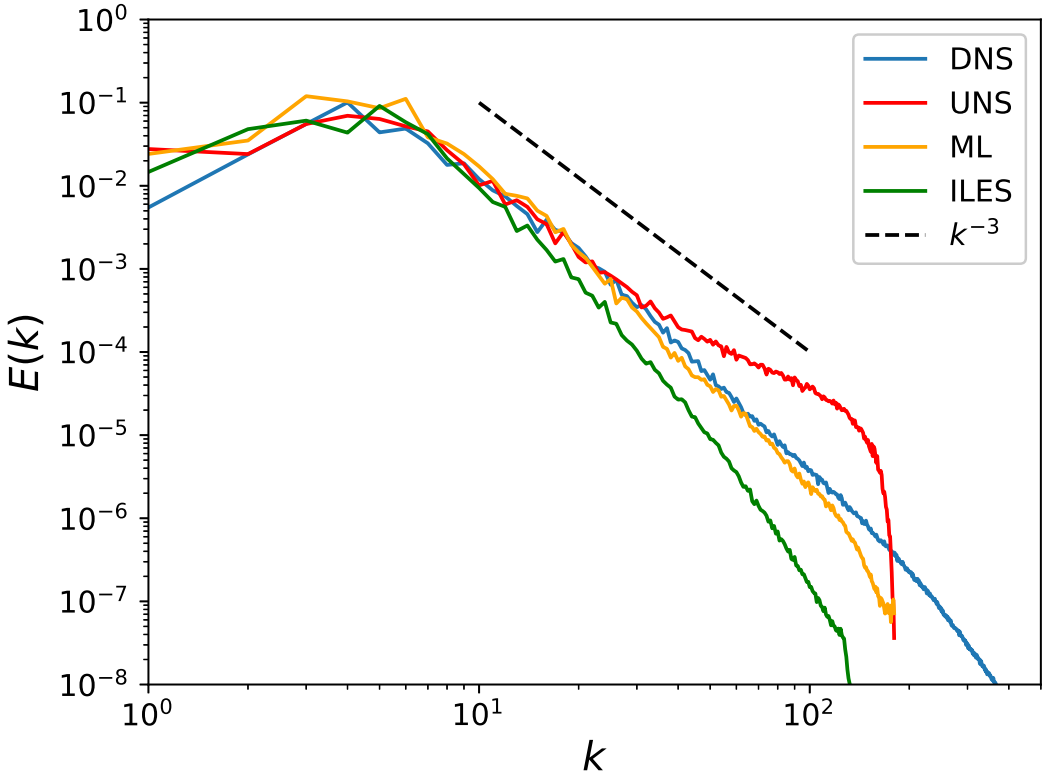


Figure 7.4: The a-posteriori performance of proposed framework (ML) for $Re = 32000$ and at $t = 4$ in terms of angle-averaged kinetic energy spectra. Comparisons with DNS, the Arakawa scheme (UNS) and the upwinded scheme (ILES) show that ML provides directed dissipation adequately.

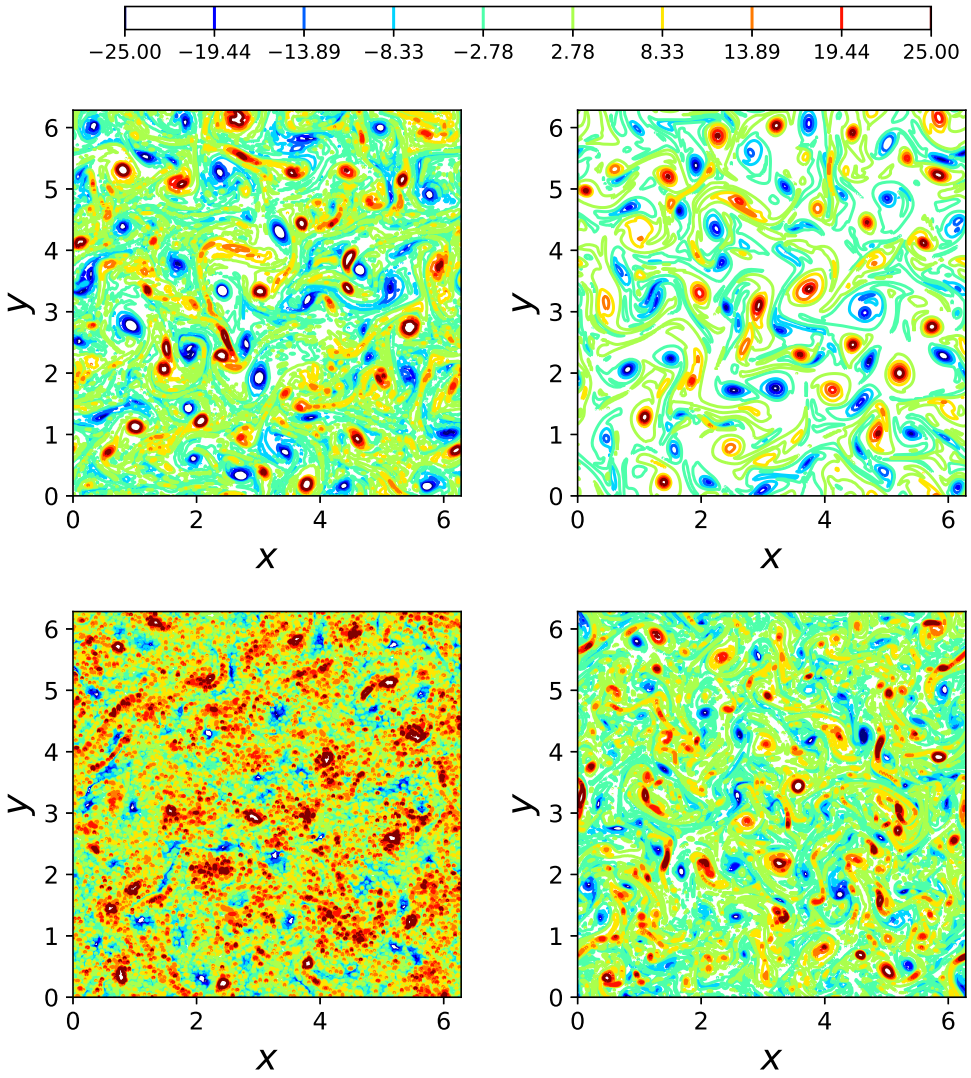


Figure 7.5: Contours for the vorticity at LES resolution and at $t = 4$. In the top-left, we have predictions from the ML approach. The top-right field has been obtained using ILES, the bottom-left field is obtained from UNS and the bottom right shows FDNS contours obtained by spectral cut-off filtering of DNS.

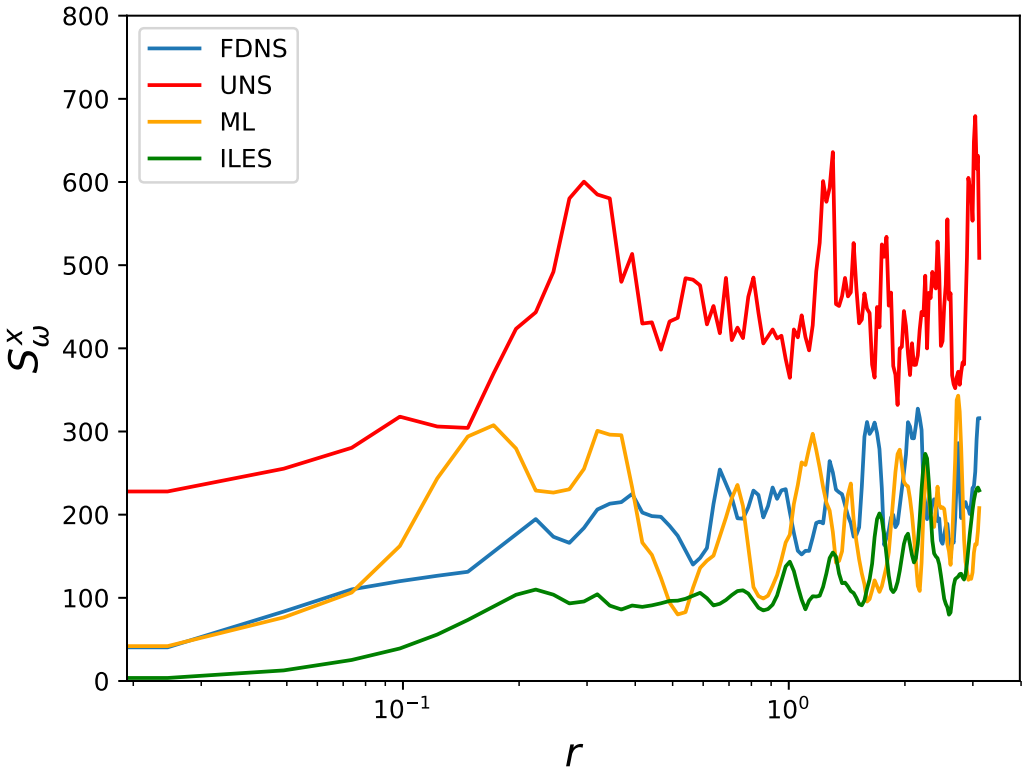


Figure 7.6: *A posteriori* vorticity structure functions in x direction of our proposed framework (ML), the Arakawa scheme (UNS) and the upwind scheme (ILES) with statistics obtained from an FDNS snapshot at $t = 4$. It is apparent that the ML method stabilizes the UNS result optimally.

and ML frameworks compared against those obtained from the final time FDNS snapshot. It is clear that the proposed framework balances between UNS and ILES deployments well to recover appropriate trends. We can thus claim that our learning is appropriate for hybrid deployments of dissipative and conservative frameworks for two-dimensional turbulence. Before moving on, we would like to point out to the reader here that the proposed methodology for closure does not require any post-processing prior to deployment in the forward simulation as utilized in several data-driven turbulence modeling studies (Beck et al., 2018; Maulik et al., 2019).

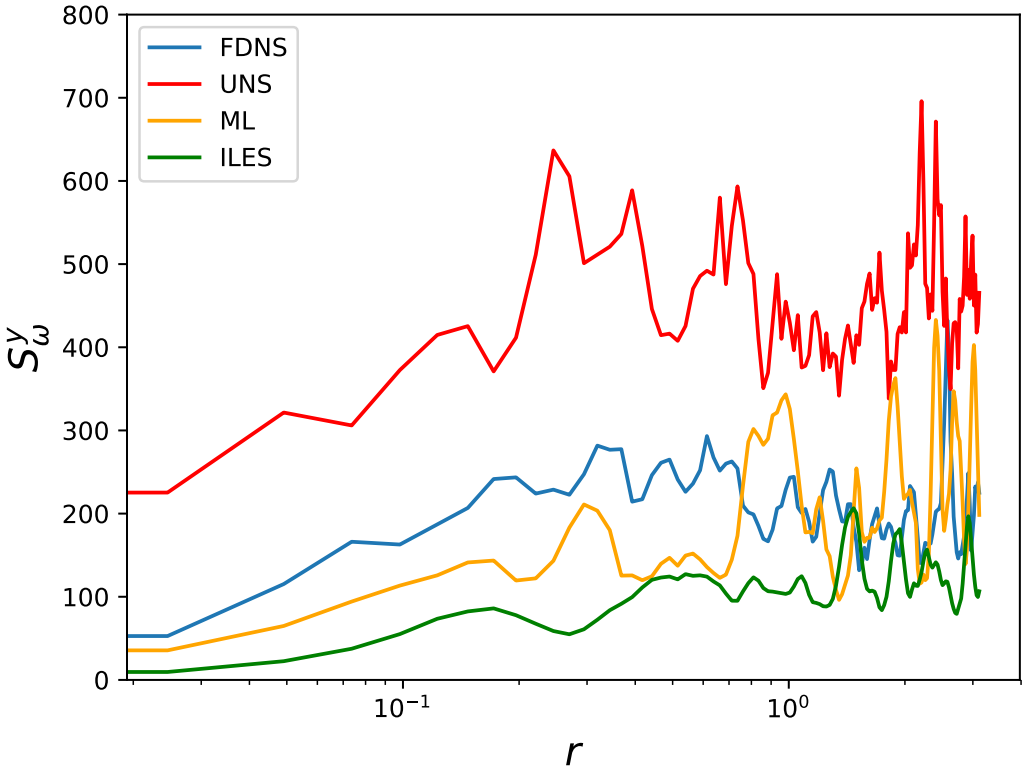


Figure 7.7: *A posteriori* vorticity structure functions in y direction of our proposed framework (ML), the Arakawa scheme (UNS) and the upwind scheme (ILES) with statistics obtained from an FDNS snapshot at $t = 4$. It is apparent that the ML method stabilizes the UNS result optimally.

7.6.2 Validation of learning

In this section, we proceed with a rigorous validation of our learning for deployment in regimes that are not a part of the training data. This is to ensure that the framework has truly learnt a classification based on the underlying physical hypothesis used for data segregation and is not memorizing data. This ensures that our classifier can be used in a more generalizable fashion. Figure 7.8 shows kinetic energy spectra obtained from the forward deployment of the ML framework for a $Re = 64000$ which represents a classification task that the framework has not previously seen (although the physics of the test-case remains similar). As observed, the proposed method performs quite well in this out-of-training data range as well. We note that a similar resolution ($N^2 = 256^2$) is utilized for this deployment. In contrast, Figure 7.9 shows the performance of the ML technique for a reduced resolution of $N^2 = 128^2$ but utilizing the same Reynolds number of 32000. The kinetic energy spectra show a successful stabilization of the flow evolution at this reduced resolution although some forcing to the large scales is observed. This suggests that the classification framework may be improved by sampling from different resolutions.

7.7 Concluding remarks

In this article, we have proposed a neural network based classifier that enables us to take decisions on the choice of non-linear term computation in the LES evolution of the Kraichnan turbulence test-case. The classifier outputs conditional probabilities for the presence (or absence) of eddy-viscosities within three different ranges during deployment and is used to switch between the Arakawa and upwind computation of the non-linear Jacobian for a hybrid upwinded deployment that optimally directs dissipation on the coarse-grained flow field. Our machine learning framework is trained by calculating a-priori eddy-viscosities which are projected onto a Gaussian distribution and segregated into three categories. Each category is devised to cap-

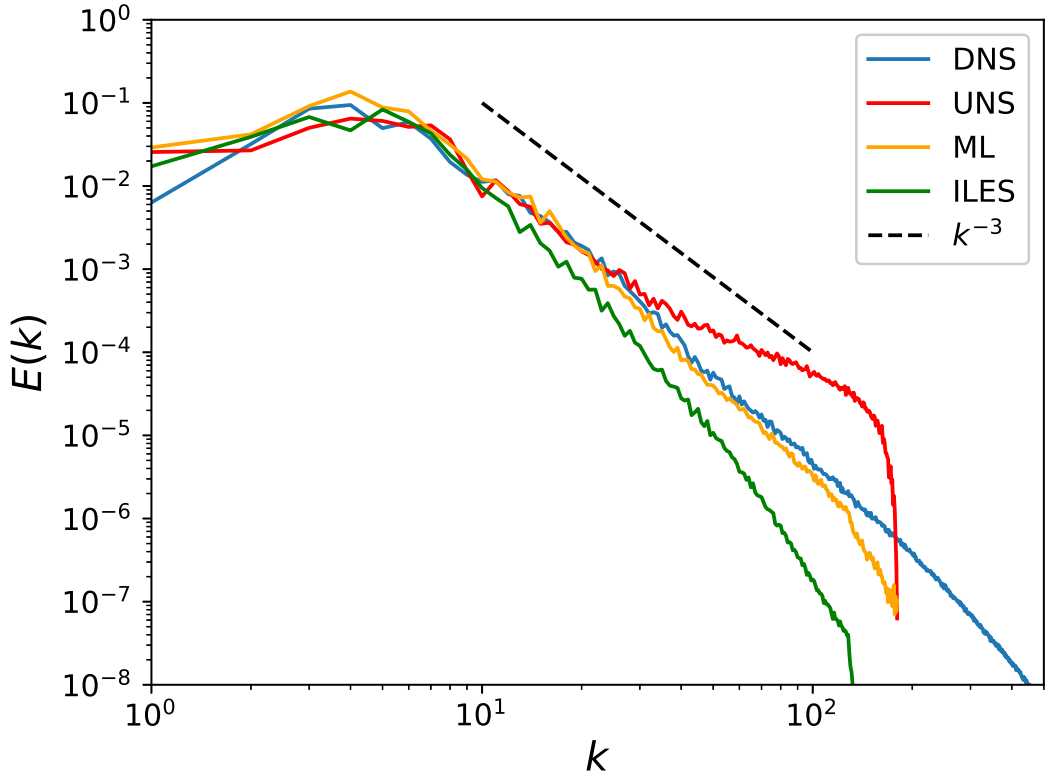


Figure 7.8: The a-posteriori performance of proposed framework (ML) for $Re = 64000$ and at $t = 4$ in terms of energy spectra. This represents deployment of our learning at a different Reynolds number than that used for generating training data.

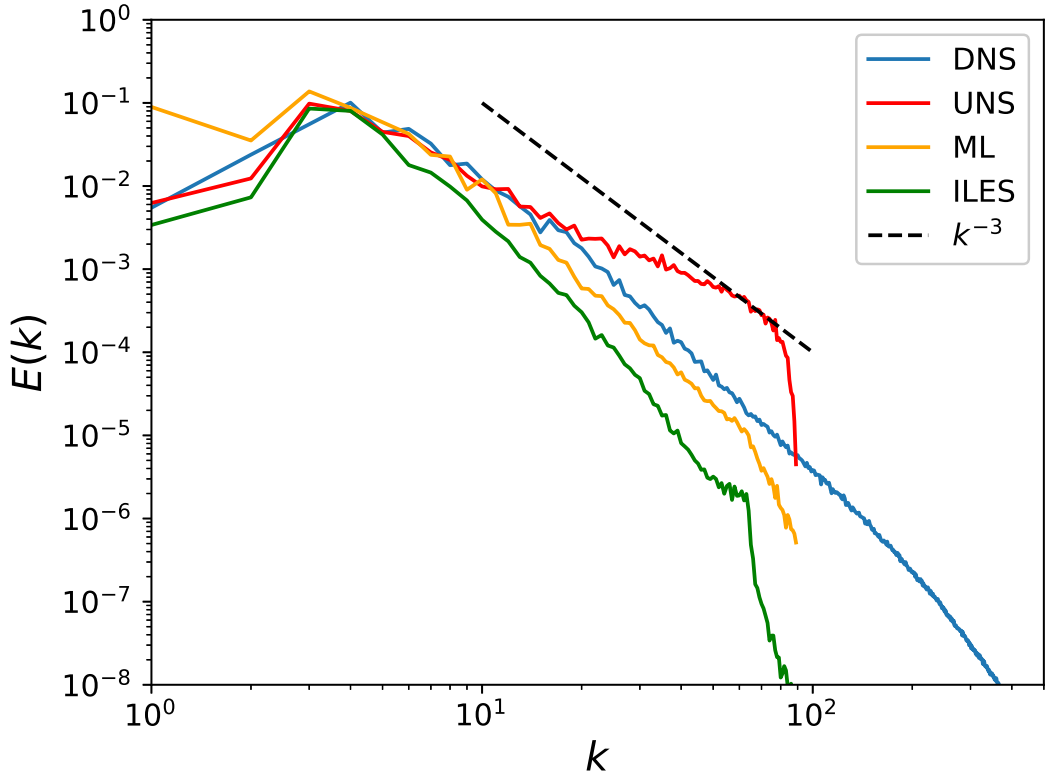


Figure 7.9: The a-posteriori performance of proposed framework (ML) for $Re = 32000$, $t = 4$ and at $N^2 = 128^2$ in terms of energy spectra. This represents deployment of our learning at a different resolution than that used for generating training data.

ture a unique behavior of the underlying sub-grid terms with negative and nearly-zero eddy-viscosity classes signifying absence of sub-grid dissipation. An optimally trained classifier is then utilized to identify if a point requires sub-grid dissipation based on if it is placed in the positive eddy-viscosity category. If so, the upwind Jacobian is calculated for imparting numerical dissipation.

We perform a-posteriori assessments on the Kraichnan turbulence test-case through statistical quantities such as the angle-averaged kinetic energy spectra and the vorticity structure functions. It is observed that the proposed framework is successful in balancing the dissipative nature of the upwind scheme and the energy-conserving Arakawa scheme to give excellent agreement with DNS statistics. Validation for out-of-training regimes also indicate that the framework is able to learn the link between grid-resolved quantities at a coarse resolution and the nature of the sub-grid forcing.

Our conclusions therefore point toward the possibility of using classifiers for the unified deployment of numerical schemes with varying dissipation through the decision making process described above. A key strength of our hypothesis stems from the fact that an ILES deployment is moderated by concepts drawn from the explicit LES ideology (i.e., that of an a-priori eddy-viscosity). The successful deployment of our method thus points towards the possibility of deploying directed numerical dissipation that preserves the statistics of turbulence without sacrificing the shock-capturing ability of many non-oscillatory schemes. Our future work lies in that particular direction.

CHAPTER VIII

Conclusions and future work

In this chapter we summarize the major facets of this study and outline some potential research that may emerge from our results.

8.1 Summary of study

We have developed and analyzed several data-driven strategies for the representation of the sub-grid forcing using various physics-informed machine learning strategies. This is motivated by our understanding of the limitations of phenomenology, which necessitates a-priori specification of model coefficients or low-pass spatial filter transfer functions which are generally correlated poorly with underlying sub-grid closure requirements. The various formulations we have explored in this subject suggest that certain learning problems can be formulated from sparsely sampled data through a hybridization with first-principles based knowledge of governing equations. Essentially, turbulence source terms may be predicted from grid-resolved data which are evolved according to the filter Navier-Stokes equations with associated benefits (in terms of the ability to choose slightly different initial conditions and physical regimes). In addition, we have also framed a machine learning methodology for enabling optimal hybridization of different closure techniques. Through this, we demonstrate that a learning may be utilized to switch between closure modeling hypothesis (or discretization techniques) for directed dissipation according to perfect closure requirements.

Some of our conclusions through the course of this research may be outlined in the following -

1. Data-driven closures are able to predict flow fields with statistical trends that mimic those obtained from high-fidelity simulations without the use of user-defined coefficients or filter transfer functions (when deployed for the purpose of calculating a perfect closure requirement). In addition, they are also capable of learning an optimal classifer to switch between dissipation scenarios and obtain DNS aligned results. This was the core motivation at the start of these investigations.

2. We also observe that network architectures may be simplified significantly if our understanding of the fundamental nature of turbulence is leveraged for instance through the use of the dual maps defining our convolution and deconvolution operators or the use of grid-resolved eddy-viscosity information. These simplified architectures result in reduced computational and memory costs.

3. A-posteriori deployment of machine learning predictions for the purpose of explicitly calculating the closure necessitates post-processing for stability. This limitation stems from the use of grid-filtering to synthesize the perfect LES data. While a perfect LES is generated through this procedure, it fails to account for numerical stability in this process. The loss of information due to this post-processing may restored (in aggregate fashion) through the use of averaging kernels as shown in chapter 3. The deployment of these strategies necessitates the understanding of the dichotomy of a-priori and a-posteriori statistical recovery. Optimal architecture selections (from traditional hyper-parameter assessment techniques in machine learning) must thus be paired with deployment-based checks on canonical problems such as Kraichnan or Kolmogorov turbulence before they may be utilized in unseen cases. This opens up another avenue for machine learning architectural analyses through canonical flow investigations.

4. One of the core advantages of formulating the data-driven closure problem as

a classification task is the obviation of post-processing requirements since the data-driven closure is spanned by a basis of stable modeling strategies which compete only through their underlying hypotheses.

5. Computational assessments for the models developed in this study indicate that there is a non-trivial increase in overall cost due to the additional floating point operations involved in neural network queries. As such, the next phase of these modeling strategies depends on successful scaling up to multiprocessing environments.

8.2 Future Work

The studies described in this document give rise to some interesting questions and may be built on for the following future research

1. The first step in extending these modeling platforms for practical use in the broader turbulence community is to assess their ability to *recognize* different classes of flow problems. While most of our studies here have utilized decaying isotropic turbulence, a data-driven framework that may adapt closure requirements according to the anisotropies it has been exposed to in a-priori holds immense potential for accurate practical deployments.

2. While the linear eddy-viscosity framework preserves time, pressure and Galilean invariance for three-dimensional closure modeling, the incorporation of translational and rotational invariance requires predictions that are spanned in a fixed space of grid-resolved quantities (generally functions of the strain-rate and rotation tensor). This would require a modified neural architecture (as against the simpler deep network described in chapter 5). In addition, the effect of modified loss-functions may also be studied to improve robustness in deployment and potentially remove the need for a-posteriori post-processing.

3. The concept of a model classifier, while successful in stable recovery for isotropic fields may be tested across physical scenarios. There is a potential that anisotropic turbulence modeling requirements may be identified and predicted by this classifier in an accurate manner.

4. As mentioned in the conclusions section, the successful deployment of the proposed models in a cost-effective fashion would require their porting to distributed environments to minimize overhead.

To conclude, this work has explored the feasibility of turbulence closure strategies being developed from data for the purpose of improved a-posteriori statistical fidelity in comparison with classical large eddy simulation closures which rely on a-priori coefficient tuning and phenomenology that may or may not be suited to the simulation requirements. Our results indicate that the techniques studied here hold promise for the development of machine learning closures that may be trained on different flow scenarios and then be deployed in cases that may be considered to be combinations of the aforementioned flow components. Our future studies will be aligned in that direction.

References

Albert, A. (1972), *Regression and the Moore-Penrose pseudoinverse*, Academic Press, New York.

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

Arakawa, A. and Lamb, V. R. (1981), 'A potential enstrophy and energy conserving scheme for the shallow water equations', *Mon. Weather Rev.* **109**(1), 18–36.

Bardina, J., Ferziger, J. H. and Reynolds, W. C. (1980), 'Improved subgrid-scale models for large-eddy simulation', *AIAA Paper 80-1357* .

Beck, A. D., Flad, D. G. and Munz, C.-D. (2018), 'Neural Networks for Data-Based Turbulence Models', *arXiv preprint arXiv:1806.04482* .

Berselli, L. C., Iliescu, T. and Layton, W. J. (2005), *Mathematics of large eddy simulation of turbulent flows*, Springer Science & Business Media, New York.

Berselli, L. C., Iliescu, T. and Layton, W. J. (2006), *Mathematics of large eddy simulation of turbulent flows*, Springer-Verlag, New York.

Bidadi, S. and Rani, S. L. (2015), 'Investigation of numerical viscosities and dissipation rates of second-order TVD-MUSCL schemes for implicit large-eddy simulation', *J. Comput. Phys.* **281**, 1003–1031.

Bishop, C. M. (2006), *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag, Berling Heidelberg.

Bos, W. J. and Bertoglio, J. (2006), 'Dynamics of spectrally truncated inviscid turbulence', *Phys. Fluids* **18**(7), 071701.

Bright, I., Lin, G. and 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. and Kutz, J. N. (2016), 'Discovering governing equations from data by sparse identification of nonlinear dynamical systems', *Proc. Natl. Acad. Sci. U.S.A.* **113**(15), 3932–3937.

Bull, J. R. and Jameson, A. (2015), 'Simulation of the Taylor-Green vortex using high-order flux reconstruction schemes', *AIAA J.* **53**, 2750–2761.

Canuto, V. and Cheng, Y. (1997), 'Determination of the Smagorinsky–Lilly constant CS', *Phys. Fluids* **9**(5), 1368–1378.

Carati, D., Winckelmans, G. S. and Jeanmart, H. (2001), 'On the modelling of the subgrid-scale and filtered-scale stress tensors in large-eddy simulation', *J. Fluid Mech.* **441**, 119–138.

Chow, F. K., Street, R. L., Xue, M. and Ferziger, J. H. (2005), 'Explicit filtering and reconstruction turbulence modeling for large-eddy simulation of neutral boundary layer flow', *J. Atmos. Sci.* **62**(7), 2058–2077.

Cichocki, A. and Amari, S. (2002), *Adaptive blind signal and image processing: learning algorithms and applications*, John Wiley & Sons, UK.

Cohen, K., Siegel, S., McLaughlin, T. and Gillies, E. (2003), 'Feedback control of a cylinder wake low-dimensional model', *AIAA J.* **41**(7), 1389–1391.

Cushman-Roisin, B. and Beckers, J.-M. (2011), *Introduction to geophysical fluid dynamics: physical and numerical aspects*, Vol. 101, Academic Press, Boston, MA.

Dabov, K., Foi, A., Katkovnik, V. and 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.

DeBonis, J. (2013), 'Solutions of the Taylor-Green vortex problem using high-resolution explicit finite difference methods', *AIAA Pap. 2013-0382* .

Demuth, H. B., Beale, M. H., De Jess, O. and Hagan, M. T. (2014), *Neural network design*, Martin Hagan, Stillwater.

Duraisamy, K., Iaccarino, G. and Xiao, H. (2019), 'Turbulence modeling in the age of data', *Annu. Rev. Fluid Mech.* **51**, 357–377.

Duraisamy, K., Zhang, Z. J. and Singh, A. P. (2015), 'New approaches in turbulence and transition modeling using data-driven techniques', *AIAA Paper 2015-1284* .

Eden, C. and Greatbatch, R. J. (2008), 'Towards a mesoscale eddy closure', *Ocean Model.* **20**(3), 223–239.

El Rafei, M., Könözsy, L. and Rana, Z. (2017), 'Investigation of Numerical Dissipation in Classical and Implicit Large Eddy Simulations', *Aerospace* **4**(4), 59.

Faller, W. E. and Schreck, S. J. (1997), 'Unsteady fluid mechanics applications of neural networks', *J. Aircraft* **34**(1), 48–55.

Foresee, F. D. and Hagan, M. T. (1997), Gauss-newton approximation to Bayesian learning, *in* 'IEEE International Conference on Neural Networks', pp. 1930–1935.

Fox-Kemper, B., Danabasoglu, G., Ferrari, R., Griffies, S., Hallberg, R., Holland, M., Maltrud, M., Peacock, S. and Samuels, B. (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. and Zidikheri, M. J. (2013), 'Subgrid modelling for geophysical flows', *Phil. Trans. R. Soc. A* **371**(1982), 20120166.

Frederiksen, J. S. and Zidikheri, M. J. (2016), 'Theoretical comparison of subgrid turbulence in atmospheric and oceanic quasi-geostrophic models', *Nonlinear Proc. Geoph.* **23**(2), 95.

Frisch, U. (1995), *Turbulence*, Cambridge University Press, Cambridge, U.K.

Frisch, U. (1996), *Turbulence: the legacy of A. N. Kolmogorov*, Cambridge University Press, UK.

Fröhlich, J. and Von Terzi, D. (2008), 'Hybrid LES/RANS methods for the simulation of turbulent flows', *Prog. Aerosp. Sci.* **44**(5), 349–377.

Fukami, K., Fukagata, K. and Taira, K. (2018), 'Super-resolution reconstruction of turbulent flows with machine learning', *arXiv preprint arXiv:1811.11328* .

Galperin, B. and Orszag, S. A. (1993), *Large eddy simulation of complex engineering and geophysical flows*, Cambridge University Press.

Gamahara, M. and 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. and 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.

Germano, M., Piomelli, U., Moin, P. and Cabot, W. H. (1991), 'A dynamic subgrid-scale eddy viscosity model', *Phys. Fluids* **3**(7), 1760–1765.

Ghosal, S., Lund, T. S., Moin, P. and Akselvoll, K. (1995), 'A dynamic localization model for large-eddy simulation of turbulent flows', *J. Fluid Mech.* **286**, 229–255.

Gottlieb, S. and Shu, C.-W. (1998), 'Total variation diminishing runge-kutta schemes', *Math. Comput.* **67**(221), 73–85.

Gottlieb, S., Shu, C.-W. and Tadmor, E. (2001), 'Strong stability-preserving high-order time discretization methods', *SIAM Rev.* **43**(1), 89–112.

Grinstein, F. F., Margolin, L. G. and Rider, W. J. (2007), *Implicit large eddy simulation: computing turbulent fluid dynamics*, Cambridge University Press, Cambridge, UK.

Grossmann, S. and Mertens, P. (1992), 'Structure functions in two-dimensional turbulence', *Z. Phys. B. Con. Mat.* **88**(1), 105–116.

Guermond, J.-L., Oden, J. T. and Prudhomme, S. (2004), 'Mathematical perspectives on large eddy simulation models for turbulent flows', *J. Math. Fluid Mech.* **6**(2), 194–248.

Guest, D., Cranmer, K. and Whiteson, D. (2018), 'Deep Learning and its Application to LHC Physics', *Annual Rev. Nucl. Part. S.* **68**, 161–181.

Habisreutinger, M. A., Bouffanais, R., Leriche, E. and Deville, M. O. (2007), 'A coupled approximate deconvolution and dynamic mixed scale model for large-eddy simulation', *J. Comp. Phys.* **224**(1), 241–266.

Hanjalic, K. (2005), 'Will RANS survive LES? A view of perspectives', *J. Fluids Eng.* **127**(5), 831–839.

Hennigh, O. (2017), 'Lat-Net: Compressing Lattice Boltzmann Flow Simulations using Deep Neural Networks', *arXiv preprint arXiv:1705.09036* .

Hickel, S., Adams, N. A. and Domaradzki, J. A. (2006), 'An adaptive local deconvolution method for implicit les', *Journal of Computational Physics* **213**(1), 413–436.

Hickel, S., Egerer, C. P. and Larsson, J. (2014), 'Subgrid-scale modeling for implicit large eddy simulation of compressible flows and shock-turbulence interaction', *Phys. Fluids* **26**(10), 106101.

Hornik, K., Stinchcombe, M. and White, H. (1989), 'Multilayer feedforward networks are universal approximators', *Neural Netw.* **2**(5), 359–366.

Hu, X., Wang, Q. and Adams, N. A. (2010), 'An adaptive central-upwind weighted essentially non-oscillatory scheme', *J. Comput. Phys.* **229**(23), 8952–8965.

Huang, G., Zhu, Q. and 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. and Siew, C. (2006), 'Extreme learning machine: theory and applications', *Neurocomput.* **70**(1), 489–501.

Hyman, J. M., Knapp, R. J. and Scovel, J. C. (1992), 'High order finite volume approximations of differential operators on nonuniform grids', *Physica D: Nonlinear Phenomena* **60**(1-4), 112–138.

Jin, X., Cheng, P., Chen, W.-L. and Li, H. (2018), 'Prediction model of velocity field around circular cylinder over various reynolds numbers by fusion convolutional neural networks based on pressure on the cylinder', *Phys. Fluids* **30**(4), 047105.

Kara, K., Balakumar, P. and Kandil, O. A. (2011), 'Effects of nose bluntness on hypersonic boundary-layer receptivity and stability over cones', *AIAA J.* **49**(12), 2593–2606.

Kim, B., Azevedo, V. C., Thuerey, N., Kim, T., Gross, M. and Solenthaler, B. (2018), 'Deep Fluids: A Generative Network for Parameterized Fluid Simulations', *arXiv preprint arXiv:1806.02071* .

King, R., Hennigh, O., Mohan, A. and Chertkov, M. (2018), 'From Deep to Physics-Informed Learning of Turbulence: Diagnostics', *arXiv preprint arXiv:1810.07785* .

King, R. N., Hamlington, P. E. and Dahm, W. J. A. (2016), 'Autonomic closure for turbulence simulations', *Phys. Rev. E* **93**(3), 031301.

Kingma, D. P. and Ba, J. (2014), 'Adam: A method for stochastic optimization', *arXiv preprint arXiv:1412.6980* .

Kolmogorov, A. N. (1941*a*), Dissipation of energy in locally isotropic turbulence, *in* 'Dokl. Akad. Nauk SSSR', Vol. 32, pp. 16–18.

Kolmogorov, A. N. (1941*b*), The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers, *in* 'Dokl. Akad. Nauk SSSR', Vol. 30, JSTOR, pp. 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.

LaBryer, A., Attar, P. and Vedula, P. (2015), 'A framework for large eddy simulation of burgers turbulence based upon spatial and temporal statistical information', *Phys. Fluids* **27**(3), 035116.

Laney, C. B. (1998), *Computational gasdynamics*, Cambridge University Press, Cambridge.

Langford, J. A. and Moser, R. D. (1999), 'Optimal LES formulations for isotropic turbulence', *J. Fluid Mech.* **398**, 321–346.

Layton, W. J. and Rebholz, L. G. (2012), *Approximate deconvolution models of turbulence: analysis, phenomenology and numerical analysis*, Springer, New York.

Layton, W. and Lewandowski, R. (2003), 'A simple and stable scale-similarity model for large eddy simulation: energy balance and existence of weak solutions', *Appl. Math. Lett.* **16**(8), 1205–1209.

Lee, C., Kim, J., Babcock, D. and Goodman, R. (1997), 'Application of neural networks to turbulence control for drag reduction', *Phys. Fluids* **9**(6), 1740–1747.

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

Lesieur, M. and Metais, O. (1996), 'New trends in large-eddy simulations of turbulence', *Annu. Rev. Fluid Mech.* **28**(1), 45–82.

Liaw, A. and Wiener, M. (2002), 'Classification and regression by randomForest', *R News* **2**(3), 18–22.

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

Ling, J., Jones, R. and Templeton, J. (2016), 'Machine learning strategies for systems with invariance properties', *J. Comput. Phys.* **318**, 22–35.

Ling, J. and Kurzawski, A. (2017), 'Data-driven Adaptive Physics Modeling for Turbulence Simulations', *AIAA Pap. 2017-3627* .

Ling, J., Kurzawski, A. and Templeton, J. (2016), 'Reynolds averaged turbulence modelling using deep neural networks with embedded invariance', *J. Fluid Mech.* **807**, 155–166.

Ling, J. and Templeton, J. (2015), 'Evaluation of machine learning algorithms for prediction of regions of high Reynolds averaged Navier Stokes uncertainty', *Phys. Fluids* **27**(8), 085103.

Ma, M., Lu, J. and Tryggvason, G. (2015), 'Using statistical learning to close two-fluid multiphase flow equations for a simple bubbly system', *Phys. Fluids* **27**(9), 092101.

MacKay, D. J. C. (1992), 'Bayesian interpolation', *Neural Comput.* **4**(3), 415–447.

Mannarino, A. and Mantegazza, P. (2014), 'Nonlinear aeroelastic reduced order modeling by recurrent neural networks', *J. Fluid. Struct.* **48**, 103–121.

Mansfield, J. R., Knio, O. M. and Meneveau, C. (1998), 'A dynamic LES scheme for the vorticity transport equation: formulation and a priori tests', *J. Comp. Phys.* **145**(2), 693–730.

Margolin, L. G. (2018), 'The reality of artificial viscosity', *Shock Waves* pp. 1–9.

Marshall, J. S. and Beninati, M. L. (2003), 'Analysis of subgrid-scale torque for large-eddy simulation of turbulence', *AIAA J.* **41**(10), 1875–1881.

Mathew, J., Lechner, R., Foysi, H., Sesterhenn, J. and Friedrich, R. (2003), 'An explicit filtering method for large eddy simulation of compressible flows', *Phys. Fluids* **15**(8), 2279–2289.

Maulik, R. and San, O. (2017*a*), 'A neural network approach for the blind deconvolution of turbulent flows', *J. Fluid Mech.* **831**, 151–181.

Maulik, R. and San, O. (2017*b*), 'Resolution and Energy Dissipation Characteristics of Implicit LES and Explicit Filtering Models for Compressible Turbulence', *Fluids* **2**(2), 14.

Maulik, R. and San, O. (2017*c*), 'A stable and scale-aware dynamic modeling framework for subgrid-scale parameterizations of two-dimensional turbulence', *Comput. Fluids* **158**, 11–38.

Maulik, R. and San, O. (2018), 'Explicit and implicit LES closures for Burgers turbulence', *J. Comput. Appl. Math.* **327**, 12–40.

Maulik, R., San, O., Rasheed, A. and Vedula, P. (2018), 'Data-driven deconvolution for large eddy simulations of Kraichnan turbulence', *Phys. Fluids* **30**(12), 125109.

Maulik, R., San, O., Rasheed, A. and Vedula, P. (2019), 'Subgrid modelling for two-dimensional turbulence using neural networks', *J. Fluid Mech.* **858**, 122–144.

Milano, M. and Koumoutsakos, P. (2002), 'Neural network modeling for near wall turbulent flow', *J. Comput. Phys.* **182**(1), 1–26.

Mitchell, T., Buchanan, B., DeJong, G., Dietterich, T., Rosenbloom, P. and Waibel, A. (1990), 'Machine Learning', *Annual Rev. Comput. Science* **4**(1), 417–433.

Mohan, A. T. and Gaitonde, D. V. (2018), 'A Deep Learning based Approach to Reduced Order Modeling for Turbulent Flow Control using LSTM Neural Networks', *arXiv preprint arXiv:1804.09269* .

Moser, R. D., Malaya, N. P., Chang, H., Zandonade, P. S., Vedula, P., Bhattacharya, A. and Haselbacher, A. (2009), 'Theoretically based optimal large-eddy simulation', *Phys. Fluids* **21**(10), 105104.

Muravleva, E., Oseledets, I. and Koroteev, D. (2018), 'Application of machine learning to viscoplastic flow modeling', *Phys. Fluids* **30**(10), 103102.

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

Orlandi, P. (2012), *Fluid flow phenomena: a numerical toolkit*, Vol. 55, Springer Science & Business Media.

Orszag, S. A. (1970), 'Analytical theories of turbulence', *J. Fluid Mech.* **41**(2), 363–386.

Parish, E. J. and 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. and Ott, E. (2018), 'Hybrid forecasting of chaotic processes: using machine learning in conjunction with a knowledge-based model', *Chaos* **28**(4), 041101.

Pearson, B. and Fox-Kemper, B. (2018), 'Log-Normal Turbulence Dissipation in Global Ocean Models', *Phys. Rev. Lett.* **120**(9), 094501.

Pearson, B., Fox-Kemper, B., Bachman, S. and Bryan, F. (2017), 'Evaluation of scale-aware subgrid mesoscale eddy models in a global eddy-rich model', *Ocean Model.* **115**, 42–58.

Piomelli, U., Cabot, W. H., Moin, P. and Lee, S. (1991), 'Subgrid-scale backscatter in turbulent and transitional flows', *Phys. Fluids* **3**(7), 1766–1771.

Pope, S. B. (2004), 'Ten questions concerning the large-eddy simulation of turbulent flows', *New J. Phys.* **6**(1), 35.

Press, W., Teukolsky, S., Vetterling, W. and Flannery, B. (1992), *Numerical recipes in FORTRAN: The art of scientific computing*, Cambridge University Press, New York.

Rahman, S., San, O. et al. (2018), 'A localized dynamic closure model for Euler turbulence', *arXiv preprint arXiv:1810.01441* .

Raissi, M. and Karniadakis, G. E. (2016), 'Deep multi-fidelity gaussian processes', *arXiv:1604.07484* .

Raissi, M. and Karniadakis, G. E. (2017), 'Hidden physics models: Machine learning of nonlinear partial differential equations', *arXiv:1708.00588* .

Raissi, M. and Karniadakis, G. E. (2018), 'Hidden physics models: Machine learning of nonlinear partial differential equations', *J. Comp. Phys.* **357**, 125–141.

Raissi, M., Perdikaris, P. and Karniadakis, G. E. (2017), 'Machine learning of linear differential equations using gaussian processes', *J. Comp. Phys.* **348**, 683–693.

Razafindralandy, D., Hamdouni, A. and Oberlack, M. (2007), 'Analysis and development of subgrid turbulence models preserving the symmetry properties of the Navier–Stokes equations', *Eur. J. Mech. B-Fluid.* **26**(4), 531–550.

Rudy, S. H., Kutz, J. N. and Brunton, S. L. (2018), 'Deep learning of dynamics and signal-noise decomposition with time-stepping constraints', *arXiv preprint arXiv:1808.02578* .

Safavian, S. R. and Landgrebe, D. (1991), 'A survey of decision tree classifier methodology', *IEEE Trans. Syst. Man. Cybern.* **21**(3), 660–674.

Sagaut, P. (2006), *Large eddy simulation for incompressible flows: an introduction*, Springer-Verlag Berlin Heidelberg.

San, O. (2014), 'A dynamic eddy-viscosity closure model for large eddy simulations of two-dimensional decaying turbulence', *Int. J. Comput. Fluid. D.* **28**(6-10), 363–382.

San, O. (2015), 'A novel high-order accurate compact stencil Poisson solver: application to cavity flows', *Int. J. Appl. Mech.* **7**(01), 1550006.

San, O. and Maulik, R. (2018), 'Neural network closures for nonlinear model order reduction', *Adv. Comput. Math* pp. 1–34.

San, O. and Staples, A. E. (2012), 'High-order methods for decaying two-dimensional homogeneous isotropic turbulence', *Comput. Fluids* **63**, 105–127.

San, O., Staples, A. E. and Iliescu, T. (2013), 'Approximate deconvolution large eddy simulation of a stratified two-layer quasigeostrophic ocean model', *Ocean Model.* **63**, 1–20.

San, O., Staples, A. E. and Iliescu, T. (2015), 'A posteriori analysis of low-pass spatial filters for approximate deconvolution large eddy simulations of homogeneous incompressible flows', *Int. J. Comput Fluid D.* **29**(1), 40–66.

San, O. and Vedula, P. (2018), 'Generalized deconvolution procedure for structural modeling of turbulence', *J. Sci. Comput.* **75**(2), 1187–1206.

Sarghini, F., De Felice, G. and 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. A* **473**(2197), 20160446.

Schmidt, M. and Lipson, H. (2009), 'Distilling free-form natural laws from experimental data', *Science* **324**(5923), 81–85.

Schneiderbauer, S. and Saeedipour, M. (2018), 'Approximate deconvolution model for the simulation of turbulent gas-solid flows: An a priori analysis', *Phys. Fluids* **30**(2), 023301.

Serre, D. (2002), *Matrices: Theory and Applications*, Springer-Verlag, New York.

Shen, D., Wu, G. and Suk, H.-I. (2017), 'Deep Learning in Medical Image Analysis', *Annual Rev. Biomed. Eng.* **19**(1), 221–248.

Shu, C.-W. (2009), 'High order weighted essentially nonoscillatory schemes for convection dominated problems', *SIAM Review* **51**(1), 82–126.

Shu, C.-W. and Osher, S. (1988), 'Efficient implementation of essentially non-oscillatory shock-capturing schemes', *J. Comput. Phys.* **77**(2), 439–471.

Silvis, M. H., Remmerswaal, R. A. and Verstappen, R. (2017), 'Physical consistency of subgrid-scale models for large-eddy simulation of incompressible turbulent flows', *Phys. Fluids* **29**(1), 015105.

Singh, A. P. and Duraisamy, K. (2016), 'Using field inversion to quantify functional errors in turbulence closures', *Phys. Fluids* **28**(4), 045110.

Singh, A. P., Medida, S. and Duraisamy, K. (2017), 'Machine-learning-augmented predictive modeling of turbulent separated flows over airfoils', *AIAA J.* pp. 2215–2227.

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

Sotgiu, C., Weigand, B. and Semmler, K. (2018), 'A turbulent heat flux prediction framework based on tensor representation theory and machine learning', *Int. Commun. Heat Mass* **95**, 74–79.

Spalart, P. R. (1997), Comments on the feasibility of LES for wings, and on a hybrid RANS/LES approach, *in* 'Proceedings of first AFOSR international conference on DNS/LES', Greyden Press.

Stolz, S. and Adams, N. A. (1999), 'An approximate deconvolution procedure for large-eddy simulation', *Phys. Fluids* **11**(7), 1699–1701.

Stolz, S., Adams, N. A. and 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. and 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.

Thornber, B., Mosedale, A. and Drikakis, D. (2007), 'On the implicit large eddy simulations of homogeneous decaying turbulence', *J. Comput. Phys.* **226**(2), 1902–1929.

Tracey, B. D., Duraisamy, K. and Alonso, J. J. (2015), A machine learning strategy to assist turbulence model development, *in* '53rd AIAA Aerospace Sciences Meeting', p. 1287.

Tracey, B., Duraisamy, K. and Alonso, J. (2013), 'Application of supervised learning to quantify uncertainties in turbulence and combustion modeling', *AIAA Paper 2013-0259* .

Tsukerman, I. (2006), 'A class of difference schemes with flexible local approximation', *J. Comput. Phys.* **211**(2), 659–699.

Van Rees, W. M., Leonard, A., Pullin, D. and Koumoutsakos, P. (2011), 'A comparison of vortex and pseudo-spectral methods for the simulation of periodic vortical flows at high Reynolds numbers', *J. Comput. Phys.* **230**(8), 2794–2805.

Vollant, A., Balarac, G. and Corre, C. (2017), 'Subgrid-scale scalar flux modelling based on optimal estimation theory and machine-learning procedures', *J. Turbul.* **18**(9), 854–878.

Vorobev, A. and Zikanov, O. (2008), 'Smagorinsky constant in LES modeling of anisotropic MHD turbulence', *Theor. Comput. Fluid Dyn.* **22**(3-4), 317–325.

Vreman, A. (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. and Sapsis, T. (2018), 'Data-assisted reduced-order modeling of extreme events in complex dynamical systems', *PloS one* **13**(5), e0197704.

Wang, J., Wu, J. and Xiao, H. (2016), 'Physics-informed machine learning for predictive turbulence modeling: Using data to improve rans modeled Reynolds stresses', *arXiv:1606.07987* .

Wang, J., Wu, J. and 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, J.-X., Wu, J., Ling, J., Iaccarino, G. and Xiao, H. (2017), 'A Comprehensive Physics-Informed Machine Learning Framework for Predictive Turbulence Modeling', *arXiv preprint arXiv:1701.07102* .

Wang, L., Huang, Y., Luo, X., Wang, Z. and Luo, S. (2011), 'Image deblurring with filters learned by extreme learning machine', *Neurocomput.* **74**(16), 2464–2474.

Wang, Y. and Zhang, J. (2009), 'Sixth order compact scheme combined with multigrid method and extrapolation technique for 2d poisson equation', *J. Comput. Phys.* **228**(1), 137–146.

Wang, Z., Luo, K., Li, D., Tan, J. and Fan, J. (2018), 'Investigations of data-driven closure for subgrid-scale stress in large-eddy simulation', *Phys. Fluids* **30**(12), 125101.

Weatheritt, J. and Sandberg, R. (2016), 'A novel evolutionary algorithm applied to algebraic modifications of the rans stress–strain relationship', *J. Comput. Phys.* **325**, 22–37.

Weatheritt, J. and Sandberg, R. (2017*a*), 'The development of algebraic stress models using a novel evolutionary algorithm', *Int. J. Heat Fluid Fl.* **68**, 298–318.

Weatheritt, J. and Sandberg, R. D. (2017*b*), 'Hybrid Reynolds-Averaged/Large-Eddy Simulation Methodology from Symbolic Regression: Formulation and Application', *AIAA J.* pp. 3734–3746.

Wu, J.-L., Xiao, H. and Paterson, E. (2018*a*), 'Data-Driven Augmentation of Turbulence Models with Physics-Informed Machine Learning', *arXiv preprint arXiv:1801.02762* .

Wu, J.-L., Xiao, H. and Paterson, E. (2018*b*), 'Physics-informed machine learning approach for augmenting turbulence models: A comprehensive framework', *Phys. Rev. Fluids* **3**(7), 074602.

Xiao, D., Fang, F., Pain, C. and Hu, G. (2015), 'Non-intrusive reduced-order modelling of the Navier–Stokes equations based on RBF interpolation', *Int. J. Numer. Meth. Fl.* **79**(11), 580–595.

Xiao, H., Wu, J.-L., Wang, J.-X., Sun, R. and Roy, C. (2016), 'Quantifying and reducing model-form uncertainties in Reynolds-averaged Navier–Stokes simulations: A data-driven, physics-informed Bayesian approach', *J. Comp. Phys.* **324**, 115–136.

Yu, C., Xiao, Z. and Li, X. (2016), 'Dynamic optimization methodology based on subgrid-scale dissipation for large eddy simulation', *Phys. Fluids* **28**(1), 015113.

Zhang, G., Patuwo, B. E. and Hu, M. Y. (1998), 'Forecasting with artificial neural networks: The state of the art', *Int. J. Forecast.* **14**(1), 35–62.

Zhang, Z. J. and Duraisamy, K. (2015), 'Machine learning methods for data-driven turbulence modeling', *AIAA Paper 2015-2460* .

Zhang, Z., Song, X.-d., Ye, S.-r., Wang, Y.-w., Huang, C.-g., An, Y.-r. and Chen, Y.-s. (2019), 'Application of deep learning method to reynolds stress models of channel flow based on reduced-order modeling of dns data', *J. Hydrodyn.* **31**, 1–8.

Zhou, Y., Grinstein, F. F., Wachtor, A. J. and Haines, B. M. (2014), 'Estimating the effective Reynolds number in implicit large-eddy simulation', *Phys. Rev. E* **89**(1), 013303.

Zhou, Z., Wang, S. and Jin, G. (2018), 'A structural subgrid-scale model for relative dispersion in large-eddy simulation of isotropic turbulent flows by coupling kinematic simulation with approximate deconvolution method', *Phys. Fluids* **30**(10), 105110.

Zhu, L., Zhang, W., Kou, J. and Liu, Y. (2019), 'Machine learning methods for turbulence modeling in subsonic flows around airfoils', *Phys. Fluids* **31**(1), 015105.

APPENDIX A

Data generation - Kraichnan turbulence

1.1 Numerical schemes

The following section outlines the various numerical schemes used in this investigation for the calculation of derivatives, nonlinear terms, solution of the Poisson equation and spatial filtering operators for DNS generation in our two-dimensional turbulence test-case.

1.1.1 Arakawa scheme

For our nonlinear Jacobian calculations, we utilized the Arakawa scheme for its conservative properties. It was proposed by Arakawa Arakawa (1966) that the conservation of energy, enstrophy and skew symmetry is sufficient to avoid computational instabilities arising from nonlinear interactions in the Jacobian. The Jacobian in the governing equations for 2D turbulence is defined as

$$J(\psi,\omega) = \frac{\partial\psi}{\partial y}\frac{\partial\omega}{\partial x} - \frac{\partial\psi}{\partial x}\frac{\partial\omega}{\partial y} \tag{1.1}$$

The second-order Arakawa scheme for the Jacobian is given by

$$J_I(\omega,\psi) = \frac{1}{3}\left(J_1(\omega,\psi) + J_2(\omega,\psi) + J_3(\omega,\psi)\right) \tag{1.2}$$

where the discrete parts of the Jacobian are

$$\begin{aligned} J_1(\psi,\omega) = \frac{1}{4h_x h_y}\,[&(\omega_{i+1,j} - \omega_{i-1,j})(\psi_{i,j+1} - \psi_{i,j-1}) \\ &-(\omega_{i,j+1} - \omega_{i,j-1})(\psi_{i+1,j} - \psi_{i-1,j})] \end{aligned} \tag{1.3}$$

$$\begin{aligned} J_2(\psi,\omega) = \frac{1}{4h_x h_y}\,[&\omega_{i+1,j}(\psi_{i+1,j+1} - \psi_{i+1,j-1}) \\ -\omega_{i-1,j}(\psi_{i-1,j+1} - \psi_{i-1,j-1}) &- \omega_{i,j+1}(\psi_{i+1,j+1} - \psi_{i-1,j+1}) \\ &+\omega_{i,j-1}(\psi_{i+1,j-1} - \psi_{i-1,j-1})] \end{aligned} \tag{1.4}$$

$$
\begin{aligned}
J_3(\psi,\omega) = \frac{1}{4h_x h_y}\left[\omega_{i+1,j+1}(\psi_{i,j+1}-\psi_{i+1,j})\right. \\
-\omega_{i-1,j-1}(\psi_{i-1,j}-\psi_{i,j-1}) - \omega_{i-1,j+1}(\psi_{i,j+1}-\psi_{i-1,j}) \\
\left.+\omega_{i+1,j-1}(\psi_{i+1,j}-\psi_{i,j-1})\right].
\end{aligned} \tag{1.5}
$$

1.1.2 Time integration scheme

An optimal third-order-accurate total variation diminishing Runge-Kutta scheme (TVDRK3) is used for explicit advancement in time. In order to implement this scheme the model equations are cast in the following form

$$
\frac{du}{dt} = \pounds(u) \tag{1.6}
$$

where $\pounds(u)$ encompasses the spatial derivatives including the nonlinear convective terms,the linear diffusive terms and the source terms modeling the subgrid scale stresses. Assuming that the numerical approximation at a time level $n+1$ is known and our time step is given by Δt, our time stepping scheme becomes Gottlieb and Shu (1998)-

$$
\begin{aligned}
u^{(1)} &= u^n + \Delta t \pounds(u^n), \\
u^{(2)} &= \frac{3}{4}u^n + \frac{1}{4}u^{(1)} + \frac{1}{4}\Delta t \pounds(u^{(1)}), \\
u^{n+1} &= \frac{1}{3}u^n + \frac{2}{3}u^{(2)} + \frac{2}{3}\Delta t \pounds(u^{(2)}).
\end{aligned} \tag{1.7}
$$

The TVDRK3 scheme has been shown to predict slightly more accurate results than some of the other third order Runge-Kutta schemes for incompressible flow problems San and Staples (2012) and has been extensively used in hyperbolic conservation laws Shu and Osher (1988); Gottlieb et al. (2001); Hu et al. (2010); Kara et al. (2011).

1.1.3 Poisson solver

The elliptic equation obtained for the relationship between the streamfunction and vorticity is solved by a FFT-based Poisson solver. The low-pass spatial filtered version of the elliptic equation can be written in the form of $\nabla^2 u = f$. The compact fourth-order discretization with nine point stencil can be written as Wang and Zhang (2009)

$$
\begin{aligned}
&au_{i,j} + b(u_{i+1,j} + u_{i-1,j}) + c(u_{i,j+1} + u_{i,j-1}) \\
&\quad + d(u_{i+1,j+1} + u_{i+1,j-1} + u_{i-1,j+1} + u_{i-1,j-1}) \\
&\quad = e(8f_{i.j} + f_{i+1,j} + f_{i-1,j} + f_{i,j+1} + f_{i,j-1}),
\end{aligned} \tag{1.8}
$$

where the coefficients are $a = -10(1+\gamma^2)$, $b = 5-\gamma^2$, $c = 5\gamma^2 - 1$, $d = (1+\gamma^2)/2$, $e = h_x^2/2$ with γ being the mesh aspect ratio defined as the ratio of grid discretization lengths h_x/h_y. In our case $\gamma = 1$ for which this scheme becomes the

famous Mehrstellen scheme Tsukerman (2006); San (2015). The presence of periodic boundary conditions in both directions suggests the use of the Fourier transform approach as against a finite sine or cosine transform. First we perform and inverse transform for the source term to get

$$\hat{f} = \frac{1}{N_x N_y} \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} f_{i,j} e^{-\mathbf{i}(k_x x_i + k_y y_j)} \tag{1.9}$$

where $k_x = 2\pi m/L_x$ and $k_y = 2\pi n/L_y$. Our elliptic equation in the Fourier space is given by

$$(-k_x^2 - k_y^2)\hat{u}_{m,n} = -\hat{f}_{m,n} \tag{1.10}$$

This equation may now be solved for the unknown $\hat{u}_{k,l}$. Once this is achieved, a forward Fourier transform into the real domain is obtained by

$$u_{i,j} = \sum_{m=-\frac{N_x}{2}}^{\frac{N_x}{2}-1} \sum_{n=-\frac{N_y}{2}}^{\frac{N_y}{2}-1} \hat{u}_{m,n} e^{\mathbf{i}(k_x x_i + k_y y_j)}. \tag{1.11}$$

For the purpose of transforming to and back from the Fourier domain, we use the FFT algorithm given in Press et al Press et al. (1992).

1.2 Initial conditions

The Kraichnan turbulence problem involves an initial energy spectrum which decays through time and we examine this decay by studying the evolution of vortices through time. In this study, we compare the energy spectra developed at the end of our simulation (i.e., $t = 4$) in order to quantify the performance of the SGS closures. The spectra are compared to the theoretical scaling expected from a perfectly homogeneous case of 2D turbulence given by KBL theory. The initial energy spectrum of the problem is given by (Orlandi, 2012)

$$E(k) = Ak^4 \exp\left(-(k/k_p)^2\right) \tag{1.12}$$

where

$$A = \frac{4k_p^{-5}}{3\pi} \tag{1.13}$$

and $k = |\mathbf{k}| = \sqrt{k_x^2 + k_y^2}$. The maximum value of the energy spectrum is designed to occur at wavenumber k_p which is assumed to occur at $k_p = 10$ in this study. The magnitude of vorticity Fourier coefficients related to the assumed initial energy

spectrum becomes

$$|\widetilde{\omega}(\mathbf{k})| = \sqrt{\frac{k}{\pi}E(k)}. \tag{1.14}$$

The initial vorticity distribution in Fourier space is then obtained through the introduction of a random phase

$$\widetilde{\omega}(\mathbf{k}) = \sqrt{\frac{k}{\pi}E(k)}e^{\mathbf{i}\zeta(\mathbf{k})} \tag{1.15}$$

where the phase function is given by $\zeta(\mathbf{k}) = \xi(\mathbf{k}) + \eta(\mathbf{k})$, where $\xi(\mathbf{k})$ and $\eta(\mathbf{k})$ are independent random values chosen in $[0, 2\pi]$ at each coordinate point in the first quadrant of the $k_x - k_y$ plane. Their behavior in the other quadrants are given by conjugate relations as shown in Fig. A.1. Once the randomization process is

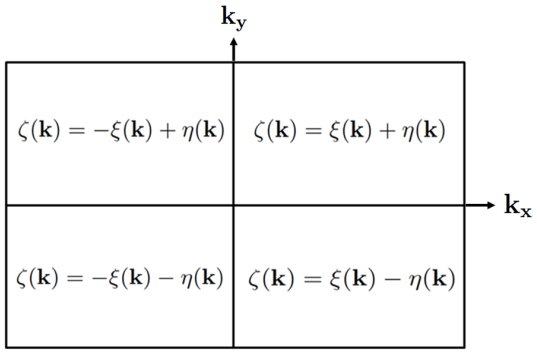


Figure A.1: Conjugate relations for the random phase function for the initial conditions.

completed, the initial vorticity distribution in the physical domain is obtained through an inverse FFT. It must be noted here that the randomization process is identical for each run (whether LES or DNS) for the purpose of comparison. The initial condition for vorticity is chosen to ensure a divergence free vorticity field. A timestep of $\Delta t = 5 \times 10^{-4}$ is selected to ensure that the DNS remains independent of errors associated with temporal discretization (i.e., negligible with respect to the spatial discretization errors).

In order to quantify the benefits of each SGS model combined with a particular solver, we define a statistical measure based on the energy spectrum in the wavenumber domain which is defined as

$$\hat{E}(\mathbf{k}, t) = \frac{1}{2}k^2|\hat{\psi}(\mathbf{k}, t)|^2 \tag{1.16}$$

and the angle averaged energy spectrum is

$$E(k,t) = \sum_{k-\frac{1}{2} \leq |\acute{\mathbf{k}}| \leq k+\frac{1}{2}} \hat{E}(\acute{\mathbf{k}}, t). \tag{1.17}$$

As mentioned previously, the result of the angle averaged energy spectrum for a run is compared to the classical KBL theory scaling which approaches k^{-3} in the limit of infinite Reynolds number.

1.3 Data sampling for learning

In the following we discuss how snapshot data is utilized from DNS results for the purpose of data-driven closure modeling. For studies undertaken on the Kraichnan turbulence test-case, we utilze 5 snapshots of data at non-dimensional times of $t = 0, 1, 2, 3, 4$ out of a potential 40000 snapshots of explicit DNS time-integration. This accounts for 0.0125% of the total snapshot data available for the purpose of data set generation. The motivation behind the extreme temporal sub-sampling of DNS data is to mimic learning data generation specific constraints such as extreme input-output costs when running full-order models on distributed computing environments. The choice of $t = 0, 1$ ensures that the learning has some experience of the initial and transitionary phases of this unsteady problem where closure needs are particularly different. We note that the non-dimensionalization of the Kraichnan turbulence problem is obtained through the usual identification of a relevant length scale given by the side length of the square-periodic domain.

APPENDIX B

Data generation - Kolmogorov turbulence

2.1 Governing equations

We shall utilize the three-dimensional Navier-Stokes equations in their conservative dimensionless form as our underlying governing laws for fluid flow evolution. These can be expressed as

$$\frac{\partial q}{\partial t}+\frac{\partial F}{\partial x}+\frac{\partial G}{\partial y}+\frac{\partial H}{\partial z}=\frac{\partial F^v}{\partial x}+\frac{\partial G^v}{\partial y}+\frac{\partial H^v}{\partial z} \tag{2.1}$$

where

$$q=\begin{bmatrix}\rho\\ \rho u\\ \rho v\\ \rho w\\ \rho e\end{bmatrix},\quad F=\begin{bmatrix}\rho u\\ \rho u^2+P\\ \rho uv\\ \rho uw\\ \rho uH\end{bmatrix},\quad G=\begin{bmatrix}\rho v\\ \rho uv\\ \rho v^2+P\\ \rho vw\\ \rho vH\end{bmatrix},\quad H=\begin{bmatrix}\rho w\\ \rho uw\\ \rho uv\\ \rho w^2+P\\ \rho wH\end{bmatrix} \tag{2.2}$$

are the inviscid contributions to the governing laws and the viscous contributions are given by

$$\begin{gathered}F^v=\begin{bmatrix}0\\ \tau_{xx}\\ \tau_{xy}\\ \tau_{xz}\\ u\tau_{xx}+v\tau_{xy}+w\tau_{xz}-q_x\end{bmatrix},\ G^v=\begin{bmatrix}0\\ \tau_{yx}\\ \tau_{yy}\\ \tau_{yz}\\ u\tau_{yx}+v\tau_{yy}+w\tau_{yz}-q_y\end{bmatrix},\\ H^v=\begin{bmatrix}0\\ \tau_{zx}\\ \tau_{yz}\\ \tau_{yz}\\ u\tau_{zx}+v\tau_{zy}+w\tau_{zz}-q_z\end{bmatrix}.\end{gathered} \tag{2.3}$$

Here ρ, P, u, v and w are the density, pressure and velocity components in the x, y and z Cartesian directions. Also, H is the total enthalpy, e is the total energy and both are linked to each other and the pressure by

$$H=e+P/\rho,\quad P=\rho(\gamma-1)\left(e-\frac{1}{2}(u^2+v^2+w^2)\right), \tag{2.4}$$

where $\gamma = 7/5$ is chosen as our ratio of specific heats. The viscous contributions can be described as

$$\begin{aligned}
\tau_{xx} &= \frac{2}{3}\frac{\mu}{Re}\left(2\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} - \frac{\partial w}{\partial z}\right) \\
\tau_{yy} &= \frac{2}{3}\frac{\mu}{Re}\left(2\frac{\partial v}{\partial y} - \frac{\partial u}{\partial x} - \frac{\partial w}{\partial z}\right) \\
\tau_{zz} &= \frac{2}{3}\frac{\mu}{Re}\left(2\frac{\partial w}{\partial z} - \frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right) \\
\tau_{xy} &= \tau_{yx} = \frac{\mu}{Re}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right) \\
\tau_{xz} &= \tau_{zx} = \frac{\mu}{Re}\left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x}\right) \\
\tau_{yz} &= \tau_{zy} = \frac{\mu}{Re}\left(\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}\right) \\
q_x &= -\frac{\mu}{RePrMa^2(\gamma-1)}\frac{\partial T}{\partial x} \\
q_y &= -\frac{\mu}{RePrMa^2(\gamma-1)}\frac{\partial T}{\partial y} \\
q_z &= -\frac{\mu}{RePrMa^2(\gamma-1)}\frac{\partial T}{\partial z}
\end{aligned} \tag{2.5}$$

where Re, Pr and Ma are the non-dimensional Reynolds, Prandtl and Mach numbers respectively and where Sutherland's law is used to evolve our dimensionless molecular viscosity

$$\mu = \frac{T^{1.5}(1 + S/T_{ref})}{T + S/T_{ref}} \tag{2.6}$$

where $S = 110.4K$ and $T_{ref} = 300K$. The convective flux Jacobian matrices for our conservation laws can be described as (Laney, 1998)

$$A = \frac{\partial F}{\partial q} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ -u^2 + \frac{\gamma-1}{2}(\mathbf{u}.\mathbf{u}) & (3-\gamma)u & -(\gamma-1)v & -(\gamma-1)w & \gamma-1 \\ -uv & v & u & 0 & 0 \\ -uw & w & 0 & u & 0 \\ -(\gamma e - (\gamma-1)\mathbf{u}.\mathbf{u})u & \gamma e - \frac{\gamma-1}{2}(2u^2 + \mathbf{u}.\mathbf{u}) & -(\gamma-1)uv & -(\gamma-1)uw & \gamma u \end{bmatrix}, \tag{2.7}$$

$$B = \frac{\partial G}{\partial q} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ -uv & v & u & 0 & 0 \\ -v^2 + \frac{\gamma-1}{2}(\mathbf{u}.\mathbf{u}) & -(\gamma-1)u & (3-\gamma)v & -(\gamma-1)w & \gamma-1 \\ -vw & 0 & w & v & 0 \\ -(\gamma e - (\gamma-1)\mathbf{u}.\mathbf{u})v & -(\gamma-1)uv & \gamma e - \frac{\gamma-1}{2}(2v^2 + \mathbf{u}.\mathbf{u}) & -(\gamma-1)vw & \gamma v \end{bmatrix}, \tag{2.8}$$

$$C = \frac{\partial H}{\partial q} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ -uw & w & 0 & u & 0 \\ -vw & 0 & w & v & 0 \\ -w^2 + \frac{\gamma-1}{2}(\mathbf{u}.\mathbf{u}) & -(\gamma-1)u & -(\gamma-1)v & (3-\gamma)w & \gamma-1 \\ -(\gamma e - (\gamma-1)\mathbf{u}.\mathbf{u})w & -(\gamma-1)uw & -(\gamma-1)vw & \gamma e - \frac{\gamma-1}{2}(2w^2 + \mathbf{u}.\mathbf{u}) & \gamma\omega \end{bmatrix}, \quad (2.9)$$

where $\mathbf{u}$ is the three-dimensional velocity vector. Each Jacobian matrix also has a similarity transformation given by

$$\begin{aligned} L_A A R_A = \Lambda_A \Rightarrow R_A \Lambda_A L_A = A \\ L_B B R_B = \Lambda_B \Rightarrow R_B \Lambda_B L_B = B \\ L_C C R_C = \Lambda_C \Rightarrow R_C \Lambda_C L_C = C \end{aligned} \quad (2.10)$$

The reader is directed to Bidadi and Rani (2015) for an elaboration of the right and left eigenvectors as well as the eigenvalues for each Jacobian matrix.

2.2 Numerical schemes

2.2.1 Finite volume framework

The semi-discrete form of the governing equations can be written as

$$\begin{aligned} &\frac{dq_{i,j,k}}{dt} + \frac{1}{\Delta x}\left(F_{i+1/2,j,k} - F_{i-1/2,j,k}\right) + \frac{1}{\Delta y}\left(G_{i,j+1/2,k} - G_{i,j-1/2,k}\right) + \frac{1}{\Delta z}\left(H_{i,j,k+1/2} - H_{i,j,k-1/2}\right) \\ &= \frac{1}{\Delta x}\left(F^v_{i+1/2,j,k} - F^v_{i-1/2,j,k}\right) + \frac{1}{\Delta y}\left(G^v_{i,j+1/2,k} - G^v_{i,j-1/2,k}\right) + \frac{1}{\Delta z}\left(H^v_{i,j,k+1/2} - H^v_{i,j,k-1/2}\right) \end{aligned} \quad (2.11)$$

with $q_{i,j,k}$ being the cell-averaged vector of dependant variables, $F_{i\pm1/2,j,k}$ representing the cell face flux reconstructions in the x direction, $G_{i,j\pm1/2,k}$ representing the cell face flux reconstructions in the y direction and $H_{i,j,k\pm1/2}$ representing the cell face flux reconstructions in the z direction. The method of lines may be used to represent our system of PDEs as an ODE through time. Our system may be expressed as

$$\frac{dq_{i,j,k}}{dt} = \pounds(q_{i,j,k}). \quad (2.12)$$

The right hand side of the above equation is representative of the combined effect of the various inviscid and viscous spatial derivatives in the conservation equations. A third-order Runge Kutta scheme (TVDRK3) (Gottlieb and Shu, 1998) is implemented for explicit time advancement as follows:

$$\begin{aligned} q^{(1)}_{i,j,k} &= q^{(n)}_{i,j,k} + \Delta t \pounds(q^{(n)}_{i,j,k}) \\ q^{(2)}_{i,j,k} &= \frac{3}{4}q^{(n)}_{i,j,k} + \frac{1}{4}q^{(1)}_{i,j,k} + \frac{1}{4}\Delta t \pounds(q^{(1)}_{i,j,k}) \\ q^{(n+1)}_{i,j,k} &= \frac{1}{3}q^{(n)}_{i,j,k} + \frac{2}{3}q^{(2)}_{i,j,k} + \frac{2}{3}\Delta t \pounds(q^{(2)}_{i,j,k}) \end{aligned} \quad (2.13)$$

with a time step Δt that is prescribed through a CFL criterion as

$$\Delta t = \min\left(\eta\frac{\Delta x}{\max(|\Lambda_A|)}, \eta\frac{\Delta y}{\max(|\Lambda_B|)}, \eta\frac{\Delta z}{\max(|\Lambda_C|)}\right). \tag{2.14}$$

where $\eta = 0.5$ is is selected as the default value for this investigation.

2.2.2 Symmetric flux reconstructions: 6^{th} order Central Schemes

This approach relies on a traditional finite volume formulation for calculating the density, energy and velocities at the cell interfaces following which the cell face fluxes are calculated as a function of these reconstructed variables. The following 6-point stencil symmetric non-dissipative scheme is used for face reconstruction of the conserved quantity (Hyman et al., 1992)

$$q_{i+1/2} = a(q_{i+1} + q_i) + b(q_{i+2} + q_{i-1}) + c(q_{i+3} + q_{i-2}) \tag{2.15}$$

where the stencil coefficients are given by

$$a = 37/60; \quad b = -2/15; \quad c = 1/60. \tag{2.16}$$

Once the relevant face quantities are determined from the nodal values, the fluxes may be calculated for use in Eq. (2.11).

2.2.3 Treatment of viscous terms

In this section, we describe the stencil used for the construction of viscous fluxes. Since viscous fluxes are functions of the primary flow quantities' derivatives, we can represent them in an abstract form as follows

$$F^v_{i+1/2} = \mathfrak{F}(q'_{i+1/2}), \tag{2.17}$$

where the superscript prime denotes the first order derivative for any quantity. A general stencil for the cell face derivative of any quantity q can be given by

$$q'_{i+1/2} = g\left(a(q_{i+1} - q_i) + b(q_{i+2} - q_{i-1}) + c(q_{i+3} - q_{i-2})\right). \tag{2.18}$$

While it is common practice to use an interpolation type approach (similar to the method of finite differences) based on the Taylor series expansion which gives us the following values for the stencil coefficients for sixth order accuracy:

$$g = \frac{1}{128}, \quad a = \frac{150}{\Delta x}, \quad b = -\frac{25}{2\Delta x}, \quad c = \frac{3}{3\Delta x}, \tag{2.19}$$

we, however, utilize the reconstruction method more suited to finite volume representation of conservation laws (Hyman et al., 1992; Shu, 2009) where our stencil

coefficients are given by

$$g = \frac{1}{180}, \quad a = \frac{245}{\Delta x}, \quad b = -\frac{75}{2\Delta x}, \quad c = \frac{10}{3\Delta x}. \tag{2.20}$$

This is because the interpolation of cell center values to obtain cell face quantities followed by the inherently central approach to flux calculation at the cell centers leads to second order accuracy.

2.3 Problem definition

In order to quantify the accuracy of a proposed LES model, a problem must be chosen that displays the fundamental mechanism of isotropic and homogeneous turbulence where an enhancement of vorticity is observed through vortex stretching and the subsequent production of small eddies. The classical cascade of energies from the smaller wavenumbers (or larger scales) to higher wavenumbers (or smaller scales) must also be observed according to the celebrated Kolmogorov (Kolmogorov, 1941*b*) scaling law. The Taylor-Green vortex is a prototype problem which models the decay of isotropic, homogeneous and turbulent periodic vortical flows by demonstrating the aforementioned physical phenomenon. The initial condition of this problem is defined on a computational domain comprised of a cubic box with periodic boundary conditions in all directions and an edge length of 2π. These initial conditions may be expressed as

$$\begin{aligned}
\rho(x,y,z,t=0) &= 1.0 \\
u(x,y,z,t=0) &= \sin(x)\cos(y)\cos(z) \\
v(x,y,z,t=0) &= -\cos(x)\sin(y)\cos(z) \\
w(x,y,z,t=0) &= 0.0 \\
p(x,y,z,t=0) &= \frac{1}{\gamma Ma^2} + \frac{(\cos(2x)+\cos(2y))(\cos(2z)+2)}{16}.
\end{aligned} \tag{2.21}$$

Our study fixes the Reynolds number at 1600 and uses $N^3 = 512^3$ as the resolution of the high fidelity results (reference values for Prandtl and Mach numbers are also set as $Pr = 0.72$, and $Ma = 0.08$). At this considerably high Reynolds number ($Re = 1600$), a large amount of scale separation is observed as well as the formation of very fine structures. This represents a comparatively greater challenge for our LES models. One of the turbulence statistical quantities chosen for LES evaluation is given by the total kinetic energy of the solution field obtained by spatially averaging the instantaneous kinetic energy at all points

$$E(t) = \frac{1}{V}\int\int\int E(x,y,z,t)dxdydz, \tag{2.22}$$

where V is the volume of the physical domain and $E(x,y,z,t)$ is the instantaneous kinetic energy per unit mass at a particular point in the solution field (i.e., $E =$

$1/2(u^2+v^2+w^2)$). Thus our integral may also be expressed as

$$E(t) = \frac{1}{N_x N_y N_z} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \sum_{k=1}^{N_z} \frac{1}{2} \left(u_{i,j,k}^2(t) + v_{i,j,k}^2(t) + w_{i,j,k}^2(t) \right) \tag{2.23}$$

where we have cell centered values for our velocity components. Another measure of comparison can be obtained from averaged kinetic energy spectral scaling which may be calculated using the following definition

$$E(\hat{\mathbf{k}}, t) = \frac{1}{2} |\hat{\mathbf{u}}(\mathbf{k}, t)|^2 \tag{2.24}$$

where $\hat{\mathbf{u}}(\mathbf{k}, t)$ is the Fourier transform of the velocity vector in the wavenumber space. Eq. (2.24) may also be written as

$$E(\hat{\mathbf{k}}, t) = \frac{1}{2} \left(|\hat{u}(\mathbf{k}, t)|^2 + |\hat{v}(\mathbf{k}, t)|^2 + |\hat{w}(\mathbf{k}, t)|^2 \right) \tag{2.25}$$

where the fast Fourier transform algorithm given in Press et al. (1992) has been used for forward and inverse transforms to and from wavenumber space. An angle averaging of the spectra is then carried out

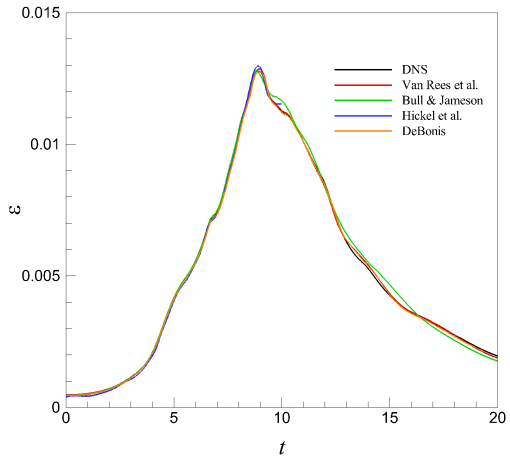
$$E(k, t) = \sum_{k-\frac{1}{2} \leq |\acute{\mathbf{k}}| < k+\frac{1}{2}} \hat{E}(\acute{\mathbf{k}}, t) \tag{2.26}$$

2.4 Validation

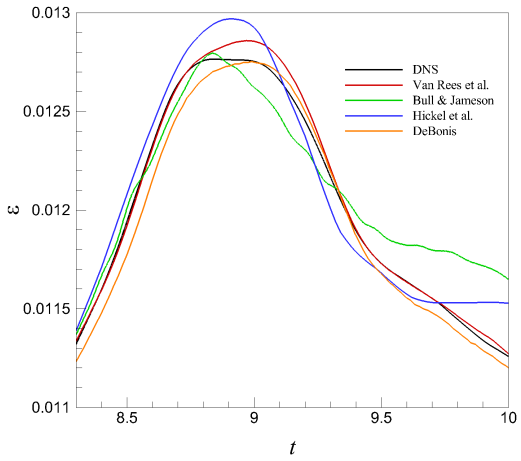
For the purpose of a validation of our benchmark numerical methods, our DNS results are compared with those available from certain famous investigations of the Taylor-Green vortex problem as shown in Fig. (B.1) where the energy dissipation rate is compared. Among the reference data shown here, Bull and Jameson (2015) have implemented a flux reconstructed spectral difference (or FR-SD) scheme at $N^3 = 256^3$ for their highest resolution, Van Rees et al. (2011) have used a pseudo-spectral simulation at $N^3 = 768^3$ for their highest fidelity data, Hickel et al. (2006) have used DNS data from the spectral simulations implemented by DeBonis (2013) have used a 13-point dispersive relation preserving (DRP) scheme at $N^3 = 512^3$ for their baseline solutions. It can be seen that our high fidelity data captures the trends of the reference data well and can thus be used as a benchmark solution.

2.5 Data sampling for learning

In a manner similar to the Kraichnan turbulence test case, any learning which utilizes full-order runs from the TGV test case uses snapshots at times $t = 5$ and $t = 10$ for the purpose of learning. This selection of snapshots ensures that the ML architecture is exposed two different closure modeling regimes. The forward deployment in a-posteriori therefore allows for greater flexibility in closure predictions. Snapshots at



(a) Energy dissipation rate



(b) Close-up for the peak dissipation rate

Figure B.1: A validation of our direct numerical simulation (DNS) using reference data from Van Rees et al. (2011), Bull and Jameson (2015), Hickel et al. (2006) & DeBonis (2013).

$t = 15$ and $t = 20$ are utilized for a-priori assessments but are not utilized for teaching a framework for a-posteriori deployment. We note that the forward deployment of closures was performed till $t = 10$ for consistent assessment. This problem is non-dimensionalized as well in a manner similar to that described for the Kraichnan turbulence test-case.

VITA

ROMIT MAULIK

Candidate for the Degree of

Doctor of Philosophy

Dissertation: DATA-DRIVEN SUB-GRID MODEL DEVELOPMENT FOR LARGE EDDY SIMULATIONS OF TURBULENCE

Major Field: MECHANICAL & AEROSPACE ENGINEERING

Biographical:

Personal Data: Born in Mohannanagar Township, TN, India in November 1989.

Education:

Completed the requirements for the degree of Doctor of Philosophy with a major in Mechanical & Aerospace Engineering at Oklahoma State University in May 2019.

Completed the requirements for the degree of Master of Science with a major in Mechanical & Aerospace Engineering at Oklahoma State University in July 2015.

Received a Bachelor of Engineering in Mechanical Engineering at BIT Mesra, India in May 2012.

Experience:

Graduate Research Assistant, Computational Fluid Dynamics Laboratory
Graduate Teaching Assistant, Mechanical & Aerospace Engineering

Professional Affiliations:

APS, SIAM