Physica D 443 (2023) 133568

Contents lists available at ScienceDirect

Physica D

journal homepage: www.elsevier.com/locate/physd

Review

Learning physics-constrained subgrid-scale closures in the small-data regime for stable and accurate LES

Yifei Guan [a,*], Adam Subel [a], Ashesh Chattopadhyay [a], Pedram Hassanzadeh [a,b]

[a] Department of Mechanical Engineering, Rice University, Houston, TX, 77005, United States
[b] Department of Earth, Environmental and Planetary Sciences, Rice University, Houston, TX, 77005, United States

ARTICLE INFO

Article history:
Received 25 May 2022
Received in revised form 29 September 2022
Accepted 19 October 2022
Available online 2 November 2022
Communicated by Dmitry Pelinovsky

Keywords:
Large eddy simulation
Deep Learning
Turbulence
Physics constraints
Small data

ABSTRACT

We demonstrate how incorporating physics constraints into convolutional neural networks (CNNs) enables learning subgrid-scale (SGS) closures for stable and accurate large-eddy simulations (LES) in the small-data regime (i.e., when the availability of high-quality training data is limited). Using several setups of forced 2D turbulence as the testbeds, we examine the *a priori* and *a posteriori* performance of three methods for incorporating physics: (1) data augmentation (DA), (2) CNN with group convolutions (GCNN), and (3) loss functions that enforce a global enstrophy-transfer conservation (EnsCon). While the data-driven closures from physics-agnostic CNNs trained in the big-data regime are accurate and stable, and outperform dynamic Smagorinsky (DSMAG) closures, their performances substantially deteriorate when these CNNs are trained with 40x fewer samples (the small-data regime). An example based on a vortex dipole demonstrates that the physics-agnostic CNN cannot account for never-seen-before samples' rotational equivariance (symmetry), an important property of the SGS term. This shows a major shortcoming of the physics-agnostic CNN in the small-data regime. We show that CNN with DA and GCNN address this issue and each produce accurate and stable data-driven closures in the small-data regime. Despite its simplicity, DA, which adds appropriately rotated samples to the training set, performs as well or in some cases even better than GCNN, which uses a sophisticated equivariance-preserving architecture. EnsCon, which combines structural modeling with aspect of functional modeling, also produces accurate and stable closures in the small-data regime. Overall, GCNN+EnsCon, which combines these two physics constraints, shows the best *a posteriori* performance in this regime. These results illustrate the power of physics-constrained learning in the small-data regime for accurate and stable LES.

Contents

* Corresponding author.
E-mail addresses: yifei.guan@rice.edu (Y. Guan), pedram@rice.edu (P. Hassanzadeh).

https://doi.org/10.1016/j.physd.2022.133568
0167-2789/

1. Introduction

Large-eddy simulation (LES) is widely used in the modeling of turbulent flows in natural and engineering systems as it offers a balance between accuracy and computational cost. In LES, the large-scale structures are explicitly resolved on a coarse-resolution grid while the subgrid-scale (SGS) eddies are parameterized in terms of the resolved flow using a closure model [1–4]. Therefore, the fidelity of LES substantially depends on the accuracy of the SGS closure, improving which has been a longstanding goal across various disciplines [e.g., 4–14]. In general, the SGS models can be classified into two categories: structural and functional [3]. Structural models, such as the gradient model [15,16], are developed to capture the structure (pattern and amplitude) of the SGS stress tensor and are known to produce high correlation coefficients (c) between the true and predicted SGS terms ($c > 0.9$) in *a priori* analysis. However, they often lead to numerical instabilities in *a posteriori* LES, for example, because of excessive backscattering and/or lack of sufficient dissipation [17–21]. Functional models, such as the Smagorinsky model [1] and its dynamic variants [22–24], are often developed by considering the inter-scale interactions (e.g., energy transfers). While producing low c (< 0.6) between the true and predicted SGS terms in *a priori* analysis [25–28], these functional models usually provide numerically stable *a posteriori* LES, at least partly due to their dissipative nature. Thus, developing SGS models that perform well in both *a priori* and *a posteriori* analyses has remained a long-lasting research focus.

In recent years, there has been a rapidly growing interest in using machine learning (ML) methods to learn data-driven SGS closure models from filtered direct numerical simulation (DNS) data [e.g., 25,29–44]. Different approaches applied to a variety of canonical fluid systems have been investigated in these studies. For example, Maulik et al. [31,45] and Xie et al. [46–48] have, respectively, developed *local* data-driven closures for 2D decaying homogeneous isotropic turbulence (2D-DHIT) and 3D incompressible and compressible turbulence using multilayer perceptron artificial neural networks (ANNs); also see [28,49–51]. Zanna and Bolton [35,52], Beck and colleagues [32,53], Pawar et al. [25], Guan et al. [26], and Subel et al. [54] developed *non-local* closures, e.g., using convolutional neural networks (CNNs), for ocean circulation, 3D-DHIT, 2D-DHIT, and forced 1D Burgers' turbulence, respectively. While finding outstanding results in *a priori* analyses, in many cases, these studies also reported instabilities in *a posteriori* analyses, requiring further modifications to the learnt closures for stabilization.

More recently, Guan et al. [26] showed that increasing the size of the training set *alone* can lead to stable and accurate *a posteriori* LES (as well as high c in *a priori* analysis) even with physics-agnostic CNNs.[1] This was attributed to the following: Big training datasets obtained from filtered DNS (FDNS) snapshots can provide sufficient information such that the physical constraints and processes (e.g., backscattering) are correctly learnt by data-driven methods, leading to a stable and accurate LES in both *a priori* and *a posteriori* analyses. However, in the *small-data* regime, the physical constraints and processes may not be captured correctly, and the inaccuracy of the data-drivenly predicted SGS term (particularly inaccuracies in backscattering) can result in unstable or unphysical LES [26]. Note that as discussed later in Section 3, whether a training dataset is *big* or *small* depends on both the number of samples and the inter-sample correlations; thus it depends on the total length of the available DNS dataset, which can be limited due to computational constraints. As a result, there is a need to be able to learn data-driven closures in the small-data regime for stable and accurate LES. This can be achieved by incorporating physics into the learning process, which is the subject of this study.

Past studies have shown that embedding physical insights or constraints can enhance the performance of data-driven models, e.g., in reduced-order models [e.g., 55–65] and in neural networks [e.g., 44,54,66–77]. There are various ways to incorporate physics in neural networks (e.g., see the reviews by Kashinath et al. [78] and Karniadakis et al. [79]). For neural network-based data-driven SGS closures, in general, three of the main ways to do this are: data augmentation (DA), physics-constrained loss functions, and physics-aware network architectures. Training datasets can be constructed to represent some aspects of physics. For example, Galilean invariance and some of the translational and rotational equivariances of the SGS term [2,80] can be incorporated through DA, i.e., built into the input and output training samples [33,54,81]. Here, "equivariance" means that the SGS terms are preserved under some coordinate transformations, resulting from properties of the Navier–Stokes equations [2]; see Appendix A for more details. Physical constraints such as conservation laws can also be included through an augmented loss function — the optimization target during training [e.g., 72,73,82–84]. Finally, physical constraints can also be enforced in the neural network architecture, e.g., by modifying particular layers [35,85] or using special components such as equivariance-preserving spatial transformers [86–88].

Building on these earlier studies, here we aim to examine the effectiveness of three methods for incorporating physics into the learning of non-local, data-driven SGS closure models, with a particular focus on the performance in the small-data regime. The three methods employed here are:

(a) DA, for incorporating rotational equivariances into the input and output training samples,
(b) Physics-constrained loss function that enforces a global enstrophy constraint (EnsCon),
(c) Group CNN (GCNN), a type of equivariance-preserving CNN that has rotational symmetries (equivariances) built into its architecture.

The test case here is a deterministically forced 2D-HIT flow. As discussed in the paper, the use of DA and GCNN are inspired by an example showing the inability of a CNN to account for rotational equivariances in the small-data regime, while the use of EnsCon is motivated by the success of similar global energy constraints in improving the performance of reduced-order and closure models in past studies [55,73]. We examine the accuracy of the learnt closure models in *a priori* (offline) tests, in terms of both predicting the SGS terms and capturing inter-scale transfers, and the accuracy and stability of LES with these SGS models in *a posteriori* (online) tests, with regard to long-term statistics.

The remaining sections of this paper are as follows. Governing equations of the forced 2D-HIT system, the filtered equations, and

[1] For simplicity, we refer to any "physics-agnostic CNN" as CNN hereafter.

the DNS and LES numerical solvers are presented in Section 2. The CNNs and their *a priori* performance trained in the small- and big-data regimes are discussed in Section 3. The physics-constrained CNN models (DA, GCNN, and EnsCon) are described in Section 4. Results of the *a priori* and *a posteriori* tests with these CNNs in the *small* data regime are shown in Section 5. Conclusions and future work are discussed in Section 6.

2. DNS and LES: Equations, numerical solvers, and filterings

2.1. Governing equations

As the testbed, we use forced 2D-HIT, which is a fitting prototype for many large-scale geophysical and environmental flows (where rotation and/or stratification dominate the dynamics). This system has been widely used as a testbed for novel techniques, including ML-based SGS modeling [e.g., 89–95]. The dimensionless governing equations in the vorticity (ω) and stream-function (ψ) formulation in a doubly periodic square domain with length $L = 2\pi$ are:

$$\frac{\partial \omega}{\partial t} + \mathcal{N}(\omega, \psi) = \frac{1}{Re}\nabla^2\omega - f - r\omega, \tag{1a}$$

$$\nabla^2\psi = -\omega. \tag{1b}$$

Here, $\mathcal{N}(\omega, \psi)$ represents the nonlinear advection term:

$$\mathcal{N}(\omega, \psi) = \frac{\partial \psi}{\partial y}\frac{\partial \omega}{\partial x} - \frac{\partial \psi}{\partial x}\frac{\partial \omega}{\partial y}, \tag{2}$$

and f represents a deterministic forcing [e.g., 90,93]:

$$f(x, y) = k_f[\cos(k_f x) + \cos(k_f y)]. \tag{3}$$

We study 5 cases, in which the forcing wavenumber (k_f) and linear friction coefficient (r) have been varied, creating a variety of flows that differ in dominant length scales and energy/enstrophy cascade regimes (Fig. 1). For all cases, the Reynolds number (Re) is set to 20 000. In DNS, as discussed in Section 2.2, Eqs. (1a)–(1b) are numerically solved at high spatio-temporal resolutions. Fig. 1(b) shows the angle-averaged spectra of turbulent kinetic energy $\hat{E}(k)$ and enstrophy $\hat{Z}(k)$, further illustrating the difference in both large and small scales of the flow systems. (See Subel et al. [96] for analyses showing that CNN-based SGS models trained on one of these systems lead to unstable *a posteriori* LES on another system (unless transfer learning is used), further suggesting that these cases cover different dynamics and data distribution.)

To derive the equations for LES, we apply Gaussian filtering [2,3,26], denoted by $\overline{(\cdot)}$, to Eqs. (1a)–(1b) to obtain

$$\frac{\partial \overline{\omega}}{\partial t} + \mathcal{N}(\overline{\omega}, \overline{\psi}) = \frac{1}{Re}\nabla^2\overline{\omega} - \overline{f} - r\overline{\omega} + \underbrace{\mathcal{N}(\overline{\omega}, \overline{\psi}) - \overline{\mathcal{N}(\omega, \psi)}}_{\Pi}, \tag{4a}$$

$$\nabla^2\overline{\psi} = -\overline{\omega}. \tag{4b}$$

The LES can be solved using a coarser resolution (compared to DNS) with the SGS term Π being the unclosed term, requiring a closure model. The power spectra of Π for all cases are shown in Fig. 1(b), illustrating the difference of the SGS terms in both large and small scales.

2.2. Numerical simulations

In DNS, we solve Eqs. (1a)–(1b). A Fourier–Fourier pseudo-spectral solver is used along with second-order Adams–Bashforth and Crank–Nicolson time-integration schemes for the advection and viscous terms, respectively [26]. The computational grid has uniform spacing $\Delta_{\text{DNS}} = L/N_{\text{DNS}}$, where $N_{\text{DNS}} = 1024$ is the number of grid points in each direction. The time-stepping size is set as $\Delta t_{\text{DNS}} = 5 \times 10^{-5}$ dimensionless time unit for all cases except for Case 5, for which $\Delta t_{\text{DNS}} = 2 \times 10^{-5}$ is used. For each case, using different random initial conditions, we conducted 3 independent DNS runs to generate the training, offline testing, and online testing datasets. Once the flow reaches statistical equilibrium after a long-term spin-up, each DNS run produces 2000 snapshots, with each consecutive snapshots $1000\Delta t_{\text{DNS}}$ apart to reduce the correlation between training samples (inter-sample correlation c_{in}, or the correlation coefficient between two consecutive Π, is below 0.75; see Section 3 for further discussions). We have also retained data sampled at $25\Delta t_{\text{DNS}}$ intervals to examine the effect of c_{in}.

For LES, we solve Eqs. (4a)–(4b) employing the same numerical solver used for DNS, but with grid resolutions N_{LES} ($= N_{\text{DNS}}/16$ or $N_{\text{DNS}}/8$) listed in Fig. 1 for each case and $\Delta t_{\text{LES}} = 10\Delta t_{\text{DNS}}$. The SGS term Π is parameterized using a data-driven closure model that is a physics-agnostic or physics-constrained CNN (Sections 3 and 4) or a physics-based dynamic Smagorinsky model (DSMAG). For DSMAG, positive clipping is used to enforce non-negative eddy-viscosity, thus providing stable *a posteriori* LES [25,26].

2.3. Filtered DNS (FDNS) data

To obtain the FDNS and to construct the training and testing data for data-driven methods, we apply a Gaussian filter and then coarse-grain the filtered variables to the LES grid, generating $\overline{\psi}$, $\overline{\omega}$, and Π [2,3]. The filtering and coarse-graining process is described in detail in our recent paper [26], and is only briefly described here. (i) *Spectral transformation*: transform the DNS variables into the spectral space by Fourier transform; (ii) *Filtering*: apply (element-wise-multiply) a Gaussian filter kernel (with filter size $\Delta_F = 2\Delta_{\text{LES}}$) in the spectral space to *filter* the high-wavenumber structures (the resulting variables still have the DNS resolution); (iii) *Coarse-graining*: truncate the wavenumbers greater than the cut-off wavenumber ($k_c = \pi/\Delta_{\text{LES}}$) of the filtered variables in the spectral space ((the resulting variables have the LES resolution); (iv) *Spectral transformation*: transfer the filtered, coarse-grained variables back to the physical space by inverse Fourier transform.

3. Convolutional neural network (CNN): Architecture and results

3.1. Architecture

In this work, we first parameterize the unclosed SGS term Π in (4a) using a physics-agnostic CNN (CNN hereafter) described in this section. The CNN used in this work has the same architecture as the one used in our previous study [26], which is 10-layer deep with fully convolutional layers, i.e., no pooling or upsampling. All layers are randomly initialized and trainable. The convolutional depth is set to be 64, and the convolutional filter size is 5×5. We have performed extensive trial and error analysis for these hyperparameters to prevent over-fitting while maintaining accuracy. For example, a CNN with more than 12 layers overfits on this dataset while a CNN with less than 8 layers results in significantly lower *a priori* correlation coefficients. The activation function of each layer is the rectified linear unit (ReLU) except for the final layer, which is a linear map.

We have standardized the input samples as

$$\left\{\overline{\psi}/\sigma_{\overline{\psi}}, \overline{\omega}/\sigma_{\overline{\omega}}\right\} \in \mathbb{R}^{2\times N_{\text{LES}}\times N_{\text{LES}}}, \tag{5}$$

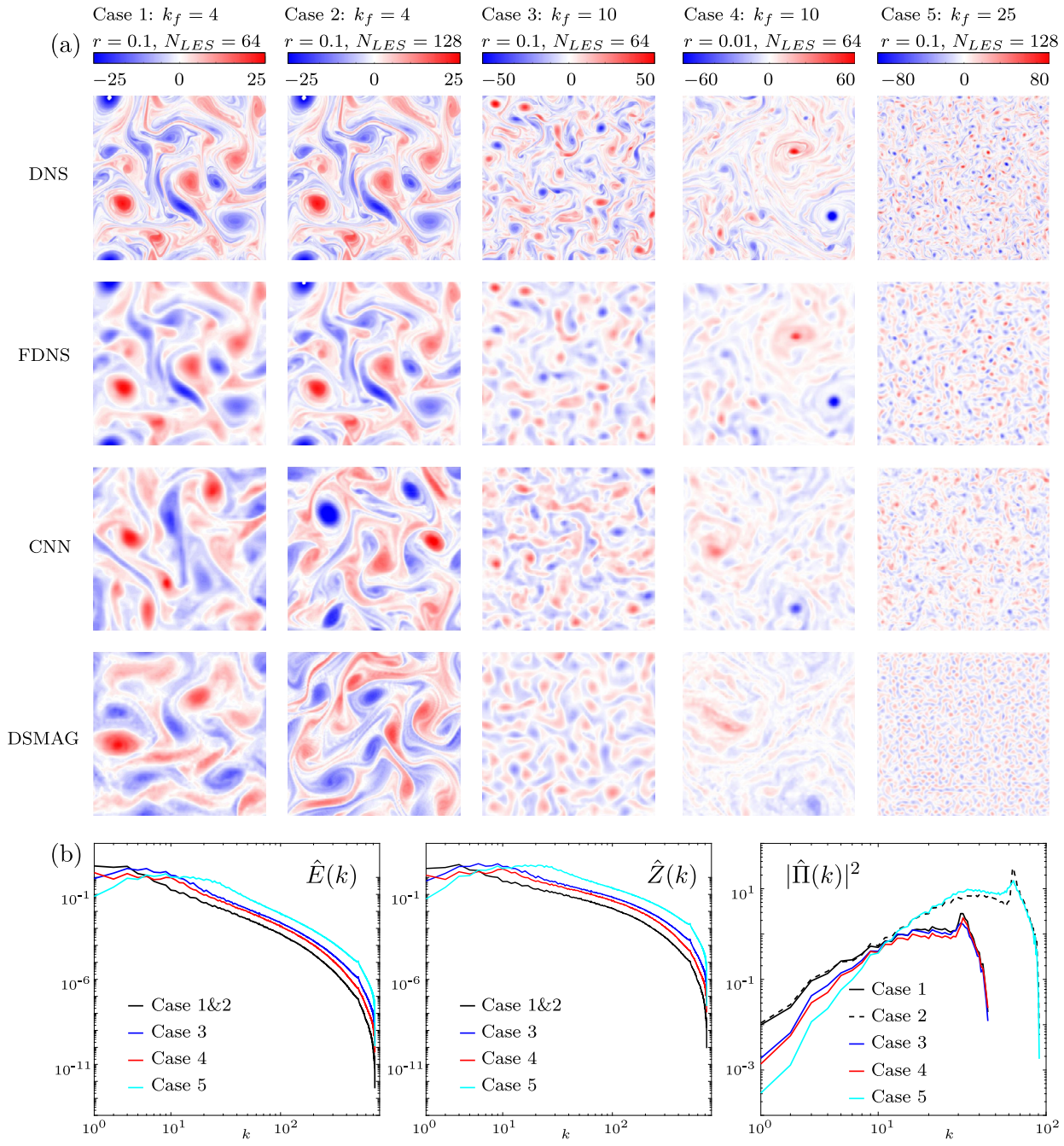

Fig. 1. (a) Examples of vorticity fields of DNS, FDNS, and LES with SGS terms modeled by CNN and DSMAG for 5 cases of forced 2D turbulence with different forcing wavenumber (k_f), friction coefficients (r), and LES resolutions (N_{LES}). For all cases, $Re = 20{,}000$ and $N_{DNS} = 1024$. The scales of the flow structures depend on k_f; the higher the k_f the smaller the scales. The linear drag coefficient, r, determines the similarity of the flow structure. When $r = 0.1$, the flow contains several large vortices of similar sizes. With $r = 0.01$ (Case 4), often two large vortices rotating in opposite directions dominate, co-existing with smaller vortices. For the LES, the CNN is trained on big data (number of training snapshots: $n_{tr} = 2000, c_{in} < 0.75$) to ensure accuracy and numerical stability. DSMAG, in general, captures the large-scale structures but underpredicts the vorticity magnitudes due to the excessive dissipation produced by the non-negative eddy viscosity. (b) Turbulent kinetic energy spectra $\hat{E}(k)$ and enstrophy spectra $\hat{Z}(k)$ of DNS; power spectra of the SGS term Π for all cases. The spectra show the difference in both small and large scales.

and the output samples as

$$\left\{\Pi/\sigma_{\Pi}\right\} \in \mathbb{R}^{N_{\text{LES}}\times N_{\text{LES}}}, \tag{6}$$

where $\sigma_{\overline{\psi}}$, $\sigma_{\overline{\omega}}$, and σ_{Π} are the standard deviations of $\overline{\psi}$, $\overline{\omega}$, and Π calculated over all training samples, respectively. In the later sections, we omit σ for clarity, but we always standardize the input/output samples. The CNN is trained as an optimal map $\mathbb{M}$ between the inputs and outputs

$$\mathbb{M}: \left\{\overline{\psi}/\sigma_{\overline{\psi}}, \overline{\omega}/\sigma_{\overline{\omega}}\right\} \in \mathbb{R}^{2\times N_{\text{LES}}\times N_{\text{LES}}} \rightarrow \left\{\Pi/\sigma_{\Pi}\right\} \in \mathbb{R}^{N_{\text{LES}}\times N_{\text{LES}}}, \tag{7}$$

by minimizing the mean-squared-error (MSE) loss function

$$MSE = \frac{1}{n_{tr}}\sum_{i=1}^{n_{tr}} \|\Pi_i^{\text{CNN}} - \Pi_i^{\text{FDNS}}\|_2^2, \tag{8}$$

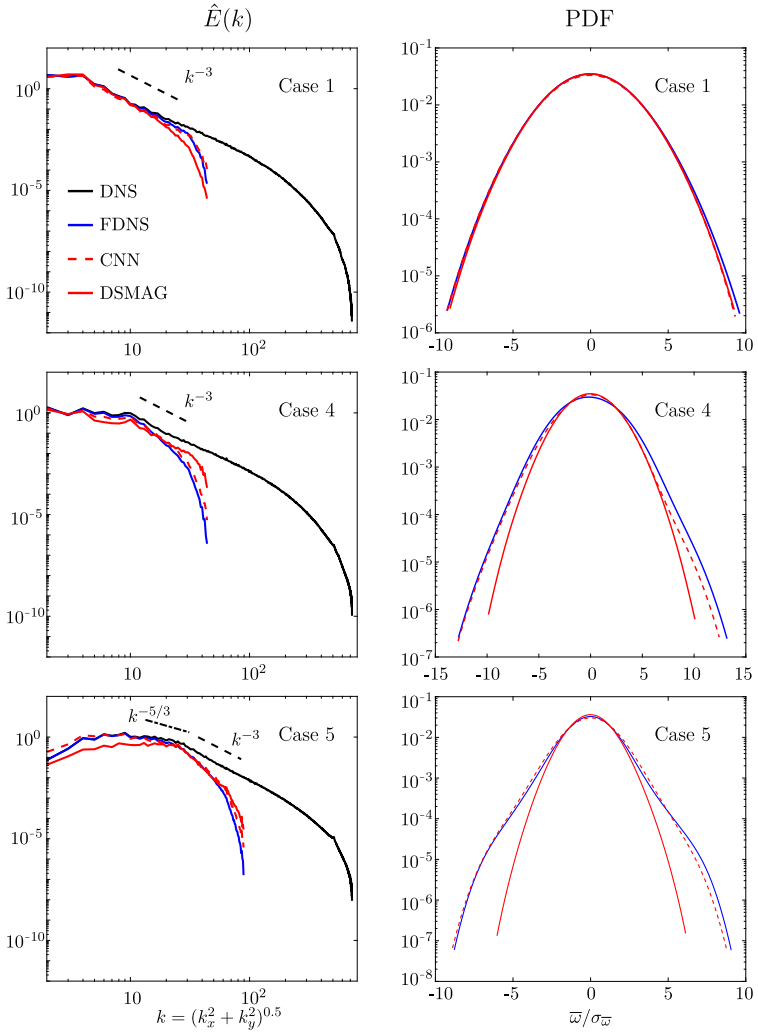

Fig. 2. Turbulent kinetic energy (TKE, $\hat{E}(k)$) spectra and probability density functions (PDFs) of $\overline{\omega}$ for representative cases (1, 4, and 5). The CNNs used here are trained on big data. The spectra and PDFs are averaged over 100 samples from the testing set. The PDF is calculated using a kernel estimator [97].

where n_{tr} is the number of training samples and $\|\cdot\|_2$ is the L_2 norm.

3.2. Results

Fig. 1 shows examples of vorticity fields from DNS and FDNS, and from *a posteriori* LES that uses CNN or DSMAG for the 5 cases. The CNN used here is trained on the full dataset (2000 snapshots with $c_{in} < 0.75$), which we will refer to as "big data" hereafter. Qualitatively, the LES with CNN more closely reproduces the small-scale features of FDNS compared to DSMAG. To better compare the *a posteriori* performance, Fig. 2 shows the turbulent kinetic energy (TKE) spectra $\hat{E}(k)$ and the probability density functions (PDF) of $\overline{\omega}$ averaged over 100 randomly chosen snapshots from LES (spanning $2 \times 10^5 \Delta t_{\text{LES}}$ or equally $2 \times 10^6 \Delta t_{\text{DNS}}$) for 3 representative cases (1, 4, and 5). Note that in forced 2D turbulence, according to the classic Kraichnan–Leith–Batchelor (KLB) similarity theory [98–101], the energy injected by the forcing at wavenumber k_f is transferred to the larger scales ($k < k_f$, energy inverse cascade) while the enstrophy redistributes to the smaller scales ($k > k_f$, enstrophy forward cascade). The KLB theory predicts a $k^{-5/3}$ slope of the TKE spectrum for $k < k_f$ and k^{-3} slope for $k > k_f$.

In general, the $\hat{E}(k)$ of LES with CNN better matches the FDNS than that of the LES with DSMAG. For Cases 1 and 4, where the enstrophy forward cascade dominates, the LES with DSMAG incorrectly captures the spectra at small scales For Case 5, where the energy inverse cascade is important too, the DSMAG fails to recover the energy at large scales correctly. Examining the PDFs of $\overline{\omega}$ shows that in Cases 4 and 5, the PDF from LES with CNN almost overlaps with the one from FDNS even at the tails, while the PDF from LES with DSMAG deviates beyond 3 standard deviations. Due to the excessive dissipation, the LES with DSMAG is incapable of capturing the extremes (tails of the PDF). Therefore, in *a posteriori* analysis, similar to decaying 2D turbulence [26], for different setups of forced 2D turbulent flows, LES with CNN *trained with big data* better reproduces the FDNS flow statistics as compared to LES with DSMAG. Note that in this study, we are comparing the CNN-based closures against DSMAG, which is more accurate and powerful than the typical baseline, the static Smagorinsky model [26,45]. Finally, we highlight that the CNN has outstanding *a priori* performance too, yielding $c > 0.9$ (Fig. 3).

Although the CNNs yield outstanding performance in both *a priori* and *a posteriori* analyses when trained with *big data*, their performance deteriorates when the training dataset is *small*. Before introducing three physics-constrained CNNs for overcoming this problem (Section 4), we first show in Fig. 3 that classifying "big" versus "small" data depends not only on the number of snapshots in the training dataset (n_{tr}) but also on the inter-sample correlation (c_{in}). In *a priori* analysis (bar plots in Fig. 3),

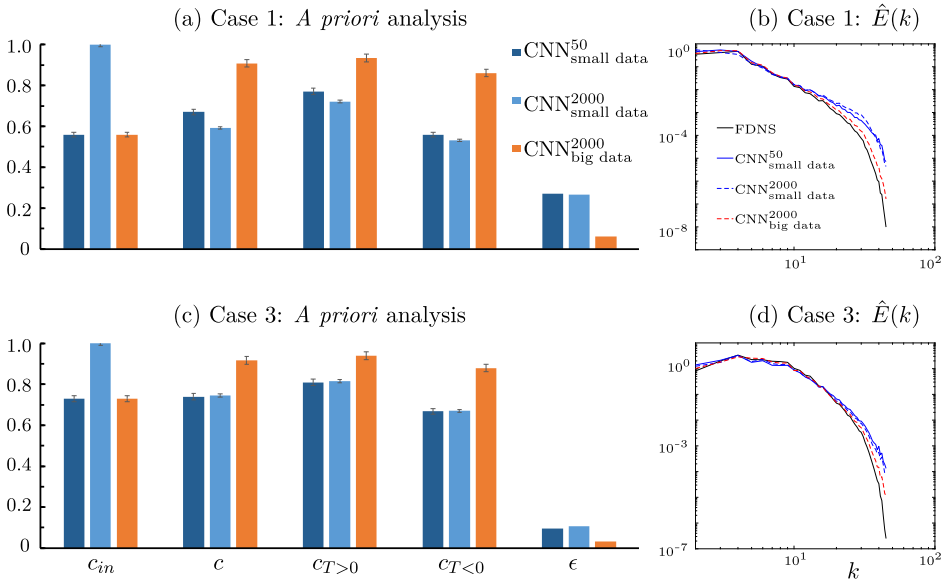

Fig. 3. (a) and (c): *a priori* and (b) and (d): *a posteriori* analyses with 3 CNNs: $\mathrm{CNN}^{50}_{\text{small data}}$ uses $n_{tr} = 50$ with each two consecutive snapshots being $1000\Delta t_{\text{DNS}}$ apart leading to $c_{in} \sim 0.6 - 0.7$, $\mathrm{CNN}^{2000}_{\text{small data}}$ uses $n_{tr} = 2000$ with each two consecutive snapshots being $25\Delta t_{\text{DNS}}$ apart, leading to $c_{in} \sim 0.99$, and $\mathrm{CNN}^{2000}_{\text{big data}}$ uses $n_{tr} = 2000$ with each two consecutive snapshots being $1000\Delta t_{\text{DNS}}$ apart, leading to $c_{in} \sim 0.6 - 0.7$. The training DNS datasets of $\mathrm{CNN}^{50}_{\text{small data}}$ and $\mathrm{CNN}^{2000}_{\text{small data}}$ span the same time range $(50,000\Delta t_{\text{DNS}})$, and their performance are very similar both in *a priori* and *a posteriori* analyses. The training DNS datasets of $\mathrm{CNN}^{2000}_{\text{big data}}$ are 40 times longer ("big data"), and these models outperform the ones trained on "small data" based on all the *a priori* metrics in (a) and (c) and *a posteriori* performance in (b) and (d). The error bars denote plus and minus standard deviation (the error bars on ϵ are small and not shown for the save of clarify).

we use four metrics. The first two are c_{in}, which is the average correlation coefficient between consecutive snapshots of Π in the training set, and c, which is the average correlation coefficient between the true (FNDS) and CNN-predicted Π over 100 random snapshots in the testing set. Following past studies [26,45], we introduce

$$T = sgn(\nabla^2\overline{\omega}) \odot \Pi, \tag{9}$$

whose sign at a grid point determines whether the SGS term is diffusive $T > 0$ or anti-diffusive $T < 0$ ($\odot$ denotes element-wise multiplication). The third metric we use is c computed separately based on the sign of T for testing samples: $c_{T>0}$, which is the average c on the grid points experiencing diffusion by SGS processes and $c_{T<0}$, which is the average c on the grid points experiencing anti-diffusion. Finally, noticing that the global enstrophy transfer due to the SGS term is $\langle\overline{\omega}\Pi\rangle$ (see Appendix B), where $\langle\cdot\rangle$ denotes domain averaging, we define ϵ, the relative error in global enstrophy transfer by the SGS term as

$$\epsilon = |\langle\overline{\omega}\Pi^{\text{CNN}}\rangle - \langle\overline{\omega}\Pi^{\text{FDNS}}\rangle|/|\langle\overline{\omega}\Pi^{\text{FDNS}}\rangle|. \tag{10}$$

Fig. 3 compares the performance of CNNs trained with three training sets for two representative cases (1 and 3). $\mathrm{CNN}^{50}_{\text{small data}}$ uses $n_{tr} = 50$ with each two consecutive snapshots being $1000\Delta t_{\text{DNS}}$ apart, leading to $c_{in} \sim 0.6 - 0.7$, while $\mathrm{CNN}^{2000}_{\text{small data}}$ uses $n_{tr} = 2000$ with each two consecutive snapshots being $25\Delta t_{\text{DNS}}$ apart, leading to highly correlated samples with $c_{in} \approx 0.99$ (note that these two DNS datasets have the same total time length). $\mathrm{CNN}^{2000}_{\text{big data}}$ uses $n_{tr} = 2000$ with each two consecutive snapshots being $1000\Delta t_{\text{DNS}}$ apart, leading to $c_{in} \sim 0.6 - 0.7$ (note that this dataset is 40 times longer than the other two).

The *a priori* results show that $\mathrm{CNN}^{50}_{\text{small data}}$ and $\mathrm{CNN}^{2000}_{\text{small data}}$ have comparable c, $c_{T>0}$, $c_{T<0}$, and ϵ, which are all worse than those of $\mathrm{CNN}^{2000}_{\text{big data}}$. This demonstrates the importance of both n_{tr} and c_{in} in determining the effective size of the training set and the performance of the learnt closure (note that similar to what was shown in our earlier work [26], in the small-data regime, $c_{T<0} < c_{T>0}$). The *a posteriori* analysis leads to the same conclusion: The TKE spectra of LES with $\mathrm{CNN}^{2000}_{\text{big data}}$ closely matches that of the FDNS while the spectra of LES with $\mathrm{CNN}^{50}_{\text{small data}}$ and $\mathrm{CNN}^{2000}_{\text{small data}}$ are comparable and do not match the spectra of FDNS at wavenumbers larger than around 10.

These analyses show that increasing the number of training samples from 50 to 2000 within the same DNS dataset does not enhance the performance of CNN. In general, the performance of CNN depends on the total DNS time length that the training dataset spans. In fact, both the number of training snapshots (n_{tr}) and the inter-sample correlation (c_{in}) determine whether we are in the "big" or "small" data regime. As "big" datasets may not be available for many problems, in the next section, we will discuss how to enhance the performance of the CNNs in the small-data regime using physics constraints.

4. Physics-constraint CNNs: Incorporating rotational equivariance and SGS enstrophy transfer

In this section, we demonstrate how incorporating rotational equivariances via DA or GCNNs, or enforcing a global SGS enstrophy transfer in the loss function can improve the *a priori* and *a posteriori* performance of the CNN-based closures in the *small-data* regime (with $n_{tr} = 50$).

The example in Fig. 4 shows one of the shortcomings of a physics-agnostic CNN: the inability to capture rotational equivariance when the training set is small. In this example, a CNN is trained on snapshots of $\bar{\omega}$ and its Π term (from Gaussian filtering) for an inviscid vertically aligned vortex dipole (first and second columns from left, first row). This dipole moves around the domain, i.e., it translates but does not rotate (thus, a horizontally aligned dipole is never seen in the training set). The trained CNN can accurately predict the Π term for out-of-sample $\bar{\omega}$ snapshots (third column for left, first row). However, for a $\bar{\omega}$ snapshot that is rotated by 90° clockwise (first column, second row), the CNN cannot accurately predict the correct Π term, which is also rotated by 90° clockwise (second column, second row). Note that there was no horizontally aligned dipole like this in the training set. The CNN, instead, predicts an incorrect Π term that is based

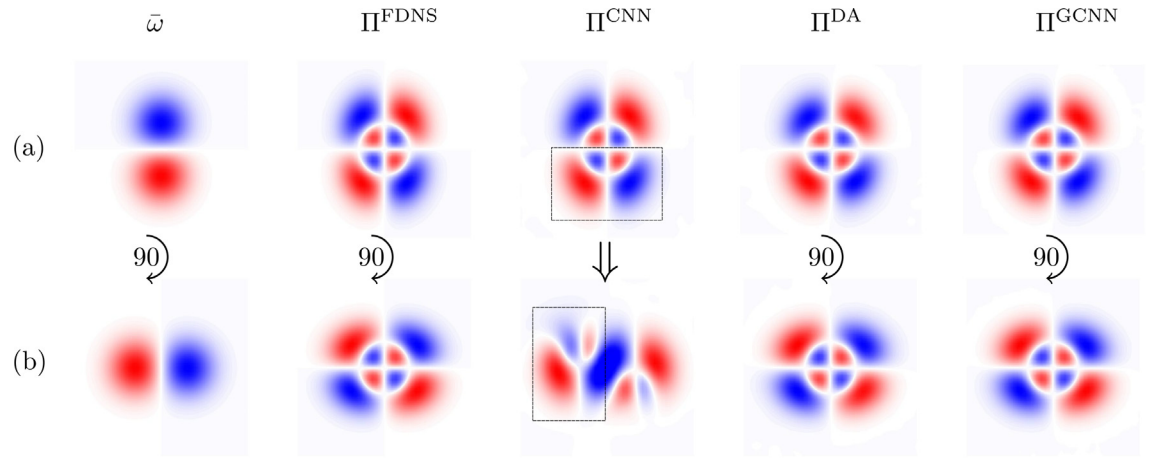

Fig. 4. A dipole vortex shows the shortcoming of a physics-agnostic CNN in capturing the rotational equivariance of the SGS term (third column). The physics-agnostic CNN regards the rotational transform between the training and testing vortex field as a translational transform (the translation of the structure in the black dashed box). However, the CNN with DA or GCNN can capture this rotational equivariance correctly (fourth and fifth columns). The symbol ↻ means rotation by 90° clockwise and ⇓ means translation.

on separate translations of the two parts of the vertically aligned dipole (third column, second row).[2]

The implication of this example is that if the training set only involves limited flow configurations (i.e., small-data regime) such as only those from the first row, then the CNN can be quite inaccurate for a testing set involving new configurations such as those in the second row. In the much more complicated 2D-HIT flow, there are various complex flow configurations. In a big training set, it is more likely that these different configurations would be present and the CNN learns their corresponding Π terms and the associated transformations; however, this is less likely in a small training set. The SGS term Π is known to be equivariant to translation and rotation, i.e., if the flow state variables are translated or rotated, Π should also be translated or rotated to the same degree [2]. While translation equivariance is already achieved in a regular CNN by weight sharing [102], rotational equivariance is not guaranteed. Recent studies show that rotational equivariance can actually be critical in data-driven SGS modeling [33,74,77,81]. To capture the rotational equivariance in the small-data regime, we propose two approaches: (1) DA, by including 3 additional rotated (by 90°, 180°, and 270°) counterparts of each original FDNS snapshot in the training set [81] and (2) by using a GCNN architecture, which enforces rotational equivariance by construction [102,103].

The GCNN uses group convolutions, which increases the degree of weight sharing by transforming and reorienting the filters such that the feature maps in GCNN are equivariant under imposed symmetry transformations, e.g., rotation and reflection [104]. In our work, the group convolutional filters are oriented at 0°, 90°, 180°, and 270° such that the feature map and the output (Π^{GCNN}) are rotationally equivariant with respect to the inputs ($\bar{\psi}$ and $\bar{\omega}$), i.e., Eq. (15) in Appendix A.

In addition to the structural modeling approaches mentioned above that achieve rotational equivariance (still with the MSE loss function, Eq. (8)), we can also modify the loss function to combine structural and functional modeling approaches to enhance the performance of CNN in the small-data regime. For example, in 2D turbulence, the SGS enstrophy transfer is critical in maintaining the accuracy and stability of LES [26,45,105]. Therefore, capturing the correct SGS enstrophy transfer $\langle\overline{\omega}\Pi\rangle$ in a CNN can be important for its performance. The *a priori* analysis in Fig. 3 already showed that the error in capturing the global SGS enstrophy transfer ϵ is small in the big-data regime, but can be large in the small-data regime. Here we propose to add a penalty term to the loss function that acts as regularization, enforcing (as a soft constraint) the global SGS enstrophy transfer. A previous study has used similar constraints on energy conservation, derived from the fact that advection terms do not produce or destroy energy [73]. Our global enstrophy transfer constraint is similar in implementation but it is derived directly from the filtered enstrophy equation (see Appendix B), and is used given the importance of inter-scale enstrophy transfer in geophysical turbulence and the fact that here we compute , the forcing on the vorticity equation. This physics-constrained loss function consists of the MSE plus the global SGS enstrophy transfer error:

$$Loss = \frac{1-\beta}{n_{tr}}\sum_{i=1}^{n_{tr}} \|\Pi_i^{\mathrm{CNN}} - \Pi_i^{\mathrm{FDNS}}\|_2^2 + \frac{\beta}{n_{tr}}\sum_{i=1}^{n_{tr}} |\langle\overline{\omega}_i\Pi_i^{\mathrm{CNN}}\rangle - \langle\overline{\omega}_i\Pi_i^{\mathrm{FDNS}}\rangle|, \tag{11}$$

where $\beta \in [0, 1]$ is an adjustable hyperparameter. We empirically find $\beta = 0.5$ to be optimal in minimizing the relative total enstrophy transfer error (ϵ) without significantly affecting c. This physics-constrained loss function (Eq. (11)) synergically combines the structural and functional modeling approaches. The computational efficiency during the training process of the various methods are included in Appendix D.

5. Results

5.1. A priori analysis

A priori analysis is performed using the following metrics: correlation coefficients (c), global enstrophy transfer error (ϵ), and scale-dependent enstrophy and energy transfers (T_Z and T_E, as defined later in this section). Fig. 5 shows the bar plots of c, $c_{T<0}$, $c_{T>0}$, and ϵ for 3 representative cases (1, 3, and 4). In the small-data regime ($n_{tr} = 50$), the use of DA or GCNN increases the correlation coefficients c, $c_{T<0}$, and $c_{T>0}$; the increases are largest for $c_{T>0}$, whose low values could lead to instabilities in *a posteriori* LES, as discussed earlier. The use of DA or GCNN also decreases the relative total enstrophy transfer error ϵ, particularly

[2] The CNN basically predicts that because the red blob of the dipole is now to the left of the blue blob, the part of the Π term corresponding to the red blob in the first row now should be to the right of the part corresponding to the blue blob (see the black box in the third column).

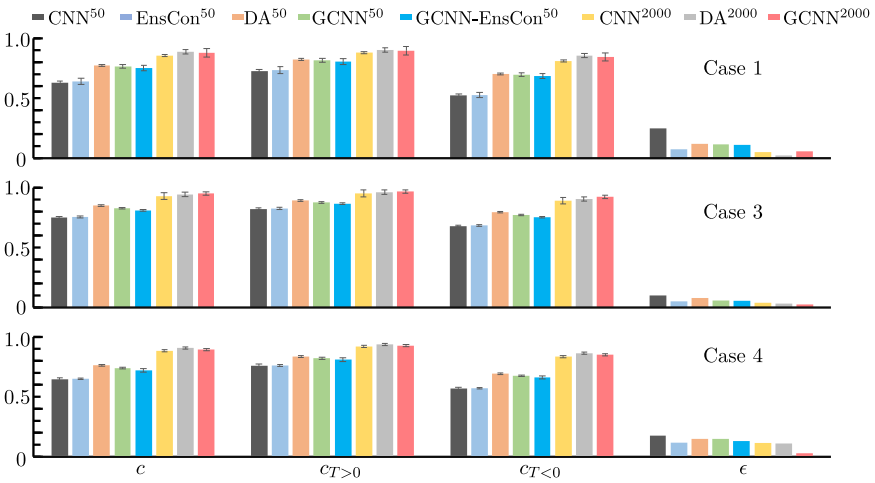

Fig. 5. *A priori* analysis in terms of correlation coefficients c, $c_{T>0}$, and $c_{T>0}$, and the relative enstrophy transfer error ϵ. It is shown that EnsCon does not improve the structural modeling metric c but significantly reduces the functional modeling error, ϵ. DA and GCNN enhance the structural modeling performance and also reduce ϵ. The superscripted number denotes the n_{tr} in the training dataset with $c_{in} < 0.75$ as in Fig. 3. The error bars denote plus and minus one standard deviation (the error bars on ϵ are small and not shown for the sake of clarify).

for Case 1. One point to highlight here is that DA can achieve the same, and in some cases even better, *a priori* accuracy compared to GCNN, while the network architecture is much simpler in DA, which builds equivariance simply in the training data. The EnsCon does not improve the correlation coefficient c (because it only adds a functional modeling component), but as expected, it decreases ϵ, which as shown later improves the *a posteriori* LES. To examine a combined approach, we build an enstrophy-constrained GCNN (GCNN-EnsCon),[3] which performs somewhere in between GCNN and EnsCon: GCNN-EnsCon has a higher c than EnsCon but lower than GCNN, and GCNN-EnsCon has higher ϵ than EnsCon but lower than GCNN. As shown later, the LES with GCNN-EnsCon has the best *a posteriori* performance among all tested models.

To summarize Fig. 5, the physics-constrained CNNs trained on small data ($n_{tr} = 50$) outperform the physics-agnostic CNN trained on small data, but none could outperform the physics-agnostic CNN trained on 40 times more data ($n_{tr} = 2000$) in these *a priori* tests. However, we emphasize that 40 is a substantial factor in terms of the amount of high-fidelity data. This figure also shows that adding physics constraints to the CNN trained in the big-data regime ($n_{tr} = 2000$) does not necessarily lead to any improvement over CNN^{2000}, suggesting that these physics constraints could be learnt by a physics-agnostic CNN from the data given enough training samples. More *a priori* analyses in terms of c and ϵ are included in Appendix C to demonstrate the dependence on n_{tr}.

To further assess the performance of the closures computed using physics-constrained CNNs trained in the small-data regime, we also examine the scale-dependent enstrophy and energy transfers (T_Z and T_E) defined as: [91,106]

$$T_Z(k) = \Re(-\hat{\Pi}_k^* \hat{\bar{\omega}}_k), \tag{12}$$

$$T_E(k) = \Re(-\hat{\Pi}_k^* \hat{\bar{\psi}}_k). \tag{13}$$

Here, $\Re(\cdot)$ means the "real part of", $\hat{(\cdot)}$ denotes Fourier transform, and the asterisks denote complex conjugate. The scale-dependent enstrophy/energy transfer is positive for enstrophy/energy backscatter (enstrophy/energy moving from subgrid scales to resolved scales) and negative for enstrophy/energy forward transfer (enstrophy/energy moving from resolved scales to subgrid scales) [106]. Note that backscatter and forward transfer here are inter-scale transfers by the SGS term (Π), and are different concepts from inverse and forward cascades (discussed earlier in Section 3.2).

Fig. 6 shows the power spectra of $|\hat{\Pi}(k)|^2$, T_Z, and T_E from FDNS and different CNNs, providing further evidence that the incorporating physics constraints improves the *a priori* performance of the data-driven closures. For Case 1 (first row), the $|\hat{\Pi}(k)|^2$ is better predicted by DA and GCNN at the high wavenumbers. The scale-dependent enstrophy forward transfer ($T_Z < 0$ in Fig. 6(b)) is underpredicted by CNN, and the deviation from FDNS is corrected by DA, GCNN, EnsCon, and GCNN-EnsCon. For Case 5 (second row), however, where the inverse energy cascade is important (see Section 3.2), the gain from the physics-constrained CNNs (DA, GCNN, EnsCon, and GCNN-EnsCon) can be seen in the scale-dependent energy transfer (T_E in Fig. 6(f)), where the physics-agnostic CNN incorrectly predicts a portion ($k < 4$) of energy backscatter ($T_E > 0$) to be forward energy transfer ($T_E < 0$).

5.2. A posteriori analysis

Figs. 7–9 show the $\hat{E}(k)$ spectra of Cases 1–5 . In general, the TKE spectrum from LES with the physics-agnostic CNN (denoted by CNN^{50}) matches the one from FDNS at low wavenumbers (large-scale structures) but severely over-predicts the TKE at high wavenumbers (small-scale structures). For example, CNN^{50} starts to deviate from FDNS at $k \approx 20$ for Cases 1, 3, and 4 as shown in Fig. 7 (left) and Fig. 8. This over-prediction can lead to unphysical and unstable numerical results. For example, the vorticity field of LES with CNN^{50} exhibits extensive noisy (i.e., very high-wavenumber) structures in several simulations (not shown). All LES runs with physics-constrained CNNs (DA^{50}, GCNN^{50}, EnsCon^{50}, and GCNN-EnsCon^{50}) outperform the LES with CNN^{50}. In particular, for Cases 2 (Fig. 7 (right)) and 3 (Fig. 8 (left)), the LES runs with DA^{50}, GCNN^{50}, EnsCon^{50}, and GCNN-EnsCon^{50} produce similar TKE spectra which are consistently better than that of the LES with CNN^{50}. For Cases 1 (Fig. 7 (left)) and 4 (Fig. 8 (right)), however, incorporating rotational equivariance

[3] Here, we present the results from GCNN-EnsCon since it is not only a combination of structural and functional modeling (in terms of LES physics [3]) but also a combination of a hard constraint (built into the architecture) and a soft constraint (incorporated in the loss function) in terms of deep learning methodology. We have found that the DA-EnsCon leads to similar *a priori* performance (not shown).

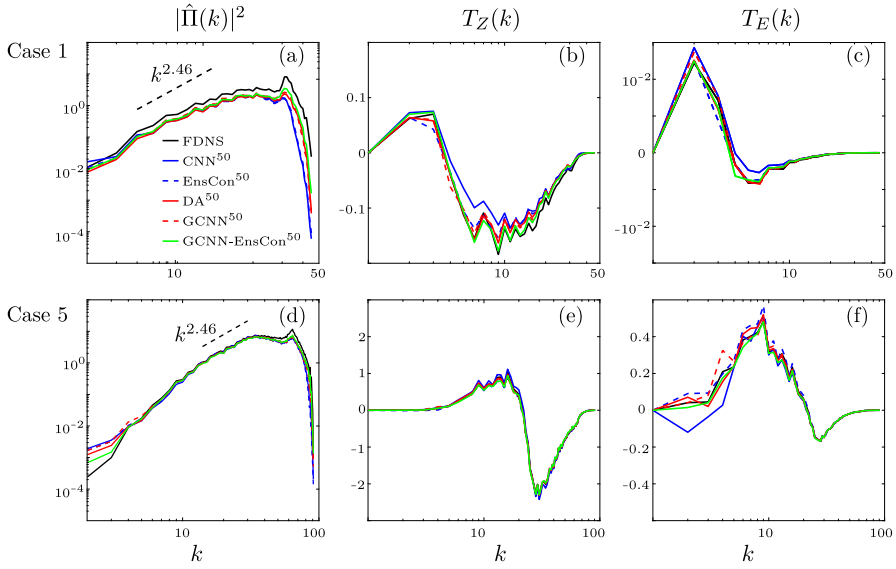

Fig. 6. *A priori* analysis in terms of scale-dependent power spectra $|\hat{\Pi}(k)|^2$, and scale-dependent enstrophy and energy transfers T_Z and T_E for two representative cases (1 and 5). The inertial part of the $|\hat{\Pi}(k)|^2$ spectrum has a slope of 2.46, consistent with previous studies [106,107]. (a)–(c): In Case 1 where the enstrophy direct cascade is important (as discussed in Section 3.2), CNN^{50} does not capture the power spectra correctly at high wavenumbers, and the enstrophy forward transfer ($T_Z < 0$) is under-predicted. (d)–(f): In Case 5 where the energy inverse cascade is important (see Section 3.2), the prediction discrepancy occurs at the low wavenumbers of the power spectra, and at the backscattering part of the energy transfer ($T_E > 0$). In general, the proposed physics-constrained CNNs (DA, GCNN, EnsCon, and GCNN-EnsCon) reduce the prediction error in both structural ($|\hat{\Pi}(k)|^2$) and functional (T_Z and T_E) modelings metrics.

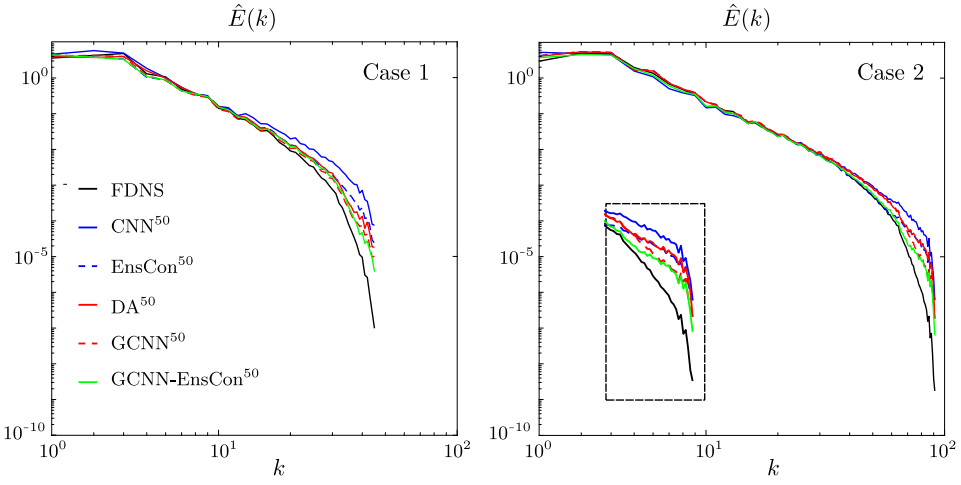

Fig. 7. The TKE spectra $\hat{E}(k)$ of Cases 1 and 2 from *a posteriori* LES run. Results are from long-term LES integrations ($2 \times 10^5 \Delta t_{\text{LES}}$ or $2 \times 10^6 \Delta t_{\text{DNS}}$). The $\hat{E}(k)$ is calculated from 100 randomly chosen snapshots and then averaged. The inset in Case 2 magnifies the tails of the $\hat{E}(k)$ spectra for better visualization. In general, the physics-constrained CNNs (DA^{50}, GCNN^{50}, EnsCon^{50}, or GCNN-EnsCon^{50}) improve the *a posteriori* performance of LES compared to the LES with the physics-agnostic CNN^{50}: The spectra from the LES with physics-constrained CNNs better match the FDNS spectra especially at the tails (high-wavenumber structures). In particular, the spectra from the LES with GCNN-EnsCon^{50} have the best match with the FDNS spectra especially at the high wavenumbers. The improvement is more prominent for the coarser-grid LES (Case 1, $N_{LES} = 64$, compared to Case 2, $N_{LES} = 128$).

(through DA, GCNN, or GCNN-EnsCon) leads to better *a posteriori* performances than incorporating the global enstrophy constraint alone (EnsCon) in terms of matching the FDNS spectra. Overall, the LES with GCNN-EnsCon^{50} has the best performance in these 4 cases, showing the advantage of combining different types of physics constraints in the small-data regime.

In Case 5, the gain from the physics constraints is less obvious from the TKE spectra, although a slight improvement at the tails can still be observed (Fig. 9 (left)). In this case, the PDF of vorticity (Fig. 9 (right)) better reveals the gain, where LES with CNN^{50} predicts spuriously large vorticity extremes due to the excessive high-wavenumber structures in the vorticity field. The physics-constrained CNNs (DA^{50}, GCNN^{50}, EnsCon^{50}, and GCNN-EnsCon^{50}) result in a stable and more accurate LES as the TKE spectrum and PDF of vorticity better match those of the FDNS. The effects of the enstrophy constraints can further be observed in the time evolution of the enstrophy and total kinetic energy (see Fig. 10). CNN^{50} predicts spuriously large enstrophy and total energy over time. GCNN^{50} also exhibits a slight increase in the total energy but shows an improvement in performance compared to CNN^{50}. DA^{50} and the enstrophy-constrained CNNs (EnsCon^{50}, GCNN-EnsCon^{50}, DA-EnsCon^{50}) all result in evolutions close to that of the DNS.

6. Summary and discussion

The objective of this paper is to learn CNN-based non-local SGS closures from filtered DNS data for stable and accurate LES, with a focus on the *small-data regime*, i.e., when the available DNS

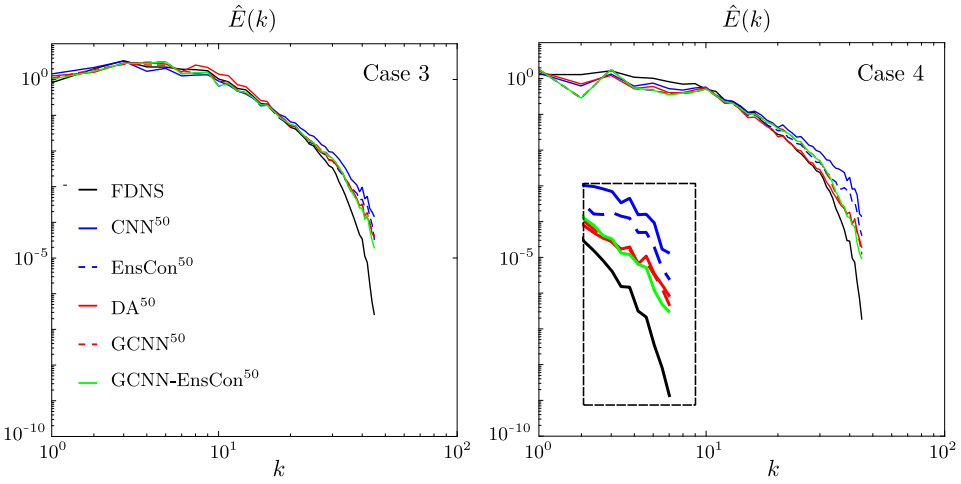

Fig. 8. Same as Fig. 7 but for Cases 3 and 4. Similar to the finding of Fig. 7, the physics-constrained CNNs (DA^{50}, GCNN^{50}, EnsCon^{50}, or GCNN-EnsCon^{50}) improve the *a posteriori* performance of LES compared to the LES with the physics-agnostic CNN^{50}. For Case 3, the improvement can only be observed at the highest wavenumber. For Case 4, however, incorporating the rotational equivariance (DA^{50}, GCNN^{50}, and GCNN-EnsCon^{50}) leads to a more accurate LES than the enstrophy constraint alone (EnsCon^{50}) alone. Also similar to the finding of Fig. 7, the spectra from the LES with GCNN-EnsCon^{50} have the best match with the FDNS spectra especially at the high wavenumbers.

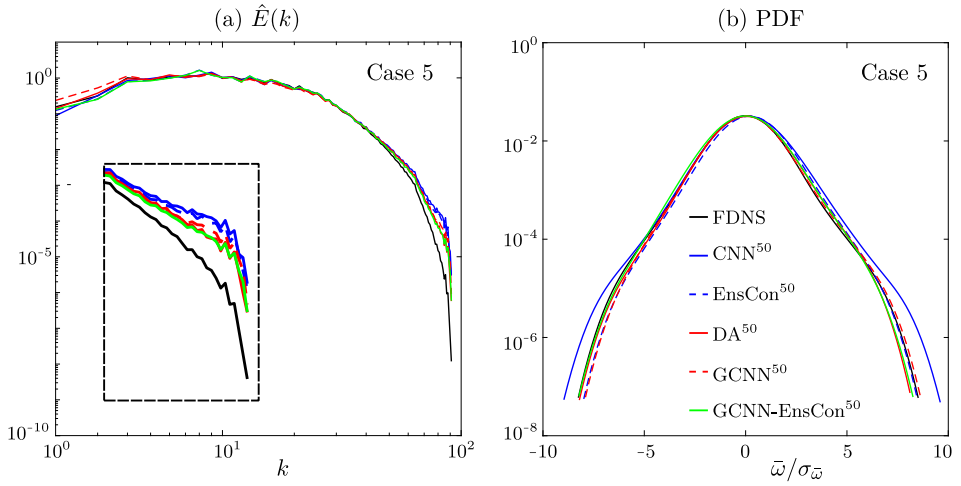

Fig. 9. Same as Fig. 7 but showing (a) The TKE spectra $\hat{E}(k)$ and (b) probability density function (PDF) of $\overline{\omega}$ for Case 5. Although, the physics-constrained CNNs (DA^{50}, GCNN^{50}, EnsCon^{50}, or GCNN-EnsCon^{50}) result in slightly improved LES in terms of the TKE spectrum, the gain from the physical constraints can be observed more clearly in the PDF of vorticity where the LES with CNN^{50} over-predicts the extreme values (tails of the PDF).

training set is small. We demonstrate that incorporating physics constraints into the CNN using three methods can substantially improve the *a priori* (offline) and *a posteriori* (online) performance of the data-driven closure model in the small-data regime. While physics-constrained deep learning-based SGS modeling has been recently explored in a couple of studies [e.g., 73,77], a systematic analysis of the improvements in stability and accuracy, particularly in the small-data regime, is still needed. This is the regime most relevant to many real-world applications in science and engineering.

In this paper, our contributions are: (1) illustrating the deficiency of physics-agnostic CNNs in the small-data regime through analyzing the rotational equivariance of the SGS term in a dipole; (2) demonstrating that the small- vs big-data regime depends not only on the number of training samples but also on their inter-sample correlations; (3) identifying the reasons behind the instabilities/inaccuracies through analyzing the global inter-scale enstrophy transfer in 2D turbulence; and (4) proposing three physics-constrained CNNs that achieve comparable *a posteriori* performances with only 1/40 of the big training data. We show that while with the big training set, even a physics-agnostic CNN yields long-term stable/accurate *a posteriori* performance, with the small training set, adding physics-constraints are essential for stability/accuracy. To the best of our knowledge, this is the first systematic analysis of the small-data regime and physics-constrained CNNs for LES modeling.

To summarize the results: We use 5 different forced 2D homogeneous isotropic turbulence (HIT) flows with various forcing wavenumbers, linear drag coefficients, and LES grid sizes as the testbeds. First, we show in Section 3 that in the "big-data" regime (with $n_{tr} = 2000$ weakly correlated training samples), the LES with physics-agnostic CNN is stable and accurate, and outperforms the LES with the physics-based DSMAG closure, particularly as the data-driven closure captures backscattering well (see Fig. 2 and Ref. [26]). Next, we show, using *a priori* (offline) and *a posteriori* (online) tests, that the performance of the physics-agnostic CNNs substantially deteriorate when they are trained in the "small-data" regime: with $n_{tr} = 2000$ highly correlated samples or with $n_{tr} = 50$ weakly correlated samples. This analysis demonstrates that the small versus big data regime depends not only on the number of training samples but also on their inter correlations.

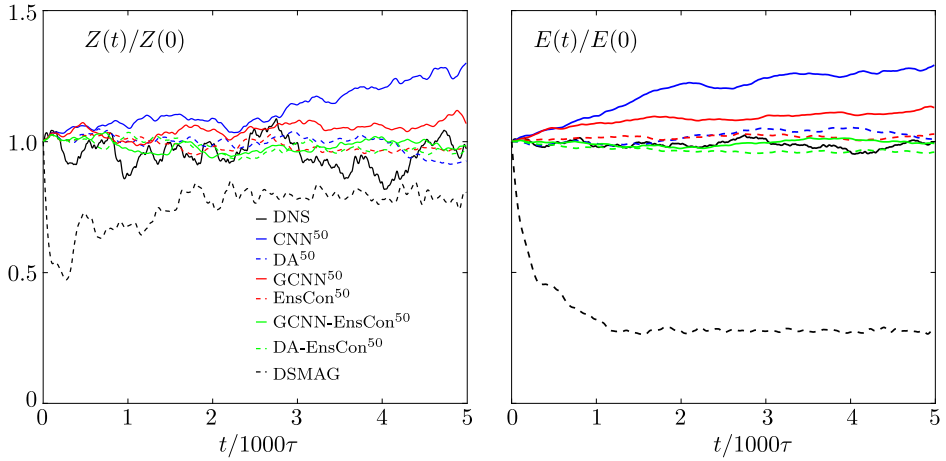

Fig. 10. Time evolution of the enstrophy $Z = \frac{\langle\omega^2\rangle}{2}$ and total energy $E = \frac{\langle\psi\omega\rangle}{2}$ of Case 5, normalized by the LES initial condition, $Z(0)$ and $E(0)$, respectively. Examining both $Z(t)$ and $E(t)$ shows that in the small-data regime, CNN^{50} leads to a gradual increase, indicating excessive noise in the vorticity field. GCNN^{50}, despite its improvement, also shows a slight increase in both $Z(t)$ and $E(t)$. All the other methods, especially those with enstrophy constraints, lead to $Z(t)/Z(0) \approx 1.0$ and $E(t)/E(0) \approx 1.0$, which is also the case for DNS, as the LES initial condition is already in the statistical equilibrium regime. DSMAG, due to its excessive diffusivity and lack of backscattering, leads to a significant decrease in both $Z(t)$ and $E(t)$. Here, the time series is scaled by the eddy-turnover time $\tau = 1/\omega_{\max}$, where $\omega_{\max}$ is the maximum vorticity in DNS.

To improve the performance of CNNs trained in the small-data regime, in Section 4, we propose incorporating physics in the CNNs through using 1) data augmentation (DA), 2) a group equivariant CNN (GCNN), or an enstrophy-constrained loss function (EnsCon). The idea behind using DA and GCNN is to account for the rotational equivariance of the SGS term. This is inspired by a simple example of a vortex dipole, which shows that for never-seen-before samples, the physics-agnostic CNN can only capture the translational equivariance, but not the rotational equivariance, another important property of the SGS term. The idea behind EnsCon is to combine structural and functional modeling approaches through a regularized loss function. *A priori* and *a posteriori* tests show that all these physics-constrained CNNs outperform the physics-agnostic CNN in the small-data regime ($n_{tr} = 50$). GCNN, which uses an equivariance-preserving architecture, improves the data-driven SGS closures. However, it should be mentioned that DA, which simply builds equivariance in the training samples and can be used with any architecture, shows comparable or even in some cases, better performance than GCNN. Also, note that a recent study suggested using an augmented GCNN which preserves rotational equivariance by multiples of $2\pi/m$ ($m > 4$), in addition to multiples of $\pi/2$, as in our GCNN [77]. This can potentially further enhance the performance of GCNN, though DA can also incorporate such equivariance in the training samples as well.

Overall, GCNN+EnsCon, which combines these two main constraints, demonstrate the best *a posteriori* performance, showing the advantage of adding physics constraints together. Note that here we focus on rotational equivariance, which is a property of the 2D turbulence test case. In other flows, other equivariance properties might exist (e.g., reflection equivariance as in Rayleigh Bénard convection), and they can be incorporated through DA or GCNN as needed. These results show the major advantage and potential of physics-constrained deep learning methods for SGS modeling in the small-data regime, which is of substantial importance for complex and high-Reynolds number flows, for which the availability of high-fidelity (e.g., DNS) data could be severely limited.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Pedram Hassanzadeh reports financial support was provided by Office of Naval Research. Pedram Hassanzadeh reports financial support was provided by National Science Foundation. Pedram Hassanzadeh reports financial support was provided by Schmidt Futures.

Data availability

Part of our codes and data are available at https://github.com/envfluids/2D-DDP. Other Data will be made available on request.

Acknowledgments

This work was supported by an award from the ONR Young Investigator Program (N00014-20-1-2722), a grant from the NSF CSSI program (OAC-2005123), and by the generosity of Eric and Wendy Schmidt by recommendation of the Schmidt Futures program. Computational resources were provided by NSF XSEDE (allocation ATM170020) and NCAR's CISL (allocation URIC0004). Our codes and data are available at https://github.com/envfluids/2D-DDP.

Appendix A. Equivariance properties of the SGS term

According to the transformation properties of the Navier–Stokes equations [2], the SGS term Π should satisfy:

$$\Pi(T_g\omega, T_g\psi) = T_g\Pi(\omega, \psi), \tag{14}$$

where T_g represents a translational or rotational transformation. ω and ψ are the vorticity and streamfunction, respectively (as described in Section 2).

In the ML literature, "equivariance" means that transforming an input (e.g., by translation or rotation, denoted by T_g) and then passing the transformed input through the learnt map (CNN in our case) should give the same result as first mapping the input and then transforming the output [102,104]:

$$\Pi^{\text{CNN}}(T_g\bar{\omega}, T_g\bar{\psi}, \theta) = T_g\Pi^{\text{CNN}}(\bar{\omega}, \bar{\psi}, \theta). \tag{15}$$

Here, θ represents a group of learnable parameters of the network. To preserve the translational and rotational equivariance of Π (Eq. (14)), the network parameters θ should be learnt such

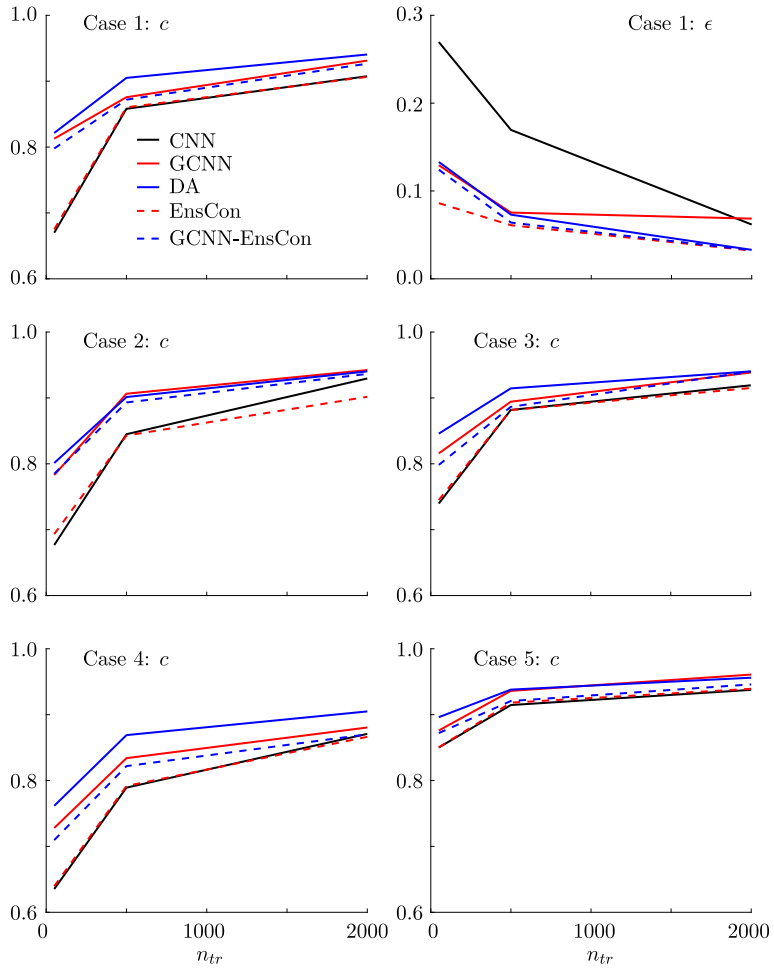

Fig. 11. *A priori* analysis of all CNNs on data dependency. n_{tr} represents the number of decorrelated samples in the training set. The correlation coefficients c show that the CNNs with rotational equivariance enhance the structural performance where the enstrophy constraint does not help. On the other hand, the enstrophy constraints (EnsCon and GCNN-EnsCon) reduce the enstrophy transfer error ϵ.

that Eq. (15) is satisfied. In turbulence modeling, "equivariance" may also be referred to as "symmetry" [2,80]. In this paper, we use the term "equivariance", and use an equivariance-preserving network (GCNN) or build this property into the training via DA.

Appendix B. The global enstrophy-transfer constraint

The equation for enstrophy transfer can be obtained by first multiplying the filtered equation (Eq. (4a)) by $\overline{\omega}$:

$$\overline{\omega}\frac{\partial\overline{\omega}}{\partial t}+\overline{\omega}\mathcal{N}(\overline{\omega},\overline{\psi})=\overline{\omega}\frac{1}{Re}\nabla^2\overline{\omega}-\overline{\omega}\overline{f}-r\overline{\omega}^2+\underbrace{\overline{\omega}\mathcal{N}(\overline{\omega},\overline{\psi})-\overline{\omega}\overline{\mathcal{N}(\omega,\psi)}}_{\overline{\omega}\Pi}. \tag{16}$$

Rearranging Eq. (16) gives:

$$\frac{1}{2}\frac{\partial\overline{\omega}^2}{\partial t}+\frac{1}{2}\mathcal{N}(\overline{\omega}^2,\overline{\psi})=\frac{1}{Re}\left[\frac{1}{2}\nabla^2\overline{\omega}^2-(\nabla\overline{\omega})^2\right]-\overline{\omega}\overline{f}-r\overline{\omega}^2+\overline{\omega}\Pi. \tag{17}$$

The evolution equation for domain-averaged enstrophy $Z=\langle\frac{1}{2}\overline{\omega}^2\rangle$ is then obtained by domain averaging Eq. (17) and invoking the domain's periodicity [108]:

$$\frac{dZ}{dt}=-\frac{1}{Re}\langle(\nabla\overline{\omega})^2\rangle-\langle\overline{\omega}\overline{f}\rangle-2rZ+\langle\overline{\omega}\Pi\rangle. \tag{18}$$

Therefore, the domain-averaged enstrophy transfer due to the SGS term is $\langle\overline{\omega}\Pi\rangle$. In Eq. (11), we enforce $\langle\overline{\omega}\Pi\rangle$ predicted by the CNN to be close to that of the FDNS as a domain-averaged (global) soft constraint.

Appendix C. *A priori* analysis on data-size dependency

Fig. 11 shows the c for a variety of CNNs (CNN, GCNN, DA, EnsCon, and GCNN-EnsCon) trained with different number of n_{tr} for Cases 1–5, and ϵ for Case 1. It can be observed that the rotational equivariance constraint enhances the structural performance in terms of c but the enstrophy constraint does not show improvement. For the global enstrophy-transfer error ϵ, the EnsCon and GCNN-EnsCon reduce the *a priori* loss.

Appendix D. Training cost

The normalized computational times of various CNNs for the training process are shown in Table 1. It can be seen that the GCNN and GCNN-EnsCon are the most costly approaches during the training process, possibly due to the extra weights and weights-sharing as well as the permutation processes. The DA is precisely four times more expensive to train compared to CNN as DA's training data set is expanded by a factor of 4. The extra global enstrophy transfer loss adds a minimal amount of computational time to the training. Note that during inference, the computational costs of these methods are the same.

Table 1
The computational time of the training process for $n_{tr} = 2000$ normalized by the that of baseline CNN.

CNN	GCNN	DA	EnsCon	GCNN-EnsCon
1.00	5.58	4.00	1.07	5.75

References

[1] J. Smagorinsky, General circulation experiments with the primitive equations: I. The basic experiment, Mon. Weather Rev. 91 (3) (1963) 99–164.
[2] S.B. Pope, Turbulent Flows, IOP Publishing, 2001.
[3] P. Sagaut, Large Eddy Simulation for Incompressible Flows: An Introduction, Springer Science & Business Media, 2006.
[4] R.D. Moser, S.W. Haering, G.R. Yalla, Statistical properties of subgrid-scale turbulence models, Annu. Rev. Fluid Mech. 53 (2020).
[5] E. Bou-Zeid, C. Meneveau, M. Parlange, A scale-dependent Lagrangian dynamic model for large eddy simulation of complex turbulent flows, Phys. Fluids 17 (2) (2005) 025105.
[6] P. Sagaut, M. Terracol, S. Deck, Multiscale and Multiresolution Approaches in Turbulence-LES, DES and Hybrid RANS/LES Methods: Applications and Guidelines, World Scientific, 2013.
[7] T. Schneider, S. Lan, A. Stuart, J. Teixeira, Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations, Geophys. Res. Lett. 44 (24) (2017) 12–396.
[8] L. Zanna, T. Bolton, Deep learning of unresolved turbulent ocean processes in climate models, in: Deep Learning for the Earth Sciences: A Comprehensive Approach To Remote Sensing, Climate Science, and Geosciences, Wiley Online Library, 2021, pp. 298–306.
[9] T. Schneider, A.M. Stuart, J.-L. Wu, Ensemble Kalman inversion for sparse learning of dynamical systems from time-averaged data, 2020, arXiv preprint arXiv:2007.06175.
[10] T. Schneider, A.M. Stuart, J.-L. Wu, Imposing sparsity within ensemble Kalman inversion, 2020, arXiv preprint arXiv:2007.06175.
[11] T. Schneider, A.M. Stuart, J.-L. Wu, Learning stochastic closures using ensemble Kalman inversion, Trans. Math. Appl. 5 (1) (2021) tnab003.
[12] H.J. Bae, P. Koumoutsakos, Scientific multi-agent reinforcement learning for wall-models of turbulent flows, Nature Commun. 13 (1) (2022) 1–9.
[13] G. Vissio, V. Lucarini, A proof of concept for scale-adaptive parametrizations: the case of the Lorenz'96 model, Q. J. R. Meteorol. Soc. 144 (710) (2018) 63–75.
[14] C.L. Franzke, T.J. O'Kane, J. Berner, P.D. Williams, V. Lucarini, Stochastic climate theory and modeling, Wiley Interdiscip. Rev. Clim. Change 6 (1) (2015) 63–78.
[15] A. Leonard, Energy cascade in large-eddy simulations of turbulent fluid flows, in: Advances in Geophysics, vol. 18, Elsevier, 1975, pp. 237–248.
[16] R.A. Clark, J.H. Ferziger, W.C. Reynolds, Evaluation of subgrid-scale models using an accurately simulated turbulent flow, J. Fluid Mech. 91 (1) (1979) 1–16.
[17] C. Meneveau, J. Katz, Scale-invariance and turbulence models for large-eddy simulation, Annu. Rev. Fluid Mech. 32 (1) (2000) 1–32.
[18] H. Lu, F. Porté-Agel, A modulated gradient model for scalar transport in large-eddy simulation of the atmospheric boundary layer, Phys. Fluids 25 (1) (2013) 015110.
[19] A. Vollant, G. Balarac, C. Corre, A dynamic regularized gradient model of the subgrid-scale stress tensor for large-eddy simulation, Phys. Fluids 28 (2) (2016) 025114.
[20] Y. Wang, Z. Yuan, C. Xie, J. Wang, A dynamic spatial gradient model for the subgrid closure in large-eddy simulation of turbulence, Phys. Fluids 33 (7) (2021) 075119.
[21] Z. Yuan, Y. Wang, C. Xie, J. Wang, Dynamic iterative approximate deconvolution models for large-eddy simulation of turbulence, Phys. Fluids 33 (8) (2021) 085125.
[22] M. Germano, U. Piomelli, P. Moin, W.H. Cabot, A dynamic subgrid-scale eddy viscosity model, Phys. Fluids A 3 (7) (1991) 1760–1765.
[23] D.K. Lilly, A proposed modification of the Germano subgrid-scale closure method, Phys. Fluids A 4 (3) (1992) 633–635.
[24] Y. Zang, R.L. Street, J.R. Koseff, A dynamic mixed subgrid-scale model and its application to turbulent recirculating flows, Phys. Fluids A 5 (12) (1993) 3186–3196.
[25] S. Pawar, O. San, A. Rasheed, P. Vedula, A priori analysis on deep learning of subgrid-scale parameterizations for Kraichnan turbulence, Theor. Comput. Fluid Dyn. (2020) 1–27.
[26] Y. Guan, A. Chattopadhyay, A. Subel, P. Hassanzadeh, Stable a posteriori LES of 2D turbulence using convolutional neural networks: Backscattering analysis and generalization to higher Re via transfer learning, J. Comput. Phys. 458 (2022) 111090.
[27] Z. Wang, K. Luo, D. Li, J. Tan, J. Fan, Investigations of data-driven closure for subgrid-scale stress in large-eddy simulation, Phys. Fluids 30 (12) (2018) 125101.
[28] Z. Zhou, G. He, S. Wang, G. Jin, Subgrid-scale model for large-eddy simulation of isotropic turbulent flows using an artificial neural network, Comput. & Fluids 195 (2019) 104319.
[29] J.-X. Wang, J.-L. Wu, H. Xiao, Physics-informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data, Phys. Rev. Fluids 2 (3) (2017) 034603.
[30] K. Duraisamy, G. Iaccarino, H. Xiao, Turbulence modeling in the age of data, Annu. Rev. Fluid Mech. 51 (2019) 357–377.
[31] R. Maulik, O. San, J.D. Jacob, C. Crick, Sub-grid scale model classification and blending through deep learning, J. Fluid Mech. 870 (2019) 784–812, http://dx.doi.org/10.1017/jfm.2019.254.
[32] A. Beck, D. Flad, C.-D. Munz, Deep neural networks for data-driven LES closure models, J. Comput. Phys. 398 (2019) 108910.
[33] A. Prat, T. Sautory, S. Navarro-Martinez, A priori sub-grid modelling using artificial neural networks, Int. J. Comput. Fluid Dyn. (2020) 1–21.
[34] S. Taghizadeh, F.D. Witherden, S.S. Girimaji, Turbulence closure modeling with data-driven techniques: physical compatibility and consistency considerations, New J. Phys. 22 (9) (2020) 093023.
[35] L. Zanna, T. Bolton, Data-driven equation discovery of ocean mesoscale closures, Geophys. Res. Lett. 47 (17) (2020) e2020GL088376.
[36] S.L. Brunton, B.R. Noack, P. Koumoutsakos, Machine learning for fluid mechanics, Annu. Rev. Fluid Mech. 52 (2020) 477–508.
[37] A. Beck, M. Kurz, A perspective on machine learning methods in turbulence modeling, GAMM-Mitt. 44 (1) (2021) e202100002.
[38] K. Duraisamy, Perspectives on machine learning-augmented Reynolds-averaged and large eddy simulation models of turbulence, Phys. Rev. Fluids 6 (5) (2021) 050504.
[39] N. Moriya, K. Fukami, Y. Nabae, M. Morimoto, T. Nakamura, K. Fukagata, Inserting machine-learned virtual wall velocity for large-eddy simulation of turbulent channel flows, 2021, arXiv preprint arXiv:2106.09271.
[40] G.D. Portwood, B.T. Nadiga, J.A. Saenz, D. Livescu, Interpreting neural network models of residual scalar flux, J. Fluid Mech. 907 (2021).
[41] R. Stoffer, C.M. Van Leeuwen, D. Podareanu, V. Codreanu, M.A. Veerman, M. Janssens, O.K. Hartogensis, C.C. Van Heerwaarden, Development of a large-eddy simulation subgrid model based on artificial neural networks: a case study of turbulent channel flow, Geosci. Model Dev. 14 (6) (2021) 3769–3788.
[42] B. Liu, H. Yu, H. Huang, X.-Y. Lu, Investigation of nonlocal data-driven methods for subgrid-scale stress modelling in large eddy simulation, 2021, arXiv preprint arXiv:2109.01292.
[43] C. Jiang, R. Vinuesa, R. Chen, J. Mi, S. Laima, H. Li, An interpretable framework of data-driven turbulence modeling using deep neural networks, Phys. Fluids 33 (5) (2021) 055133.
[44] Y. Tian, D. Livescu, M. Chertkov, Physics-informed machine learning of the Lagrangian dynamics of velocity gradient tensor, Phys. Rev. Fluids 6 (9) (2021) 094607.
[45] R. Maulik, O. San, A. Rasheed, P. Vedula, Subgrid modelling for two-dimensional turbulence using neural networks, J. Fluid Mech. 858 (2019) 122–144.
[46] C. Xie, K. Li, C. Ma, J. Wang, Modeling subgrid-scale force and divergence of heat flux of compressible isotropic turbulence by artificial neural network, Phys. Rev. Fluids 4 (10) (2019) 104605.
[47] C. Xie, J. Wang, K. Li, C. Ma, Artificial neural network approach to large-eddy simulation of compressible isotropic turbulence, Phys. Rev. E 99 (5) (2019) 053113.
[48] C. Xie, J. Wang, H. Li, M. Wan, S. Chen, Artificial neural network mixed model for large eddy simulation of compressible isotropic turbulence, Phys. Fluids 31 (8) (2019) 085112.
[49] C. Xie, X. Xiong, J. Wang, Artificial neural network approach for turbulence models: A local framework, Phys. Rev. Fluids 6 (8) (2021) 084612.
[50] Y. Wang, Z. Yuan, C. Xie, J. Wang, Artificial neural network-based spatial gradient models for large-eddy simulation of turbulence, AIP Adv. 11 (5) (2021) 055216.
[51] R. Maulik, H. Sharma, S. Patel, B. Lusch, E. Jennings, A turbulent eddy-viscosity surrogate modeling framework for Reynolds-Averaged Navier-Stokes simulations, Comput. & Fluids 227 (2021) 104777.
[52] T. Bolton, L. Zanna, Applications of deep learning to ocean data inference and subgrid parameterization, J. Adv. Modelling Earth Syst. 11 (1) (2019) 376–399.
[53] M. Kurz, A. Beck, A machine learning framework for LES closure terms, Electron. Trans. Numer. Anal. 56 (2022) 117–137.
[54] A. Subel, A. Chattopadhyay, Y. Guan, P. Hassanzadeh, Data-driven subgrid-scale modeling of forced Burgers turbulence using deep learning with generalization to higher Reynolds numbers via transfer learning, Phys. Fluids 33 (3) (2021) 031702.

[55] U. Achatz, G. Schmitz, On the closure problem in the reduction of complex atmospheric models by PIPs and EOFs: A comparison for the case of a two-layer model with zonally symmetric forcing, J. Atmos. Sci. 54 (20) (1997) 2452–2474.

[56] J.-C. Loiseau, S.L. Brunton, Constrained sparse Galerkin regression, J. Fluid Mech. 838 (2018) 42–67.

[57] Z.Y. Wan, P. Vlachas, P. Koumoutsakos, T. Sapsis, Data-assisted reduced-order modeling of extreme events in complex dynamical systems, PLoS One 13 (5) (2018) e0197704.

[58] Y. Guan, S.L. Brunton, I. Novosselov, Sparse nonlinear models of chaotic electroconvection, R. Soc. Open Sci. 8 (2021) 202367.

[59] A.A. Kaptanoglu, K.D. Morgan, C.J. Hansen, S.L. Brunton, Physics-constrained, low-dimensional models for magnetohydrodynamics: First-principles and data-driven approaches, Phys. Rev. E 104 (1) (2021) 015206.

[60] R. Vinuesa, S.L. Brunton, The potential of machine learning to enhance computational fluid dynamics, 2021, arXiv preprint arXiv:2110.02085.

[61] M. Khodkar, P. Hassanzadeh, A data-driven, physics-informed framework for forecasting the spatiotemporal evolution of chaotic dynamics with nonlinearities modeled as exogenous forcings, J. Comput. Phys. 440 (2021) 110412.

[62] R. Maulik, V. Rao, J. Wang, G. Mengaldo, E. Constantinescu, B. Lusch, P. Balaprakash, I. Foster, R. Kotamarthi, AIEADA 1.0: Efficient high-dimensional variational data assimilation with machine-learned reduced-order models, Geosci. Model Dev. Discuss. (2022) 1–20.

[63] R. Mojgani, M. Balajewicz, P. Hassanzadeh, Lagrangian PINNs: A causality-conforming solution to failure modes of physics-informed neural networks, 2022, arXiv preprint arXiv:2205.02902.

[64] R. Mojgani, A. Chattopadhyay, P. Hassanzadeh, Closed-form discovery of structural errors in models of chaotic systems by integrating Bayesian sparse regression and data assimilation, 2021, arXiv preprint arXiv:2110.00546.

[65] M. Santos Gutiérrez, V. Lucarini, M.D. Chekroun, M. Ghil, Reduced-order models for coupled dynamical systems: Data-driven methods and the Koopman operator, Chaos 31 (5) (2021) 053116.

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

[67] R. King, O. Hennigh, A. Mohan, M. Chertkov, From deep to physics-informed learning of turbulence: Diagnostics, 2018, arXiv preprint arXiv:1810.07785.

[68] R. Sharma, A.B. Farimani, J. Gomes, P. Eastman, V. Pande, Weakly-supervised deep learning of heat transport via physics informed loss, 2018, arXiv preprint arXiv:1807.11374.

[69] S. Pan, K. Duraisamy, Long-time predictive modeling of nonlinear dynamical systems using neural networks, Complexity 2018 (2018).

[70] A. Chattopadhyay, E. Nabizadeh, P. Hassanzadeh, Analog forecasting of extreme-causing weather patterns using deep learning, J. Adv. Modelling Earth Syst. 12 (2) (2020) e2019MS001958.

[71] K. Meidani, A.B. Farimani, Data-driven identification of 2D partial differential equations using extracted physical features, Comput. Methods Appl. Mech. Engrg. 381 (2021) 113831.

[72] T. Beucler, M. Pritchard, S. Rasp, J. Ott, P. Baldi, P. Gentine, Enforcing analytic constraints in neural networks emulating physical systems, Phys. Rev. Lett. 126 (9) (2021) 098302.

[73] A.-T.G. Charalampopoulos, T.P. Sapsis, Machine-learning energy-preserving nonlocal closures for turbulent fluid flows and inertial tracers, Phys. Rev. Fluids 7 (2) (2022) 024305.

[74] A. Prakash, K.E. Jansen, J.A. Evans, Invariant data-driven subgrid stress modeling in the strain-rate eigenframe for large eddy simulation, 2021, arXiv preprint arXiv:2106.13410.

[75] C. Yan, H. Li, Y. Zhang, H. Chen, Data-driven turbulence modeling in separated flows considering physical mechanism analysis, 2021, arXiv preprint arXiv:2109.09095.

[76] R. Magar, Y. Wang, C. Lorsung, C. Liang, H. Ramasubramanian, P. Li, A.B. Farimani, AugLiChem: Data augmentation library of chemical structures for machine learning, 2021, arXiv preprint arXiv:2111.15112.

[77] S. Pawar, O. San, A. Rasheed, P. Vedula, Frame invariant neural network closures for kraichnan turbulence, 2022, arXiv preprint arXiv:2201.02928.

[78] K. Kashinath, M. Mustafa, A. Albert, J. Wu, C. Jiang, S. Esmaeilzadeh, K. Azizzadenesheli, R. Wang, A. Chattopadhyay, A. Singh, et al., Physics-informed machine learning: case studies for weather and climate modelling, Phil. Trans. R. Soc. A 379 (2194) (2021) 20200093.

[79] G.E. Karniadakis, I.G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, L. Yang, Physics-informed machine learning, Nat. Rev. Phys. 3 (6) (2021) 422–440.

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

[81] H. Frezat, G. Balarac, J. Le Sommer, R. Fablet, R. Lguensat, Physical invariance in neural networks for subgrid-scale scalar flux modeling, Phys. Rev. Fluids 6 (2) (2021) 024607.

[82] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J. Comput. Phys. 378 (2019) 686–707.

[83] Y. Zhu, N. Zabaras, P.-S. Koutsourelakis, P. Perdikaris, Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data, J. Comput. Phys. 394 (2019) 56–81.

[84] J.-L. Wu, K. Kashinath, A. Albert, D. Chirila, H. Xiao, et al., Enforcing statistical constraints in generative adversarial networks for modeling chaotic dynamical systems, J. Comput. Phys. 406 (2020) 109209.

[85] A.T. Mohan, N. Lubbers, D. Livescu, M. Chertkov, Embedding hard physical constraints in neural network coarse-graining of 3D turbulence, 2020, arXiv preprint arXiv:2002.00021.

[86] A. Chattopadhyay, M. Mustafa, P. Hassanzadeh, K. Kashinath, Deep spatial transformers for autoregressive data-driven forecasting of geophysical turbulence, in: Proceedings of the 10th International Conference on Climate Informatics, 2020, pp. 106–112.

[87] R. Wang, R. Walters, R. Yu, Incorporating symmetry into deep dynamics models for improved generalization, 2020, arXiv preprint arXiv:2002.03061.

[88] A. Chattopadhyay, M. Mustafa, P. Hassanzadeh, E. Bach, K. Kashinath, Towards physically consistent data-driven weather forecasting: Integrating data assimilation with equivariance-preserving spatial transformers in a case study with ERA5, Geosci. Model Dev. Discuss. (2021) 1–23.

[89] G.K. Vallis, Atmospheric and Oceanic Fluid Dynamics, Cambridge University Press, 2017.

[90] G.J. Chandler, R.R. Kerswell, Invariant recurrent solutions embedded in a turbulent two-dimensional Kolmogorov flow, J. Fluid Mech. 722 (2013) 554–595.

[91] J. Thuburn, J. Kent, N. Wood, Cascades, backscatter and conservation in numerical models of two-dimensional turbulence, Q. J. R. Meteorol. Soc. 140 (679) (2014) 626–638.

[92] W.T. Verkley, C.A. Severijns, B.A. Zwaal, A maximum entropy approach to the interaction between small and large scales in two-dimensional turbulence, Q. J. R. Meteorol. Soc. 145 (722) (2019) 2221–2236.

[93] D. Kochkov, J.A. Smith, A. Alieva, Q. Wang, M.P. Brenner, S. Hoyer, Machine learning–accelerated computational fluid dynamics, Proc. Natl. Acad. Sci. 118 (21) (2021).

[94] A. Alexakis, C.R. Doering, Energy and enstrophy dissipation in steady state 2D turbulence, Phys. Lett. A 359 (6) (2006) 652–657.

[95] C.R. Doering, J.D. Gibbon, Note on the constantin-foias-temam attractor dimension estimate for two-dimensional turbulence, Physica D 48 (2–3) (1991) 471–480.

[96] A. Subel, Y. Guan, A. Chattopadhyay, P. Hassanzadeh, Explaining the physics of transfer learning a data-driven subgrid-scale closure to a different turbulent flow, 2022, arXiv preprint arXiv:2206.03198.

[97] R.R. Wilcox, Fundamentals of Modern Statistical Methods: Substantially Improving Power and Accuracy, Springer, 2010.

[98] G.K. Batchelor, Computation of the energy spectrum in homogeneous two-dimensional turbulence, Phys. Fluids 12 (12) (1969) II–233.

[99] R.H. Kraichnan, Inertial ranges in two-dimensional turbulence, Phys. Fluids 10 (7) (1967) 1417–1423.

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

[101] P.A. Perezhogin, A.V. Glazunov, A.S. Gritsun, Stochastic and deterministic kinetic energy backscatter parameterizations for simulation of the two-dimensional turbulence, Russian J. Numer. Anal. Math. Modelling 34 (4) (2019) 197–213.

[102] T. Cohen, M. Welling, Group equivariant convolutional networks, in: International Conference on Machine Learning, PMLR, 2016, pp. 2990–2999.

[103] M.M. Bronstein, J. Bruna, T. Cohen, P. Veličković, Geometric deep learning: Grids, groups, graphs, geodesics, and gauges, 2021, arXiv preprint arXiv:2104.13478.

[104] B.S. Veeling, J. Linmans, J. Winkens, T. Cohen, M. Welling, Rotation equivariant CNNs for digital pathology, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 210–218.

[105] P.A. Perezhogin, 2D turbulence closures for the barotropic jet instability simulation, Russian J. Numer. Anal. Math. Modelling 35 (1) (2020) 21–35.

[106] P. Perezhogin, A. Glazunov, A priori and a posteriori analysis in large eddy simulation of the two-dimensional decaying turbulence using explicit filtering approach, in: EGU General Assembly Conference Abstracts, 2021, EGU21–2382.

[107] J. Berner, G. Shutts, M. Leutbecher, T. Palmer, A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system, J. Atmos. Sci. 66 (3) (2009) 603–626.

[108] P.A. Davidson, Turbulence: An Introduction for Scientists and Engineers, Oxford University Press, 2015.