AGU ADVANCING EARTH AND SPACE SCIENCE

JAMES | Journal of Advances in Modeling Earth Systems®

RESEARCH ARTICLE

10.1029/2021MS002847

Key Points:

- Deep learning-based subgrid-scale models outperform traditional eddy-viscosity models in *a priori* (offline) tests
- Deep learning-based subgrid-scale models can account for density stratification effects in boundary-layer turbulence
- The DNN model can capture key turbulence statistics in *a posteriori* (online) tests when applied to large-eddy simulations of the atmospheric boundary layer

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

P. Gentine,
pg2328@columbia.edu

Citation:

Cheng, Y., Giometto, M. G., Kauffmann, P., Lin, L., Cao, C., Zupnick, C., et al. (2022). Deep learning for subgrid-scale turbulence modeling in large-eddy simulations of the convective atmospheric boundary layer. *Journal of Advances in Modeling Earth Systems*, *14*, e2021MS002847. https://doi.org/10.1029/2021MS002847

Received 28 SEP 2021
Accepted 5 MAY 2022

Deep Learning for Subgrid-Scale Turbulence Modeling in Large-Eddy Simulations of the Convective Atmospheric Boundary Layer

Yu Cheng[1], Marco G. Giometto[2], Pit Kauffmann[3], Ling Lin[3], Chen Cao[3], Cody Zupnick[3], Harold Li[1], Qi Li[4], Yu Huang[1], Ryan Abernathey[5], and Pierre Gentine[1,5]

[1]Department of Earth and Environmental Engineering, Columbia University, New York, NY, USA, [2]Department of Civil Engineering and Engineering Mechanics, Columbia University, New York, NY, USA, [3]Capstone Project at Data Science Institute, Columbia University, New York, NY, USA, [4]School of Civil and Environmental Engineering, Cornell University, Ithaca, NY, USA, [5]Department of Earth and Environmental Sciences, Columbia University, Palisades, NY, USA

Abstract In large-eddy simulations, subgrid-scale (SGS) processes are parameterized as a function of filtered grid-scale variables. First-order, algebraic SGS models are based on the eddy-viscosity assumption, which does not always hold for turbulence. Here we apply supervised deep neural networks (DNNs) to learn SGS stresses from a set of neighboring coarse-grained velocity from direct numerical simulations of the convective boundary layer at friction Reynolds numbers Re_τ up to 1243 without invoking the eddy-viscosity assumption. The DNN model was found to produce higher correlation between SGS stresses compared to the Smagorinsky model and the Smagorinsky-Bardina mixed model in the surface and mixed layers and can be applied to different grid resolutions and various stability conditions ranging from near neutral to very unstable. The DNN model can capture key statistics of turbulence in *a posteriori* (online) tests when applied to large-eddy simulations of the atmospheric boundary layer.

Plain Language Summary Using grid-scale state variables to represent subgrid-scale (SGS) processes is a major step in large-eddy simulations of turbulence, which typically rely on some physically based assumptions that do not always apply. Here we replace physically based assumptions with deep learning in SGS modeling and show the latter performs better in *a priori* (offline) tests of atmospheric turbulence. In addition, the deep neural networks model well captures key statistics of turbulence in *a posteriori* (online) tests when applied to in large-eddy simulations of atmospheric boundary layers.

1. Introduction

Turbulent flows are ubiquitous in nature and in engineering systems. Geophysical turbulence at very high Reynolds numbers ($Re{\sim}10^7$ in the atmospheric boundary layer (Bradley et al., 1981)) exhibits a large range of scales that cannot be resolved by direct numerical simulations (DNSs) of the Navier-Stokes equations with current computational resources (Lee & Moser, 2015; Pope, 2000; Speziale, 1991; Voller & Porte-Agel, 2002; Yamamoto & Tsuji, 2018). The large-eddy simulation (LES) technique (Deardorff, 1970; Lilly, 1967; Smagorinsky, 1963) only resolves large-scale, filtered, eddy-motions and parameterizes the impact of subgrid-scale (SGS) turbulence (Smagorinsky, 1963) as a function of the resolved flow (Text S1 in Supporting Information S1). The accuracy of LESs is highly dependent on the properties of the SGS model (Germano et al., 1991; Lilly, 1992).

Widely used SGS models, such as the dynamic Smagorinsky model (Germano et al., 1991; Lilly, 1992), the Smagorinsky-Bardina mixed models (Bardina et al., 1980), Wall-Adapting Local Eddy-viscosity (WALE) models (Nicoud & Ducros, 1999), and Lagrangian scale dependent model (LASD) (Bou-Zeid et al., 2005), are first-order, algebraic closure models, and rely on the eddy-viscosity assumption (a gradient-diffusion model), which is justified only if the mixing length is much smaller than the scale of the mean flow variations. However, this condition is not always met in real flows (Schmitt, 2007), so that there is no clear length scale separation (Bernard & Handler, 1990; Corrsin, 1975; Egolf, 1994; Lilly, 1967; Speziale, 1991; Tennekes et al., 1972) for turbulent flows. Previous studies (Clark et al., 1979; Liu et al., 1994) showed that the correlation ρ between SGS stress calculated from the Smagorinsky model and from DNSs was only around 0.2. The Bardina similarity model (Bardina et al., 1980) resulted in a correlation coefficient $\rho = 0.8$, amongst the highest correlations obtained in *a priori* tests. It is worth noting that the Bardina similarity model does not rely on the eddy-viscosity assumption.

The Smagorinsky-Bardina mixed model (Bardina et al., 1980) was proposed shortly after to enhance the dissipative properties of the standard Bardina model, while still preserving a high correlation around 0.8 in neutrally stratified, incompressible channel flow (Horiuti, 1989).

The atmospheric boundary layer (ABL) can be unstable (stable) with vertical motion strengthened (damped) when there is heating (cooling) from the surface generating large coherent structures which are transporting momentum and scalars on large scales. In LESs relying on scale-invariant Smagorinsky models (where the Smagorinsky coefficient C_s does not change with the scale (Germano et al., 1991; Porté-Agel et al., 2000)), scales of motion smaller than the grid stencil (Canuto & Cheng, 1997; Lilly, 1967) should fall within the isotropic inertial subrange (Kolmogorov, 1941). As the Dougherty-Ozmidov length scale (Dougherty, 1961; Grachev et al., 2015; Ozmidov, 1965) is typically modeled rather than resolved due to the limitation of computational resources (Cheng et al., 2020), LESs of the strongly stable ABL can hardly resolve isotropic turbulence and stably stratified conditions will not be considered in this manuscript. On the other hand, LESs have been widely applied to study the unstable ABL (Deardorff, 1972; Li et al., 2018; Moeng, 1984; Nieuwstadt et al., 1993) as the energy-containing eddies are dominant and larger compared to those of neutral and unstable ABLs. Yet, various unstable (from weakly to highly unstable) conditions can fundamentally alter turbulent transport and its coherent structures. The eddy-viscosity formulation in SGS models can be modified to account for such stability effects (Chung & Matheou, 2014; Mason, 1989; Mason & Brown, 1999), but the quality and reliability of available parameterizations remain an open problem (Chamecki et al., 2007; Chung & Matheou, 2014; Kleissl et al., 2003).

Recently, deep neural networks (DNNs) with convolutional operations (Krizhevsky et al., 2012; LeCun et al., 1989) have been successfully applied for recognition and detection of images and objects (Girshick et al., 2014; Sermanet et al., 2013; Taigman et al., 2014). In general DNNs with convolutional operators are efficient (as they reduce the dimensionality) for processing data with a known, grid-like topology (Goodfellow et al., 2016). Moreover, Bar-Sinai et al. (2018) showed that DNNs can represent spatial gradients well. Recent studies have shown that DNNs may perform well in modeling turbulence stresses and how they relate to the flow gradients, which are key to SGS stresses. For instance, Ling et al. (2016) successfully applied DNNs to model Reynolds stress in Reynolds-averaged Navier-Stokes (RANS) models. Yang et al. (2019) developed artificial neutral networks for wall modeling in LESs for neutral channel flow. Gamahara and Hattori (2017) trained an artificial neural network to represent SGS stresses in LESs for a channel flow and the gradient was imposed and calculated rather than learned. Compared to gradient-based method, additional non-adjacent neighboring pixels can be added to include non-local effects (e.g., curvature) in DNNs.

Here we apply DNNs to learn the SGS stress in LESs (details in Supporting Information S1) with varying degrees of stability from near neutral to very unstable conditions ($z_i/L = -678.2$, where z_i is boundary layer height and L Obukhov length (Obukhov, 1946)) in the ABL, using a series of coarse-grained DNSs as the training data set. The results of the DNN models are compared to those of the Smagorinsky model (Lilly, 1967; Smagorinsky, 1963) and of the Smagorinsky-Bardina mixed model (Bardina et al., 1980). One focus of this study is to gain insight into DNN-based SGS modeling across stability conditions and grid resolutions. We also examine if the DNN model could be applied to higher Reynolds numbers by evaluating the predicted SGS stresses in the log-law region (von Kármán, 1930) of the boundary layer (details in Supporting Information S1 (Monin & Obukhov, 1954; Obukhov, 1946; Panofsky, 1963; Paulson, 1970)), which is "universal" (Marusic et al., 2013) across high Reynolds numbers. We test the proposed DNN model explicitly in *offline* (*a priori*) tests (SGS models applied to DNS data to calculate stresses) and also conduct *online* (*a posteriori*) tests (SGS models applied to an LES experiment to calculate stresses, replacing traditional representation) by implementing the model in an LES solver.

2. Methodology

2.1. Introduction to DNS Data

Four different DNS datasets are processed to provide the input data for training and test, including three simulations of convective boundary layers and one simulation of neutral channel flow. Details of the DNS setups can be found in Table S1 and Text S2 in Supporting Information S1 (Cheng et al., 2021; Chorin, 1968; Giometto et al., 2017; Li et al., 2018; Nieuwstadt et al., 1993; Orlandi, 2012; Shah & Bou-Zeid, 2014; Wray, 1990). The viscous layer has been removed from the training and prediction datasets. In convective boundary layers, the initial velocity field is set by the geostrophic wind U_g. The Reynolds number is defined as $Re_\tau = \frac{u_\tau z_i}{\nu}$, where u_τ

is the friction velocity, z_i the boundary layer height and ν the kinematic viscosity. The three simulations of the convective boundary layers are named Sh2, Sh5 and Sh20, with corresponding shear Reynolds numbers Re_τ of 309 ($z_i/L = -678.2$), 554 ($z_i/L = -105.1$) and 1243 ($z_i/L = -7.1$), respectively. Each data set consists of streamwise velocity u, spanwise velocity v, vertical velocity w, potential temperature θ and modified pressure p^* fields at different time steps ($p^* = p + \frac{\rho}{2} u_l u_l$ (Li et al., 2018)). For example, $Sh5_{t=1}$ denotes data set Sh5 at time $t = 1$. $\Delta_z^+ \equiv \Delta_z u_\tau / \nu$ is the spatial grid resolution denoted in terms of inner units in the vertical direction. The fully developed incompressible planar channel flow data set (Giometto et al., 2017) is named Channel1. The Reynolds number is defined as $Re_h = \frac{u_\tau h}{\nu} = 546$, where h is the channel half-width.

Each DNS data set is spatially filtered (Leonard, 1975) to provide coarse-grained data with different spatial grid resolutions in inner units. Similarly to Clark et al. (1979), we use a top-hat spatial filter (Clark et al., 1979) (as detailed in the prediction datasets of Table 1). Similarly to Clark et al. (1979), the velocity $\overline{u_l}(x, y, z)$ with a coarse-graining factor of $2k + 1$ is therefore calculated as

$$\overline{u_l}(x, y, z) = \frac{1}{(2k+1)^3} \sum_{x'=x-k}^{x'=x+k} \sum_{y'=y-k}^{y'=y+k} \sum_{z'=x-k}^{z'=z+k} u_l\left(x', y', z'\right)$$

where (x,y,z) are the coordinates of the points on the fine grid. The filter is spatial averaging of $2k + 1$ points in each direction, thus the ratio of horizontal and vertical grid resolutions in the coarse-grained data set used for the LES is the same as that in the original DNS data. Unlike scale-invariant Smagorinsky models, the cutoff grid sizes of the DNN model need not fall in the inertial subrange. The cutoff grid sizes of coarse-grained datasets $Sh20_{t=4,\Delta_z^+=42}$ and $Sh20_{t=4,\Delta_z^+=64}$ are in the inertial subrange of the corresponding raw DNS data as in typical LESs, while the cutoff sizes of other coarse-grained DNSs are in the dissipation range (Table 1), which is due to the narrow inertial subrange of low DNS Reynolds numbers.

The deviatoric part of the SGS stress tensor τ_{lk}^d is parameterized in LESs as a function of resolved quantities (Lilly, 1967; Smagorinsky, 1963). $\tau_{lk,DNS}^d$ is calculated using the DNS fine-grid velocity. It is worth noting that aliasing errors occur (Chow & Moin, 2003) when we compute SGS stresses using the DNS output, thus the aliasing errors in $\tau_{lk,DNS}^d$ may be learned by the DNN model. Removing the aliasing errors will be implemented in a future study. The estimated stress by DNNs is named $\tau_{lk,NN}^d$. As an example, the velocity in the x-y plane at $z^+ = \frac{z}{(\nu/u_\tau)} = 129.8$ in the logarithmic equilibrium layer of the original and spatially filtered data set $Sh20_{t=1}$ are shown (Figure 1).

2.2. Deep Neural Networks

A mapping $f(\boldsymbol{x}) \rightarrow \boldsymbol{y}$ is called neural network if $f(\boldsymbol{x})$ consists of a chain of functions, $f(\boldsymbol{x}) = f^n(\cdots f^k(\cdots f^2(f^1(\boldsymbol{x}))))$, where $f^1, f^2, \ldots f^k, \ldots f^n$ are different functions (Goodfellow et al., 2016) and where f^k is the kth layer and f^n the output layer. If the information only flows from $\boldsymbol{x}$ to $\boldsymbol{y}$ the networks are called feedforward (Goodfellow et al., 2016). The name neural networks arises because of the analogy with neural sciences (McCulloch & Pitts, 1943). Deep neural networks (DNN) typically refer to neural networks with more than two layers (Goodfellow et al., 2016).

The first dense layer utilizes the whole volume of input data, which can be regarded as a convolutional operation with kernel size being equal to the input volume in convolutional neural networks (CNN) (Goodfellow et al., 2016; LeCun et al., 1989). The cost function is the mean squared errors between $\tau_{lk,NN}^d$ from DNN models and $\tau_{lk,DNS}^d$ from DNS. The mean squared errors have been widely used as the cost function in modeling SGS stresses (Gamahara & Hattori, 2017; Xie et al., 2020) or subgrid parameterization in climate models (Rasp et al., 2018). A validation data set (20% of the total training data) is separated from the training data set to detect overfitting. The input data is normalized by its standard deviation and mean in the whole DNS field. Each dense layer has a nonlinear Rectified Linear Unit (ReLU) (Nair & Hinton, 2010) activation function $f(x) = \max(0,x)$, which typically follows affine transformations in each layer (Goodfellow et al., 2016). Dropout (Hinton et al., 2012) is used to avoid overfitting and to limit the influence of large weights. A schematic of the DNN set-up is shown in Figure 2. A mini-batch gradient descent method (Ioffe & Szegedy, 2015) is used to update the weight and to introduce some stochasticity in the descent by using small batches. The batch size is 1000 and number of epochs is 50. We have tested different batch sizes but did not find improved results when changing the value. A batch size

Table 1
Training Datasets, Input Variables and Prediction Datasets of Different DNN Models

DNN model name	Training data set	Input variables	Prediction data set	Correlation between $\tau_{13,DNS}$ and $\tau_{13,NN}$
M_uvw_multiSh	$Sh20_{t=1,\Delta_z^+=21}$,	u, v, w	$Sh20_{t=4,\Delta_z^+=11}$,	0.834
	$Sh20_{t=2,\Delta_z^+=21}$,		$Sh20_{t=4,\Delta_z^+=21}$,	0.840
	$Sh2_{t=1,\Delta_z^+=6}$,		$Sh20_{t=4,\Delta_z^+=42}$,	0.816
	$Sh2_{t=2,\Delta_z^+=6}$,		$Sh20_{t=4,\Delta_z^+=64}$,	0.778
			$Sh2_{t=4,\Delta_z^+=6}$,	0.824
			$Sh5_{t=1,\Delta_z^+=10}$,	0.818
			$Channel1_{\Delta_z^+=23}$,	0.415
M_uvw	$Sh20_{t=1,\Delta_z^+=21}$	u, v, w	$Sh20_{t=4,\Delta_z^+=21}$,	0.838
			$Sh20_{t=4,\Delta_z^+=42}$,	0.817
			$Sh20_{t=4,\Delta_z^+=64}$,	0.781
			$Sh2_{t=4,\Delta_z^+=6}$,	0.769
			$Sh5_{t=1,\Delta_z^+=10}$,	0.775
M_uvwθ	$Sh20_{t=1,\Delta_z^+=21}$	u, v, w, θ	$Sh20_{t=4,\Delta_z^+=21}$,	0.843
			$Sh20_{t=4,\Delta_z^+=42}$,	0.825
			$Sh20_{t=4,\Delta_z^+=64}$,	0.798
			$Sh2_{t=4,\Delta_z^+=6}$,	0.777
			$Sh5_{t=1,\Delta_z^+=10}$,	0.787
M_uvw$p$	$Sh20_{t=1,\Delta_z^+=21}$	$u, v, w, p*$	$Sh20_{t=4,\Delta_z^+=21}$,	0.842
			$Sh20_{t=4,\Delta_z^+=42}$,	0.822
			$Sh20_{t=4,\Delta_z^+=64}$,	0.791
			$Sh2_{t=4,\Delta_z^+=6}$,	0.763
			$Sh5_{t=1,\Delta_z^+=10}$,	0.771
M_uvw_multiT	$Sh20_{t=1,\Delta_z^+=21}$,	u, v, w	$Sh20_{t=4,\Delta_z^+=21}$,	0.850
	$Sh20_{t=2,\Delta_z^+=21}$,		$Sh20_{t=4,\Delta_z^+=42}$,	0.829
	$Sh20_{t=3,\Delta_z^+=21}$,		$Sh20_{t=4,\Delta_z^+=64}$,	0.797
	$Sh20_{t=5,\Delta_z^+=21}$		$Sh2_{t=4,\Delta_z^+=6}$,	0.814
			$Sh5_{t=1,\Delta_z^+=10}$,	0.787
M_uvw_multiC	$Sh20_{t=1,\Delta_z^+=11}$,	u, v, w	$Sh20_{t=4,\Delta_z^+=21}$,	0.773
	$Sh20_{t=1,\Delta_z^+=21}$,		$Sh20_{t=4,\Delta_z^+=42}$,	0.803
	$Sh20_{t=1,\Delta_z^+=42}$,		$Sh20_{t=4,\Delta_z^+=64}$,	0.703
	$Sh20_{t=1,\Delta_z^+=64}$,		$Sh2_{t=4,\Delta_z^+=6}$,	0.757
			$Sh5_{t=1,\Delta_z^+=10}$,	0.758
M_uvw_multiSh_box555	$Sh20_{t=1,\Delta_z^+=21}$,	u, v, w	$Sh20_{t=4,\Delta_z^+=21}$,	0.876
	$Sh20_{t=2,\Delta_z^+=21}$,		$Sh20_{t=4,\Delta_z^+=42}$,	0.847
	$Sh2_{t=1,\Delta_z^+=6}$,		$Sh20_{t=4,\Delta_z^+=64}$,	0.808
	$Sh2_{t=2,\Delta_z^+=6}$,		$Sh2_{t=4,\Delta_z^+=6}$,	0.862
			$Sh5_{t=1,\Delta_z^+=10}$,	0.858
M_uvw_multiSh_box777	$Sh20_{t=1,\Delta_z^+=21}$,	u, v, w	$Sh20_{t=4,\Delta_z^+=21}$,	0.881
	$Sh20_{t=2,\Delta_z^+=21}$,		$Sh20_{t=4,\Delta_z^+=42}$,	0.851
	$Sh2_{t=1,\Delta_z^+=6}$,		$Sh20_{t=4,\Delta_z^+=64}$,	0.810
	$Sh2_{t=2,\Delta_z^+=6}$,		$Sh2_{t=4,\Delta_z^+=6}$,	0.873
			$Sh5_{t=1,\Delta_z^+=10}$,	0.867

Table 1
Continued

DNN model name	Training data set	Input variables	Prediction data set	Correlation between $\tau_{13,DNS}$ and $\tau_{13,NN}$

Note. Correlation coefficients of $\tau_{13,DNS}$ from DNS data and $\tau_{13,NN}$ from DNN models in the whole field of the coarse-grained DNS data are also shown. The first 6 DNN models have an input dimension of $3 \times 3 \times 3$, model M_uvw_multiSh_box555 has an input dimension of $5 \times 5 \times 5$ and model M_uvw_multiSh_box777 has an input dimension of $7 \times 7 \times 7$.

of 1000 was also used in previous neural network models for SGS stresses (Xie et al., 2020). The training loss approaches a steady state after 20 epochs.

The DNNs are trained to reproduce the SGS stress $\tau^{d}_{lk,DNS}$ that are generated from the DNS data, that is, we use supervised learning. For example, the SGS stress $\tau^{d}_{lk,DNS}$ at each spatial point is modeled using the coarse-grained variables $\overline{u_l}$,$\overline{v_l}$ and $\overline{w_l}$ ($\overline{\theta}$ or $\overline{p^*}$ is also added in some cases as the input of DNNs) in the neighboring $3 \times 3 \times 3$ box, which is used as the input layer of DNNs (schematic in Figure 2) with a total dimension of $3 \times 3 \times 3 \times 3$ (here the last "3" is number of input variables: $\overline{u_l}$,$\overline{v_l}$ and $\overline{w_l}$).

We tested different dimensions of the input box (Figure S1 in Supporting Information S1). Increasing input dimension from $3 \times 3 \times 3$ to $5 \times 5 \times 5$ or $7 \times 7 \times 7$ does not lead to much improvement for the correlations of $\tau^{d}_{lk,DNS}$ and $\tau^{d}_{lk,NN}$. We therefore selected input dimension of $3 \times 3 \times 3$ for the final model as a tradeoff between high correlation between SGS stresses and computational efficiency. We have also tested several architectures (not shown here), including typical convolutional neural networks (CNNs) and other dimensions of input box but we finally chose the architecture leading to the highest correlation between $\tau^{d}_{lk,DNS}$ and $\tau^{d}_{lk,NN}$. We tested 8 different combinations of input datasets (with different stabilities), input variables (u, v, w, θ, p^*) and dimensions of input box for training the DNNs (Table 1). The correlation coefficients (Clark et al., 1979) of $\tau^{d}_{lk,DNS}$ and $\tau^{d}_{lk,NN}$ in the turbulent region and zooming into the log-law layer are used to evaluate the performance of the models.

3. Results

3.1. DNNs Trained on Datasets With Different Stability Conditions

The deviatoric SGS stresses of data set $Sh5_{t=1,\Delta_z^+=10}$ calculated from the coarse-grained DNS data (denoted by $\tau^{d}_{lk,DNS}$), the Smagorinsky model (Smagorinsky, 1963) (denoted by $\tau^{d}_{lk,S}$), the Smagorinsky-Bardina mixed model (Bardina et al., 1980) (denoted by $\tau^{d}_{lk,SB}$) and the DNN model M_uvw_multiSh (denoted by $\tau^{d}_{lk,NN}$) at $z^+ = \frac{z}{(\nu/u_\tau)} = 77.3$ in the log-law region are shown in Figure 3 for illustration purposes. The SGS stress τ_{13} averaged in the x-y plane at different z^+ of data set $Sh5_{t=1,\Delta_z^+=10}$ are compared (Figure 4a). Similarly, another SGS stress is compared in Figure S2 in Supporting Information S1. The horizontal mean of τ_{13} is better captured by the DNN model in the majority of the domain including the log-law region between the two dotted lines, compared to the Smagorinsky model and the Smagorinsky-Bardina mixed model. The spatial variation of τ_{13} in the DNN prediction is similar to that in the DNS data, while the Smagorinsky model and the Smagorinsky-Bardina mixed model underestimate the variance of τ_{13} in the z direction (Figure 4a) and do not capture the vertical trend. This is further confirmed by the probability distribution function (Figure 4b): the τ_{13} distribution in Smagorinsky and Smagorinsky-Bardina predictions are too concentrated around the mean compared to that in the DNS data and DNN prediction. The latter indeed exhibits a broader distribution and better captures the tails of the distribution (Figure 4b), although there might still be underestimation of the variance. The correlation between $\tau_{13,DNS}$ and $\tau_{13,S}$, $\tau_{13,SB}$ and $\tau_{13,NN}$ in x-y plane at different z^+ are compared (Figure 4c). The correlation between $\tau_{13,S}$ and $\tau_{13,DNS}$ is about 0.2 in the middle of the vertical domain, in agreement with previous studies (Clark et al., 1979; Liu et al., 1994), while the correlation between $\tau_{13,NN}$ and $\tau_{13,DNS}$ is over 0.8, which is about 0.05 higher than that of the Smagorinsky-Bardina mixed model (Bardina et al., 1980; Horiuti, 1989) in the majority domain including the log-law region and the mixed layer that lies above the log-law region.

The streamwise power spectra of $\tau_{13,DNS}$, $\tau_{13,S}$, $\tau_{13,SB}$ and $\tau_{13,NN}$ averaged in the x-y plane at $z^+ = \frac{z}{(\nu/u_\tau)} = 77.3$ in the log-law region are then compared (Figure 4d). The power spectra produced by the DNN model are very close to the DNS data across the normalized wavenumber domain $0.05 < kz < 6$, where k is the streamwise wavenumber and z is distance to the wall. However, the power spectra produced by the Smagorinsky model and Smagorinsky-Bardina mixed model deviate much more from the DNS data compared to the DNN model, especially for

10.1029/2021MS002847

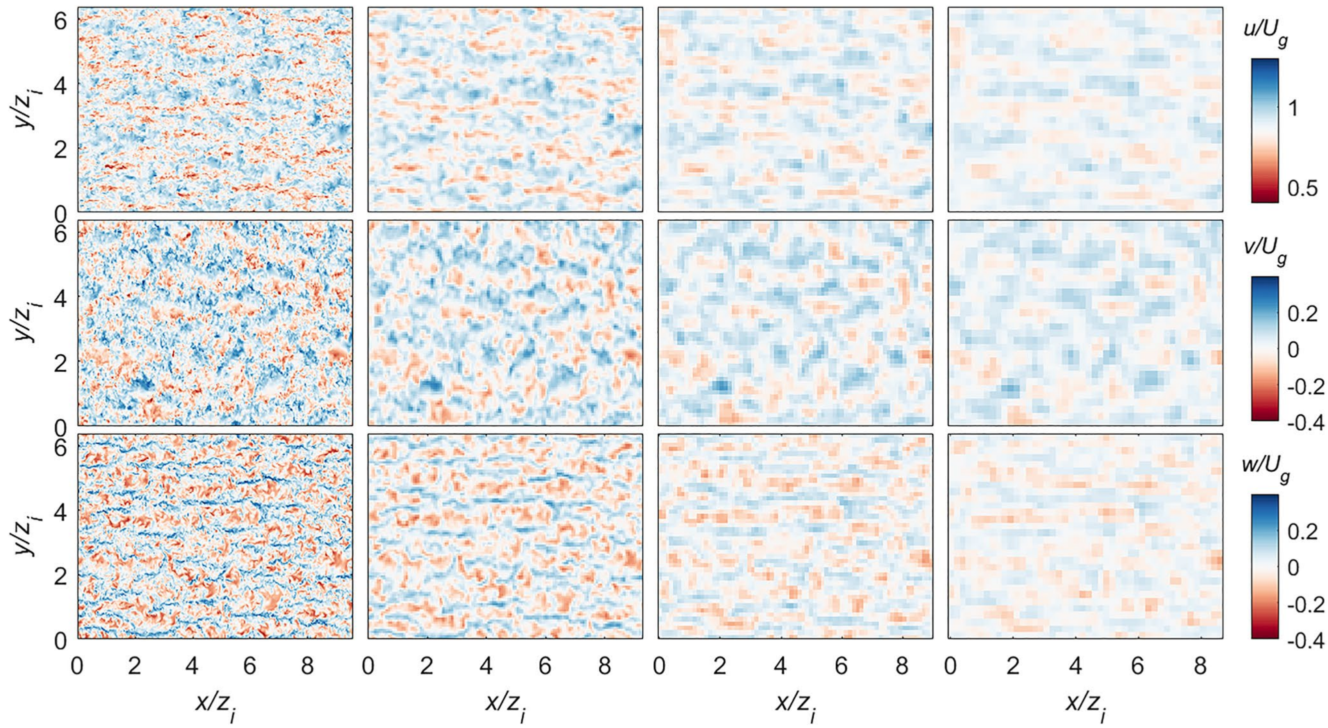

Figure 1. Velocity u/U_g (first row), v/U_g (second row) and w/U_g (third row) of x-y plane at $z^+ = \frac{z}{(\nu/u_\tau)} = 129.8$ in the log-law region of data set Sh20. U_g is geostrophic wind, ν is kinematic viscosity, u_τ is friction velocity and z_i is boundary layer height. From left to right the selected datasets are $Sh20_{t=1,\Delta_z^+=3}$, $Sh20_{t=1,\Delta_z^+=21}$, $Sh20_{t=1,\Delta_z^+=42}$ and $Sh20_{t=1,\Delta_z^+=64}$.

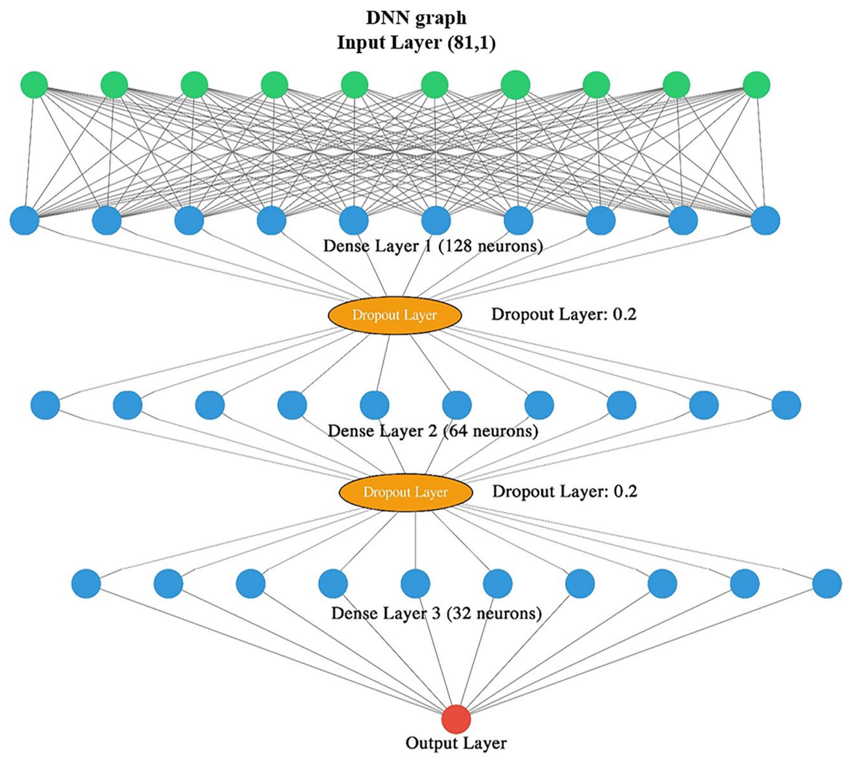

Figure 2. Schematic of the deep neural networks in this study.

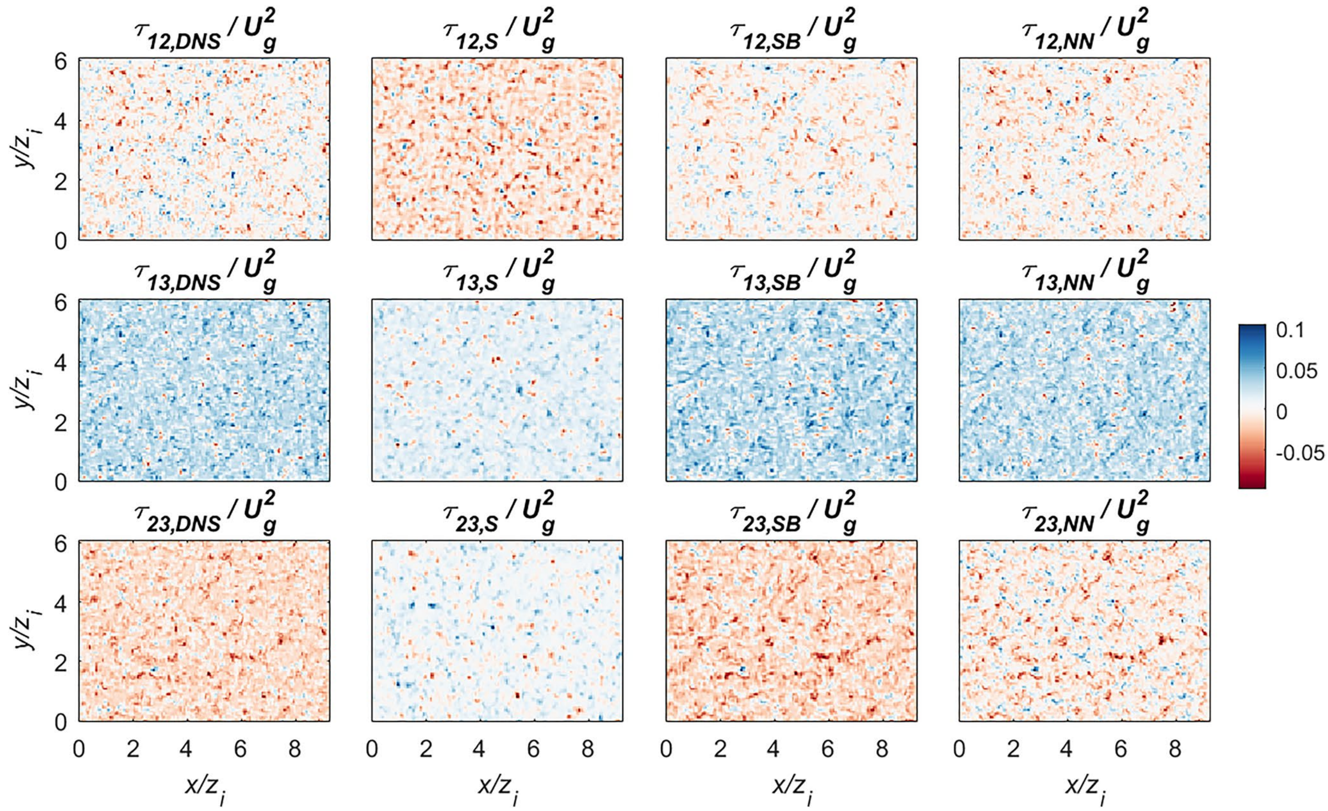

Figure 3. SGS stresses of x-y plane normalized by U_g^2 at $z^+ = \frac{z}{(\nu/u_\tau)} = 77.3$ in the log-law region of data set $Sh5_{t=1,\Delta_z^+=10}$. U_g is geostrophic wind, z_i is boundary layer height, ν is kinetic viscosity and u_τ is friction velocity. From left to right: SGS stresses calculated from DNS data, the Smagorinsky model, the Smagorinsky-Bardina mixed model and the DNN model M_uvw_multiSh.

$0.05 < kz < 0.3$ and $kz > 4$. Therefore, the DNN model M_uvw_multiSh predicts more accurately the SGS stress tensor components when compared to the Smagorinsky model and the Smagorinsky-Bardina mixed model both in the whole DNS field and the log-law region.

The correlation coefficients ρ calculated from $\tau_{13,S}$, $\tau_{13,SB}$ and $\tau_{13,NN}$ in the whole DNS field are compared (Figure 5a) across different test datasets. The overall correlation coefficients $\rho \approx 0.2$ between $\tau_{13,S}$ and $\tau_{13,DNS}$, while $\rho \approx 0.8$ between $\tau_{13,NN}$ and $\tau_{13,DNS}$ except for data set $Channel1_{\Delta_z^+=23}$ which has a different boundary condition compared to other DNS datasets. The produced correlation between the DNN model is around 0.05 higher than that of the Smagorinsky-Bardina mixed model (Bardina et al., 1980). Similar results can be found for the correlation only for the log-law region (Figure S3 in Supporting Information S1), emphasizing the quality of the DNN across turbulent regions of the flow.

To further evaluate the potential of the DNN model, we focus on the correlation ρ in the log-law region (Table 2). The DNN model M_uvw_multiSh produces smaller ρ for τ_{13} (Table 2) for most prediction datasets in the log-law region compared to the whole DNS turbulent region. The model M_uvw_multiSh (trained on datasets with $\Delta_z^+ = 21$ and $\Delta_z^+ = 6$) produces a correlation between 0.780 in the log-law region for $Sh5_{t=1,\Delta_z^+=10}$, which is relatively high when considering previous studies (Bardina et al., 1980; Clark et al., 1979; Horiuti, 1989). The largest correlation decreases in the log-law region compared to the whole DNS field is around 0.170 for the data set $Sh20_{t=4,\Delta_z^+=64}$, where there are though only three vertical grid levels in the former. In fact, the model M_uvw_multiSh produces correlation coefficients not smaller than 0.780 in all datasets with more than 5 vertical grid levels in the log-law region. Therefore, we conclude that the DNN model M_uvw_multiSh can reproduce SGS stresses well in the log-law region in *a priori* tests.

The DNN model M_uvw_multiSh, the Smagorinsky model, and the Smagorinsky-Bardina mixed model all produce much higher root mean square error (RMSE) in the more buoyant cases: $Sh2_{t=4,\Delta_z^+=6}$ ($z_i/L = -678.2$), and

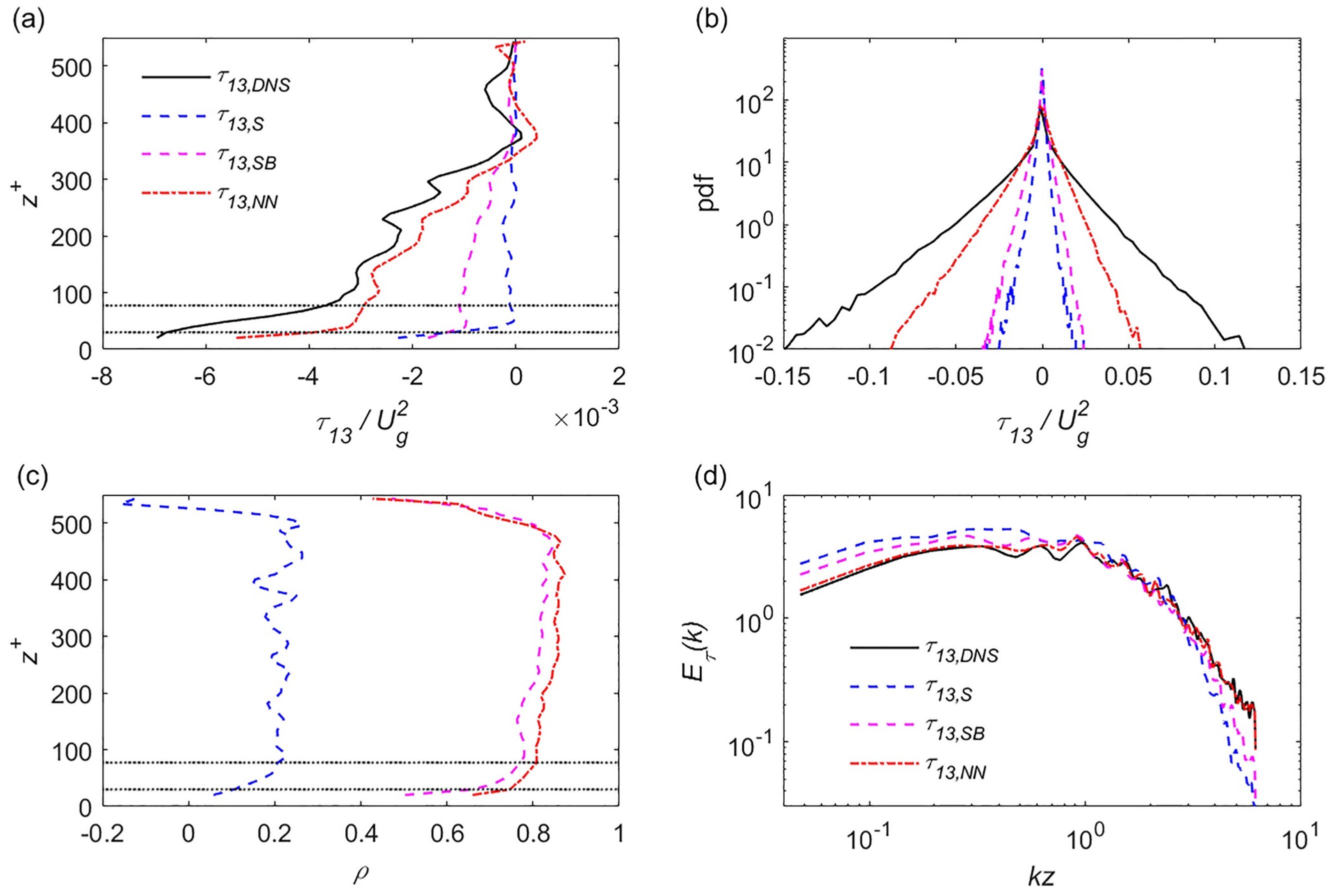

Figure 4. Comparison of $\tau_{13,DNS}$, $\tau_{13,S}$, $\tau_{13,SB}$ and $\tau_{13,NN}$. (a) Mean SGS stresses of *x-y* plane at different z^+. (b) Probability distribution function (pdf) of SGS stresses in the whole DNS field. (c) Correlation between SGS stresses in *x-y* plane at different z^+. (d) Streamwise spectra of SGS stresses averaged in *x-y* plane at $z^+ = \frac{z}{(\nu/u_\tau)} = 77.3$ in the log-law region. The data set is $Sh5_{t=1,\Delta_z^+=10}$ and DNN model is M_uvw_multiSh. U_g is geostrophic wind, ρ is correlation, ν is kinematic viscosity, u_τ is friction velocity, k is streamwise wavenumber, z is distance to the wall and E_τ is power spectra. z^+ between the two dotted lines is the log-law region in (a) and (c).

$Sh5_{t=1,\Delta_z^+=10}$ ($z_i/L = -105.1$). The difference in RMSE across different stabilities is related to topological changes in the coherent structures in the unstable cases (transitioning from rolls to thermals) with highly varying $\tau^d_{lk,DNS}$ in space due to buoyancy forces. When RMSE is normalized by the standard deviation of $\tau^d_{lk,DNS}$ (we divide the RMSE by the standard deviation of $\tau^d_{lk,DNS}$ in the whole DNS turbulent region), its magnitude is of order 1 across all datasets (Figure 5b). Therefore, similarly to correlation coefficients that are normalized by the standard deviation, we also use normalized RMSE to evaluate the DNN performance and find that the DNN model produces the lowest normalized RMSE (Figure 5b).

The model M_uvw_multiSh is trained using both the highly convective Sh2 ($z_i/L = -678.2$) case and the high-shear Sh20 ($z_i/L = -7.14$) case, but predicts SGS stresses for the Sh5 (moderately convective, $z_i/L = -105.1$) case better than the Smagorinsky model and the Smagorinsky-Bardina mixed model in terms of correlation, probability distribution, mean value, power spectra and normalized RMSE. Therefore, it appears that the DNN model M_uvw_multiSh has learned the topology of the turbulent structures from the two extreme shear and convective cases and can appropriately be used to make predictions on datasets with intermediate stability. In other words, the model trained on extreme cases can interpolate turbulence behavior in between.

3.2. Sensitivity to Input Parameters

In addition to using u, v and w as the input to the DNNs, we also added potential temperature or modified pressure as an input variable for the training data $Sh20_{t=1,\Delta_z^+=21}$. The resulted correlation between models M_uvwθ

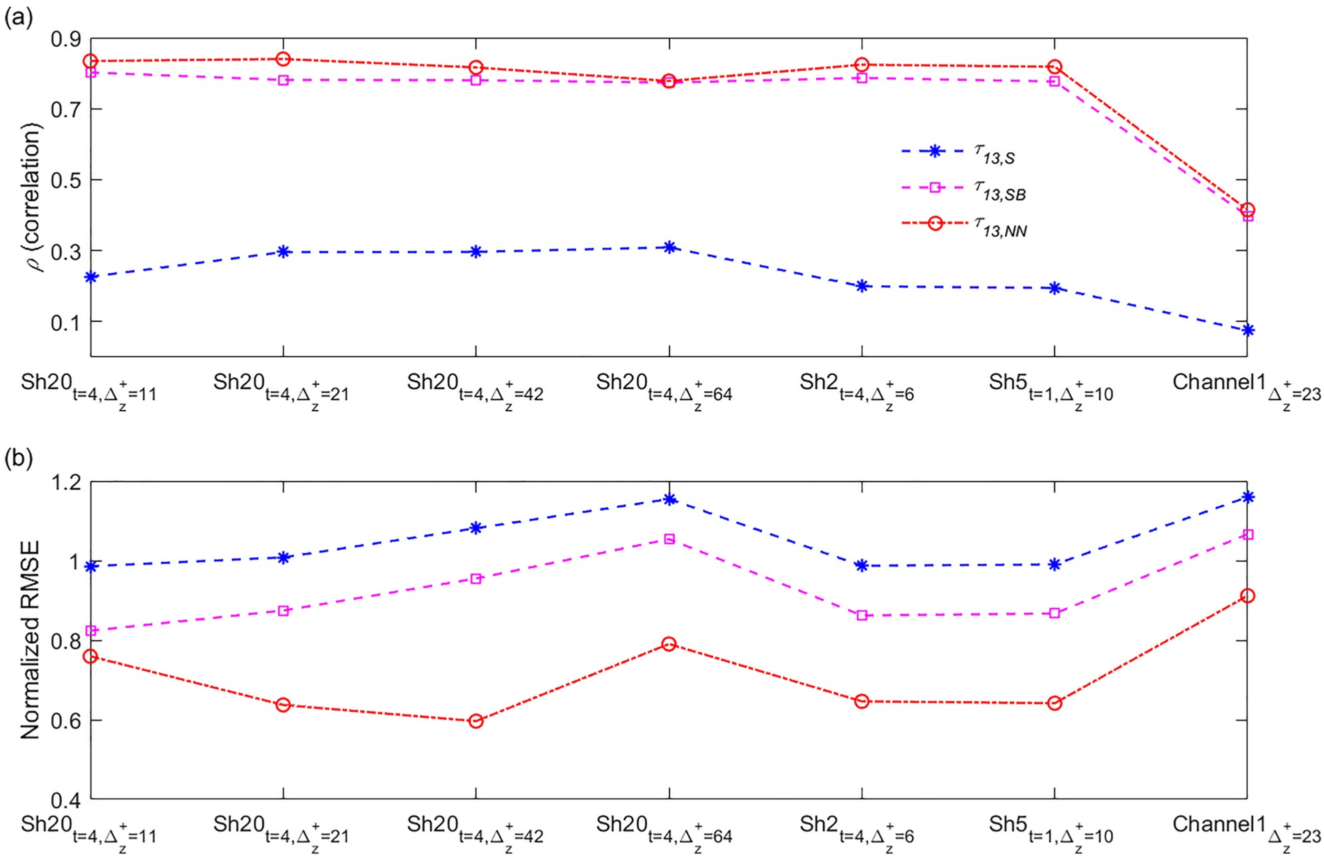

Figure 5. (a) Correlation coefficients calculated from $\tau_{13,S}$, $\tau_{13,SB}$ and $\tau_{13,NN}$ predicted by model M_uvw_multiSh for the whole DNS field. (b) Normalized root mean square error (RMSE) calculated from $\tau_{13,S}$, $\tau_{13,SB}$ and $\tau_{13,NN}$ for the whole DNS field. The prediction datasets are listed in *x* axis.

and M_uvw$p$ (Table 1) with additional inputs are compared to that of the model trained only with u, v, w (model M_uvw). As a reference, we also show results of model M_uvw_multiSh (Figure 6a, Figures S4a and S5a in Supporting Information S1).

For datasets $Sh20_{t=4,\Delta_z^+=21}$, $Sh20_{t=4,\Delta_z^+=42}$ and $Sh20_{t=4,\Delta_z^+=64}$, model M_uvwθ (using u, v, w and θ as input) produces a slightly higher correlation for τ_{13} (Figure 6a) but slightly lower correlation for τ_{23} (Figure S5a in Supporting Information S1) compared to model M_uvw (using u, v and w as input). For datasets $Sh2_{t=4,\Delta_z^+=6}$, and $Sh5_{t=1,\Delta_z^+=10}$, model M_uvwθ produces close correlation for τ_{13} and τ_{23} (Figures 6a and S5a in Supporting Information S1) compared to model M_uvw. In the log-law region, models M_uvwθ and M_uvw produce very

Table 2
Correlation Coefficients of $\tau_{13,DNS}$ and $\tau_{13,NN}$ Predicted by Different DNN Models for Different Prediction Datasets in Both the Log-Law Region and the Whole DNS Field

Correlation for τ_{13} log-law region (whole DNS field)	M_uvw_ multiSh	M_uvw	M_uvwθ	M_uvwp	M_uvw_ multiT	M_uvw_ multiC	M_uvw_ multiSh_box555	M_uvw_ multiSh_box777
$Sh20_{t=4,\ \Delta_z^+=11}$	0.847(0.834)	0.849(0.837)	0.847(0.829)	0.850(0.851)	0.850(0.829)	0.871(0.884)	0.890(0.862)	0.882(0.853)
$Sh20_{t=4,\ \Delta_z^+=21}$	0.806(0.840)	0.804(0.838)	0.812(0.843)	0.809(0.842)	0.820(0.850)	0.750(0.773)	0.852(0.876)	0.858(0.881)
$Sh20_{t=4,\ \Delta_z^+=42}$	0.716(0.816)	0.723(0.817)	0.738(0.825)	0.730(0.822)	0.744(0.829)	0.707(0.803)	0.765(0.847)	0.768(0.851)
$Sh20_{t=4,\ \Delta_z^+=64}$	0.608(0.778)	0.611(0.781)	0.639(0.798)	0.636(0.791)	0.642(0.797)	0.476(0.703)	0.673(0.808)	0.669(0.810)
$Sh2_{t=4,\ \Delta_z^+=6}$	0.793(0.824)	0.732(0.769)	0.753(0.777)	0.731(0.763)	0.788(0.814)	0.730(0.757)	0.838(0.862)	0.850(0.873)
$Sh5_{t=1,\ \Delta_z^+=10}$	0.780(0.818)	0.732(0.775)	0.756(0.787)	0.734(0.771)	0.745(0.787)	0.718(0.758)	0.830(0.858)	0.840(0.867)

Note. Correlation coefficients for the whole DNS field are in the brackets. $\Delta_z^+ \equiv \Delta_z u_\tau / \nu$ is the spatial grid resolution in the vertical direction in inner units.

10.1029/2021MS002847

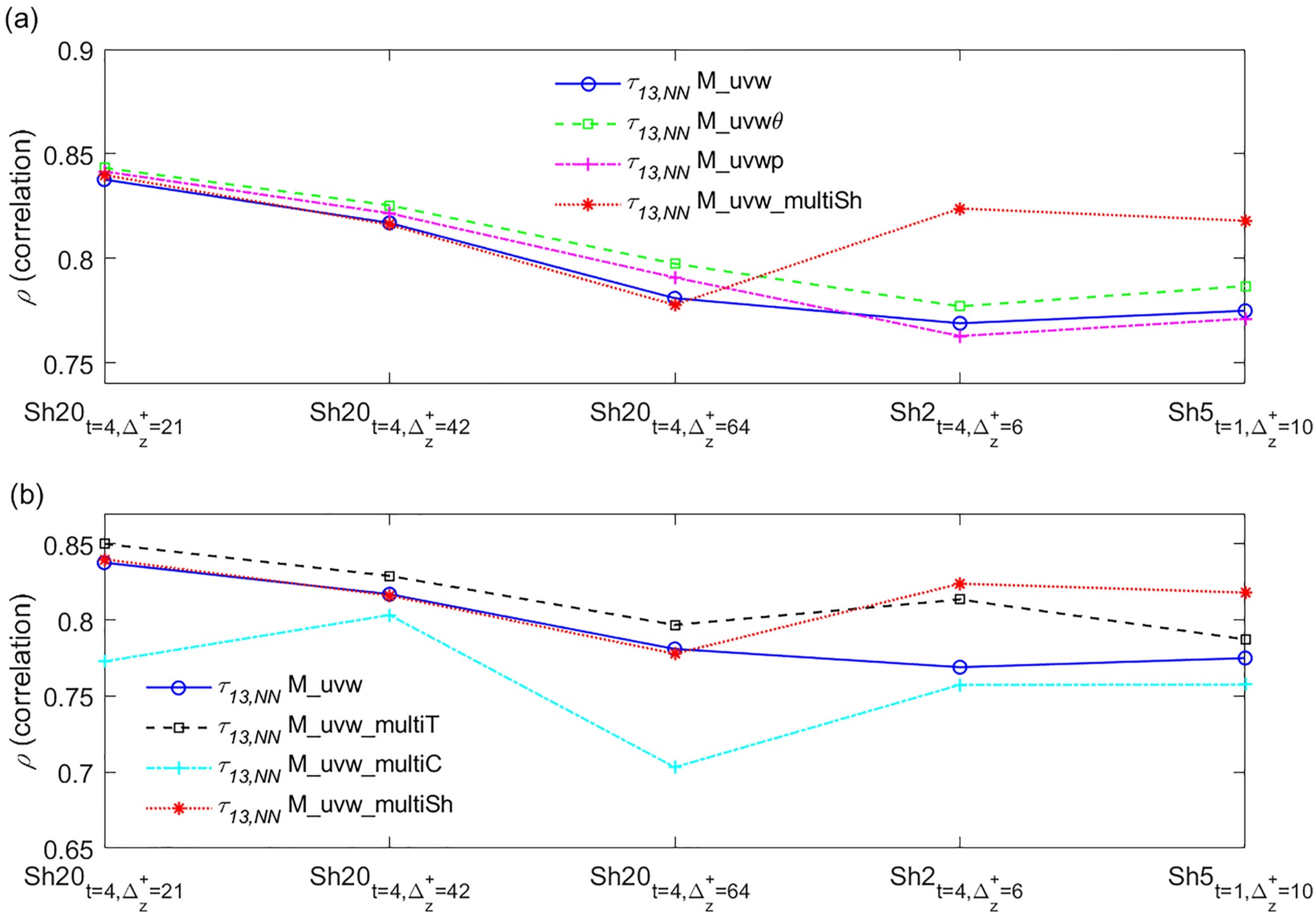

Figure 6. Correlation coefficients between $\tau_{13,DNS}$ and $\tau_{13,NN}$ predicted from different DNN models in the whole DNS field. (a) Comparison between the DNN models M_uvw_multiSh, M_uvw, M_uvwθ and M_uvw$p$. (b) Comparison between the DNN models M_uvw_multiSh, M_uvw, M_uvw_multiT and M_uvw_multiC. The prediction datasets are listed in x axis.

close correlation for τ_{13} (Table 2) with differences less than 0.02, so that the effects of potential temperature is not significant. Therefore, adding potential temperature as an extra input variable to include stability effects is not needed as additional input, across different unstable conditions. This is in contrast with some SGS models that additionally consider stability effects (due to temperature fluctuations) in unstable conditions (Mason, 1989; Mason & Brown, 1999) to calculate SGS stresses. Here the stability information might already be encoded in the velocity field.

For datasets $Sh20_{t=4,\Delta_z^+=21}$, $Sh20_{t=4,\Delta_z^+=42}$, and $Sh20_{t=4,\Delta_z^+=64}$, model M_uvw$p$ (using u, v, w and p^* as input) including pressure effects produces very close correlation for τ_{13} (Figure 6a), τ_{12} (Figure S4a in Supporting Information S1) and τ_{23} (Figure S4a in Supporting Information S1) compared to model M_uvw (using u, v and w as input). In the log-law region, models M_uvw$p$ and M_uvw produce very close correlation for τ_{13} (Table 2) with differences less than 0.02, so that the effects of pressure is not significant. Bernard and Handler (1990) suggested that pressure might contribute to acceleration transport, which could be a key component of the Reynolds stress. However, we here show that pressure need not be explicitly taken into account to calculate SGS stresses as the velocity field seems to have inherently incorporated pressure effects, which is not surprising given that the pressure field can be expressed as a function of the velocity field via the pressure Poisson equation. Besides, potential temperature and pressure do not additionally increase the correlation by changing the input box to $5 \times 5 \times 5$ and $7 \times 7 \times 7$ (not shown here). Based on those results, we decided to only keep the velocity field as input to the DNNs.

3.3. Effects of the Stability Range of Training Data

To test if the initial training data need to cover different stability conditions to capture various coherent structures (Agee et al., 1973; Atkinson & Zhang, 1996; Li & Bou-Zeid, 2011; Li et al., 2018; Lumley, 1981; Salesky et al., 2017), we also train a DNN model (denoted as model M_uvw_multiT) using only $Sh20_{\Delta_z^+=21}$ ($z_i/L = -7.14$) at different time steps.

For datasets $Sh20_{t=4,\Delta_z^+=21}$ ($z_i/L = -7.14$) and $Sh20_{t=4,\Delta_z^+=42}$ ($z_i/L = -7.14$), model M_uvw_multiT produces a higher correlation (around 0.02) for τ_{13} (Figure 6b) compared to model M_uvw_multiSh, since the former model is trained using Sh20 ($z_i/L = -7.14$) while the latter model is trained using both Sh20 and Sh2 ($z_i/L = -678.2$). For data set $Sh5_{t=1,\Delta_z^+=10}$ ($z_i/L = -105.1$), model M_uvw_multiT produces a lower correlation by around 0.04 for τ_{13} (Figure 6b) compared to model M_uvw_multiSh. The above analysis of the performance of models M_uvw_multiSh and M_uvw_multiT in the whole DNS domain also holds for the log-law region (Table 2). Therefore, to make predictions on datasets with the same stability as that of existing training data, it is enough to just train on one data set. However, to make predictions in other stability conditions, it is essential to train on a range of stability conditions from unstable to shear-driven conditions that cover the prediction stability.

3.4. Effects of Grid Resolutions–Scale Awareness

An important consideration for an SGS model is if it works well for LESs with different grid resolutions. Therefore, we also train model M_uvw_multiC on coarse-grained DNS datasets with different grid resolutions to test the scale awareness of the DNN model. For the coarse-resolution data set $Sh20_{t=4,\Delta_z^+=64}$ (Figure 6b and Table 2), model M_uvw (trained on $Sh20_{t=1,\Delta_z^+=21}$) produces a higher correlation by 0.078 for τ_{13} than model M_uvw_multiC (trained on $Sh20_{t=1,\Delta_z^+=11}$, $Sh20_{t=1,\Delta_z^+=21}$, $Sh20_{t=1,\Delta_z^+=42}$ and $Sh20_{t=1,\Delta_z^+=64}$) in the whole DNS field, although the model M_uvw_multiC is trained on extra datasets with different grid resolutions. In particular, the model M_uvw produces a higher correlation by 0.135 for τ_{13} compared to the model M_uvw_multiC in the log-law region of data set $Sh20_{t=4,\Delta_z^+=64}$. One possible explanation could be that the DNS datasets with coarser resolution (e.g., $Sh20_{t=1,\Delta_z^+=42}$ and $Sh20_{t=1,\Delta_z^+=64}$) lose some of the smaller-scale turbulence information thus leading to relatively worse performance when used to train the DNN models.

For the fine-resolution data set $Sh20_{t=4,\Delta_z^+=11}$ (Table 2), model M_uvw produces a lower correlation by 0.047 than model M_uvw_multiC, which is due to the fine-resolution training data $Sh20_{t=1,\Delta_z^+=11}$ applied to the latter. Including training data with the same grid resolution as test data set can contribute to improved correlation for prediction on fine-resolution datasets. Therefore, training on datasets with the finest grid resolutions is able to produce good correlation between SGS stresses on datasets across different grid resolutions.

3.5. Online Tests

We then implement a DNN model trained on $Sh20_{\Delta_z^+=21}$ ($z_i/L = -7.1$) and $Sh2_{\Delta_z^+=6}$ ($z_i/L = -678.2$) in the existing LES algorithm, and use it in simulations of a convective ABL ($z_i/L = -8.7$), that is, online (*a posteriori*) tests. The LES code used here has been applied to simulate the convective ABL (Cheng et al., 2017; Li et al., 2018; Salesky et al., 2017), with detailed descriptions in Kumar et al. (2006). In the LES experiments, a mixed pseudospectral approach and second-order accurate staggered centered finite difference scheme are adopted to discretize the system of equations in the horizontal and vertical directions, respectively. As the DNN model is trained without a sponge layer (Nieuwstadt et al., 1993) that prevents the reflection of gravity waves at the top 25% of the computational domain, we do not include SGS stresses in that region of the domain. As described in Li et al. (2018), Monin-Obukhov similarity theory is used to compute the surface stress. To smooth the stresses computed by the DNN model and Monin-Obukhov similarity theory, vertical averaging of SGS stresses is applied only at the second lowest grid (the grid just above the most bottom grid). The *a posteriori* LES experiments have been run for a temporal period that is equivalent to 4.1 hr in the ABL, similar to the simulation periods in previous LES studies (Cancelli et al., 2014; Salesky et al., 2017). A constant heat flux is prescribed at the bottom in most research (including our study) to simulate a quasi-steady convective atmospheric boundary layer (Kumar et al., 2006), but this is not the case in the real-world ABL. We note that only a quasi-equilibrium state can be obtained as the state of the convective boundary layer varies with time.

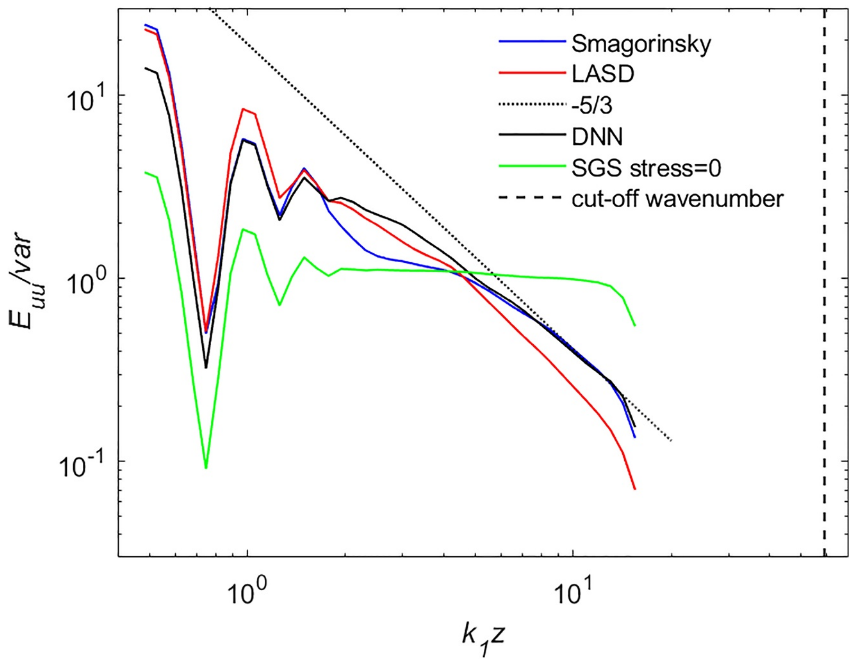

Figure 7. Wavelet spectra (E_{uu}) of streamwise velocity u in the x axis averaged in the spanwise direction at 50% of the height of the computational domain. k_1 is the streamwise wavenumber, z is the distance to the wall, and "var" denotes the variance of velocity series. The cut-off wavenumber is denoted by the vertical dashed line. "Smagorinsky", "LASD", and "DNN" denote the spectra computed after running a sufficient time for equilibrium using the Smagorinsky model, the Lagrangian scale-dependent (LASD) model, and the DNN model, respectively. "SGS stress = 0" denotes the spectrum in an LES where the SGS stresses are set as 0. The −5/3 slope of Kolmogorov's velocity spectra is also denoted.

The proposed DNN model seems to correctly captures the −5/3 scaling of velocity spectra in the convective ABL after a quasi-equilibrium is reached (Figure 7). As the code produces the right −5/3 slope, it is likely that the DNN model should have the right dissipative properties (otherwise we would see a pile-up of energy in the high wavenumber range). We note that the spectra slopes at the highest wavenumbers are much steeper than −5/3, which represents numerical dissipation and is different from turbulence dissipation. In addition, the domain-averaged SGS term (characterizing SGS dissipation (Meneveau & Katz, 2000)) in the kinetic energy budget produced by the DNN model resembles that produced by the Smagorinsky model and model LASD model (Figure S6 in Supporting Information S1) in the quasi-equilibrium period. In particular, the SGS dissipation produced by the DNN model is about 1.3 times of that produced by the LASD model. The stabilization of the LES code is due to the implicit filtering and interaction with a discretized dynamical system. We have conducted an additional LES experiment where the SGS stresses are set as 0, while the rest of the LES setup is the same as that in the DNN-LES experiment. The velocity spectrum of the 0-stress experiment remains almost flat at high wavenumbers, which is significantly different from the −5/3 slope in the DNN model, emphasizing that the high-wavenumber dependence is critically affected by the LES closure and not by numerical dissipation. We do not intend to argue whether the DNN, Smagorinsky or LASD performs best but rather want to highlight that the DNN model produces a reasonable inertial subrange while also capturing flow statistics that are relevant to the ABL community. In addition, the DNN model well captures the SGS stresses as shown in *a priori* tests. This is promising since previous studies (Bastiaans et al., 1998; Jimenez & Moser, 2000; Li & Meneveau, 2004) suggested that SGS stresses and dissipation are generally not simultaneously captured by SGS models, although both are necessary (Meneveau, 1994). For example, the similarity model proposed by Bardina et al. (1980) can produce high correlation between SGS stresses but has to be used together with the Smagorinsky model to capture dissipation. Similarly, the neural network-based SGS model proposed in Beck et al. (2019) did not capture dissipation well and had to be combined with the Smagorinsky model. It is worth noting that there are substantial differences in the spectra between the Smagorinsky model and the LASD model. Compared to the Smagorinsky model, the DNN model and the LASD model produce a shallower transition of spectra above the inertial subrange.

10.1029/2021MS002847

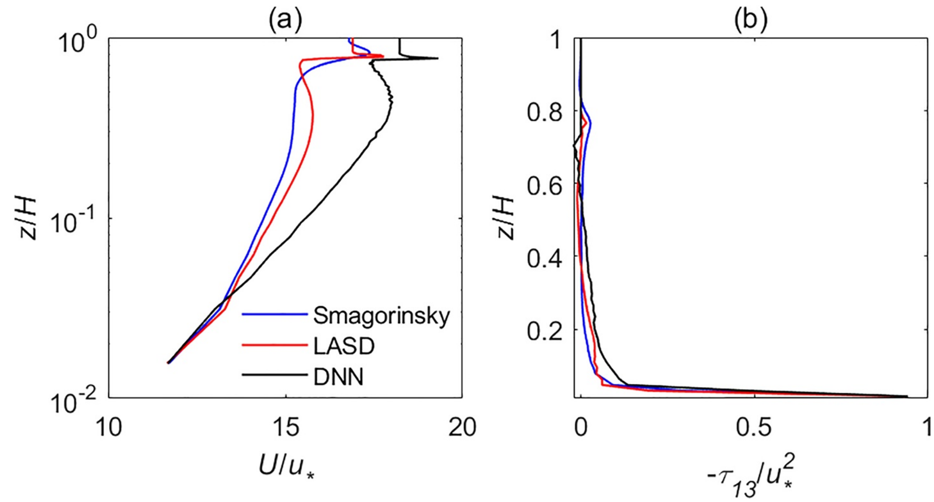

Figure 8. Vertical profile of horizontally averaged (a) streamwise velocity U and (b) normalized SGS stresses in a convective ABL ($z_i/L = -8.7$). "Smagorinsky", "LASD", and "DNN" denote various SGS models. H is height of computational domain.

We also compare mean streamwise velocity from different SGS models (Figure 8a). Near the wall at $\frac{z}{H} < 0.1$ (where H is height of computational domain), the mean velocity produced by the DNN model increases at an overall higher rate than that produced by the Smagorinsky model and LASD model. Thus, the DNN model leads to a higher velocity gradient near the wall and a larger mean velocity in the mixed layer, which is consistent with previous studies (Bou-Zeid et al., 2005; Mason & Thomson, 1992; Porté-Agel et al., 2000) suggesting that more resolved kinetic energy dissipated in the near-wall region leads to a larger mean horizontal velocity in the core of the flow. Near the wall, the SGS stresses produced by the DNN model is generally larger that produced by the Smagorinsky model (Figure 8b), consistent with the *a priori* tests (Figure 4a) showing that the mean of τ_{13} is better captured by the DNN model in most of the domain compared to the Smagorinsky model. We note that many more LES tests are needed to determine whether the proposed SGS model is applicable across most conditions of the ABL. However, the proposed DNN models and tests serve as viable first steps to construct SGS models for LESs of the ABL using deep neural networks.

4. Conclusions

We use deep neural networks to construct SGS models for LESs of turbulence across stability conditions from near neutral to very unstable in the atmospheric boundary layer. The SGS stress at each spatial point is predicted using resolved velocity in the neighboring box in a similar way to convolutional operation. We also use the DNN model to test hypotheses on the structure and physics of the turbulence SGS model. Our primary objective is to shed light on the development of DNN-based SGS modeling across stability conditions. The ultimate goal is not to simply "replace physically based closures with deep learning" in SGS modeling, but to combine these two perspectives in order to use the strengths of deep learning to efficiently harness data and the power of physically rooted equations for better extrapolation. The main findings and suggestions from our study are as follows.

1. SGS models based on DNNs produce more accurate SGS stresses when compared to the Smagorinsky model and the Smagorinsky-Bardina mixed model in terms of correlation, normalized RMSE, mean value, probability distribution and power spectra. This was not necessarily expected *a priori*, given the fact that we used a regular mean square error cost function, which could for instance mostly target the mean state.
2. The DNN model trained on datasets with two extreme stability conditions (high shear and highly convective) performs well when applied to datasets with intermediate stability. This is of particular importance for stability-aware models in LESs.
3. The DNN model produces high correlation in the log-law region (atmospheric surface layer) without inclusion of the wall distance as input parameter.
4. The input dimension of $7 \times 7 \times 7$ and $5 \times 5 \times 5$ do not lead to much improvement over input dimension of $3 \times 3 \times 3$. This suggests that the SGS stress essentially depends on the local neighboring velocity field.

5. Potential temperature and modified pressure do not provide additional predictive skill beyond that provided by velocity.
6. To make predictions on datasets with the same stability (buoyancy effects) as existing training data, it is enough to just train on one data set. To make predictions on other unstable cases at different stability conditions, it is better to train on a wide range of stability conditions that cover the prediction stability.
7. The DNN-based SGS model can be applied to a range of grid resolutions.
8. Based on online (*a posteriori*) tests in LESs, we find that the DNN model can capture the −5/3 scaling for velocity spectra in the convective ABL, which is key to turbulence modeling and correct model dissipation.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

The data are available from Columbia Academic Commons at the following link: https://doi.org/10.7916/d8-y850-6w35.

Acknowledgments
PG would like to acknowledge funding from the National Science Foundation (NSF CAREER, EAR-1552304) and from the Department of Energy (DOE Early Career, DE-SC00142013) and NSF Learning the Earth with Artificial intelligence and Physics (LEAP) Science and Technology Center grant.

References

Agee, E. M., Chen, T., & Dowell, K. (1973). A review of mesoscale cellular convection. *Bulletin of the American Meteorological Society*, *54*(10), 1004–1012. https://doi.org/10.1175/1520-0477(1973)054<1004:AROMCC>2.0.CO;2

Atkinson, B., & Zhang, J. (1996). Mesoscale shallow convection in the atmosphere. *Reviews of Geophysics*, *34*(4), 403–431. https://doi.org/10.1029/96RG02623

Bardina, J., Ferziger, J., & Reynolds, W. (1980). Improved subgrid-scale models for large-eddy simulation. In *Paper presented at 13th fluid and PlasmaDynamics conference*. https://doi.org/10.2514/6.1980-1357

Bar-Sinai, Y., Hoyer, S., Hickey, J., & Brenner, M. P. (2018). Data-driven discretization: A method for systematic coarse graining of partial differential equations, arXiv preprint arXiv:1808.04930.

Bastiaans, R. J., Rindt, C., & Van Steenhoven, A. (1998). Experimental analysis of a confined transitional plume with respect to subgrid-scale modelling. *International journal of heat and mass transfer*, *41*(23), 3989–4007. https://doi.org/10.1016/S0017-9310(98)00075-1

Beck, A., Flad, D., & Munz, C.-D. (2019). Deep neural networks for data-driven LES closure models. *Journal of Computational Physics*, *398*, 108910. https://doi.org/10.1016/j.jcp.2019.108910

Bernard, P. S., & Handler, R. A. (1990). Reynolds stress and the physics of turbulent momentum transport. *Journal of Fluid Mechanics*, *220*, 99–124. https://doi.org/10.1017/S0022112090003202

Bou-Zeid, E., Meneveau, C., & Parlange, M. (2005). A scale-dependent Lagrangian dynamic model for large eddy simulation of complex turbulent flows. *Physics of Fluids*, *17*(2), 025105. https://doi.org/10.1063/1.1839152

Bradley, E. F., Antonia, R., & Chambers, A. (1981). Turbulence Reynolds number and the turbulent kinetic energy balance in the atmospheric surface layer. *Boundary-Layer Meteorol*, *21*(2), 183–197. https://doi.org/10.1007/BF02033936

Cancelli, D. M., Chamecki, M., & Dias, N. L. (2014). A large-eddy simulation study of scalar dissimilarity in the convective atmospheric boundary layer. *Journal of the Atmospheric Sciences*, *71*(1), 3–15. https://doi.org/10.1175/JAS-D-13-0113.1

Canuto, V., & Cheng, Y. (1997). Determination of the Smagorinsky–Lilly constant CS. *Physics of Fluids*, *9*(5), 1368–1378. https://doi.org/10.1063/1.869251

Chamecki, M., Meneveau, C., & Parlange, M. B. (2007). The local structure of atmospheric turbulence and its effect on the Smagorinsky model for large eddy simulation. *Journal of the Atmospheric Sciences*, *64*(6), 1941–1958. https://doi.org/10.1175/JAS3930.1

Cheng, Y., Li, Q., Argentini, S., Sayde, C., & Gentine, P. (2020). A model for Turbulence Spectra in the Equilibrium Range of the Stable Atmospheric Boundary Layer. *Journal of Geophysical Research: Atmospheres*, *125*(5), e2019JD032191. https://doi.org/10.1029/2019JD032191

Cheng, Y., Li, Q., Li, D., & Gentine, P. (2021). Logarithmic profile of temperature in sheared and unstably stratified atmospheric boundary layers. *Physical Review Fluids*, *6*(3), 034606. https://doi.org/10.1103/PhysRevFluids.6.034606

Cheng, Y., Sayde, C., Li, Q., Basara, J., Selker, J., Tanner, E., & Gentine, P. (2017). Failure of Taylor's hypothesis in the atmospheric surface layer and its correction for eddy-covariance measurements. *Geophysical Research Letters*, *44*(9), 4287–4295. https://doi.org/10.1002/2017GL073499

Chorin, A. J. (1968). Numerical solution of the Navier-Stokes equations. *Mathematics of Computation*, *22*(104), 745–762. https://doi.org/10.1090/S0025-5718-1968-0242392-2

Chow, F. K., & Moin, P. (2003). A further study of numerical errors in large-eddy simulations. *Journal of Computational Physics*, *184*(2), 366–380. https://doi.org/10.1016/S0021-9991(02)00020-7

Chung, D., & Matheou, G. (2014). Large-eddy simulation of stratified turbulence. Part I: A vortex-based subgrid-scale model. *Journal of the Atmospheric Sciences*, *71*(5), 1863–1879. https://doi.org/10.1175/JAS-D-13-0126.1

Clark, R. A., Ferziger, J. H., & Reynolds, W. C. (1979). Evaluation of subgrid-scale models using an accurately simulated turbulent flow. *Journal of Fluid Mechanics*, *91*(1), 1–16. https://doi.org/10.1017/S002211207900001X

Corrsin, S. (1975). Limitations of gradient transport models in random walks and in turbulence. In F. N. Frenkiel & R. E. Munn (Eds.), *Advances in geophysics* (Vol. 18, pp. 25–60). Elsevier. https://doi.org/10.1016/S0065-2687(08)60451-3

Deardorff, J. W. (1970). A numerical study of three-dimensional turbulent channel flow at large Reynolds numbers. *Journal of Fluid Mechanics*, *41*(2), 453–480. https://doi.org/10.1017/S0022112070000691

Deardorff, J. W. (1972). Numerical investigation of neutral and unstable planetary boundary layers. *Journal of the Atmospheric Sciences*, *29*(1), 91–115. https://doi.org/10.1175/1520-0469(1972)029<0091:NIONAU>2.0.CO;2

Dougherty, J. (1961). The anisotropy of turbulence at the meteor level. *Journal of Atmospheric and Terrestrial Physics*, *21*(2–3), 210–213. https://doi.org/10.1016/0021-9169(61)90116-7

Egolf, P. W. (1994). Difference-quotient turbulence model: A generalization of Prandtl's mixing-length theory. *Physical Review E*, *49*(2), 1260–1268. https://doi.org/10.1103/PhysRevE.49.1260

Gamahara, M., & Hattori, Y. (2017). Searching for turbulence models by artificial neural network. *Physical Review Fluids*, *2*(5), 054604. https://doi.org/10.1103/PhysRevFluids.2.054604

Germano, M., Piomelli, U., Moin, P., & Cabot, W. H. (1991). A dynamic subgrid-scale eddy viscosity model. *Physics of Fluids A: Fluid Dynamics*, *3*(7), 1760–1765. https://doi.org/10.1063/1.857955

Giometto, M., Lozano-Durán, A., Park, G., & Moin, P. (2017). Three-dimensional transient channel flow at moderate Reynolds numbers: Analysis and wall modeling. *CTR Annu. Res. Briefs*, 193–205. Report.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Paper presented at Proceedings of the IEEE conference on computer vision and pattern recognition*. https://doi.org/10.1109/CVPR.2014.81

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning*. MIT press Cambridge.

Grachev, A. A., Andreas, E. L., Fairall, C. W., Guest, P. S., & Persson, P. O. G. (2015). Similarity theory based on the Dougherty-Ozmidov length scale. *Q J R Meteorol Soc*, *141*(690), 1845–1856. https://doi.org/10.1002/qj.2488

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors, arXiv preprint arXiv:1207.0580.

Horiuti, K. (1989). The role of the Bardina model in large eddy simulation of turbulent channel flow. *Physics of Fluids A: Fluid Dynamics*, *1*(2), 426–428. https://doi.org/10.1063/1.857465

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift, arXiv preprint arXiv:1502.03167.

Jimenez, J., & Moser, R. D. (2000). Large-eddy simulations: Where are we and what can we expect? *AIAA journal*, *38*(4), 605–612. https://doi.org/10.2514/2.1031

Kleissl, J., Meneveau, C., & Parlange, M. B. (2003). On the magnitude and variability of subgrid-scale eddy-diffusion coefficients in the atmospheric surface layer. *Journal of the Atmospheric Sciences*, *60*(19), 2372–2388. https://doi.org/10.1175/1520-0469(2003)060<2372:OTMAVO>2.0.CO;2

Kolmogorov, A. N. (1941). The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *Dokl. Akad. Nauk SSSR*, *30*, 299–303.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Paper presented at Advances in neural information processing systems*.

Kumar, V., Kleissl, J., Meneveau, C., & Parlange, M. B. (2006). Large-eddy simulation of a diurnal cycle of the atmospheric boundary layer: Atmospheric stability and scaling issues. *Water Resources Research*, *42*(6), W06D09. https://doi.org/10.1029/2005WR004651

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, *1*(4), 541–551. https://doi.org/10.1162/neco.1989.1.4.541

Lee, M., & Moser, R. D. (2015). Direct numerical simulation of turbulent channel flow up to $Re\tau \approx 5200$. *Journal of Fluid Mechanics*, *774*, 395–415. https://doi.org/10.1017/jfm.2015.268

Leonard, A. (1975). Energy cascade in large-eddy simulations of turbulent fluid flows. In F. N. Frenkiel & R. E. Munn (Eds.), *Advances in geophysics* (Vol. 18, pp. 237–248). Elsevier. https://doi.org/10.1016/S0065-2687(08)60464-1

Li, D., & Bou-Zeid, E. (2011). Coherent structures and the dissimilarity of turbulent transport of momentum and scalars in the unstable atmospheric surface layer. *Boundary-Layer Meteorology*, *140*(2), 243–262. https://doi.org/10.1007/s10546-011-9613-5

Li, Q., Gentine, P., Mellado, J. P., & McColl, K. A. (2018). Implications of nonlocal transport and conditionally averaged statistics on Monin–Obukhov similarity theory and Townsend's attached eddy hypothesis. *Journal of the Atmospheric Sciences*, *75*(10), 3403–3431. https://doi.org/10.1175/JAS-D-17-0301.1

Li, Y., & Meneveau, C. (2004). Analysis of mean momentum flux in subgrid models of turbulence. *Physics of Fluids*, *16*(9), 3483–3486. https://doi.org/10.1063/1.1773846

Lilly, D. (1967). The representation of small-scale turbulence in numerical simulation experiments. *Paper presented at Proceedings of IBM Scientific Computation on Symposium for Environmental Science*. White Plains, IBM.

Lilly, D. K. (1992). A proposed modification of the Germano subgrid-scale closure method. *Physics of Fluids A: Fluid Dynamics*, *4*(3), 633–635. https://doi.org/10.1063/1.858280

Ling, J., Kurzawski, A., & Templeton, J. (2016). Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. *Journal of Fluid Mechanics*, *807*, 155–166. https://doi.org/10.1017/jfm.2016.615

Liu, S., Meneveau, C., & Katz, J. (1994). On the properties of similarity subgrid-scale models as deduced from measurements in a turbulent jet. *Journal of Fluid Mechanics*, *275*, 83–119. https://doi.org/10.1017/S0022112094002296

Lumley, J. L. (1981). Coherent structures in turbulence. In R. E. Meyer (Ed.), *Transition and turbulence* (pp. 215–242). Elsevier. https://doi.org/10.1016/B978-0-12-493240-1.50017-X

Marusic, I., Monty, J. P., Hultmark, M., & Smits, A. J. (2013). On the logarithmic region in wall turbulence. *Journal of Fluid Mechanics*, *716*, R3. https://doi.org/10.1017/jfm.2012.511

Mason, P. J. (1989). Large-eddy simulation of the convective atmospheric boundary layer. *Journal of the Atmospheric Sciences*, *46*(11), 1492–1516. https://doi.org/10.1175/1520-0469(1989)046<1492:LESOTC>2.0.CO;2

Mason, P. J., & Brown, A. R. (1999). On subgrid models and filter operations in large eddy simulations. *Journal of the Atmospheric Sciences*, *56*(13), 2101–2114. https://doi.org/10.1175/1520-0469(1999)056<2101:OSMAFO>2.0.CO;2

Mason, P. J., & Thomson, D. J. (1992). Stochastic backscatter in large-eddy simulations of boundary layers. *Journal of Fluid Mechanics*, *242*, 51–78. https://doi.org/10.1017/S0022112092002271

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, *5*(4), 115–133. https://doi.org/10.1007/BF02478259

Meneveau, C. (1994). Statistics of turbulence subgrid-scale stresses: Necessary conditions and experimental tests. *Physics of Fluids*, *6*(2), 815–833. https://doi.org/10.1063/1.868320

Meneveau, C., & Katz, J. (2000). Scale-invariance and turbulence models for large-eddy simulation. *Annual Review of Fluid Mechanics*, *32*(1), 1–32. https://doi.org/10.1146/annurev.fluid.32.1.1

Moeng, C.-H. (1984). A large-eddy-simulation model for the study of planetary boundary-layer turbulence. *Journal of the Atmospheric Sciences*, *41*(13), 2052–2062. https://doi.org/10.1175/1520-0469(1984)041<2052:ALESMF>2.0.CO;2

Monin, A., & Obukhov, A. (1954). Basic laws of turbulent mixing in the surface layer of the atmosphere. *Contrib. Geophys. Inst. Acad. Sci. USSR*, *151*, 163–187.

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. *Paper presented at proceedings of the 27th international conference on machine learning (ICML-10)*.

Nicoud, F., & Ducros, F. (1999). Subgrid-scale stress modelling based on the square of the velocity gradient tensor. *Flow, Turbulence and Combustion*, *62*(3), 183–200. https://doi.org/10.1023/A:1009995426001

Nieuwstadt, F. T., Mason, P. J., Moeng, C.-H., & Schumann, U. (1993). Large-eddy simulation of the convective boundary layer: A comparison of four computer codes. In F. Durst, R. Friedrich, B. E. Launder, F. W. Schmidt, U. Schumann, & J. H. Whitelaw (Eds.), *Turbulent shear flows 8* (pp. 343–367). Springer. https://doi.org/10.1007/978-3-642-77674-8_24

Obukhov, A. (1946). Turbulence in thermally inhomogeneous atmosphere. *Trudy Inst. Teor. Geofiz. Akad. Nauk SSSR*, *1*, 95–115.

Orlandi, P. (2012). *Fluid flow phenomena: A numerical toolkit*. Springer Science & Business Media. https://doi.org/10.1007/978-94-011-4281-6

Ozmidov, R. (1965). On the turbulent exchange in a stably stratified ocean. *Izv. Acad. Sci. USSR, Atmos. Oceanic Phys.*, *1*, 861–871.

Panofsky, H. A. (1963). Determination of stress from wind and temperature measurements. *Quarterly Journal of the Royal Meteorological Society*, *89*(379), 85–94. https://doi.org/10.1002/qj.49708937906

Paulson, C. A. (1970). The mathematical representation of wind speed and temperature profiles in the unstable atmospheric surface layer. *Journal of Applied Meteorology*, *9*(6), 857–861. https://doi.org/10.1175/1520-0450(1970)009<0857:TMROWS>2.0.CO;2

Pope, S. (2000). *Turbulent flows* (p. 771). Cambridge University Press. https://doi.org/10.1017/CBO9780511840531

Porté-Agel, F., Meneveau, C., & Parlange, M. B. (2000). A scale-dependent dynamic model for large-eddy simulation: Application to a neutral atmospheric boundary layer. *Journal of Fluid Mechanics*, *415*, 261–284. https://doi.org/10.1017/S0022112000008776

Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. In *Proceedings of the national academy of sciences* (Vol. 115(39), pp. 9684–9689). https://doi.org/10.1073/pnas.1810286115

Salesky, S. T., Chamecki, M., & Bou-Zeid, E. (2017). On the nature of the transition between roll and cellular organization in the convective boundary layer. *Boundary-Layer Meteorology*, *163*(1), 41–68. https://doi.org/10.1007/s10546-016-0220-3

Schmitt, F. G. (2007). About Boussinesq's turbulent viscosity hypothesis: Historical remarks and a direct evaluation of its validity. *Comptes Rendus Mécanique*, *335*(9–10), 617–627. https://doi.org/10.1016/j.crme.2007.08.004

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., & LeCun, Y. (2013). Overfeat: Integrated recognition, localization and detection using convolutional networks, arXiv preprint arXiv:1312.6229.

Shah, S. K., & Bou-Zeid, E. (2014). Direct numerical simulations of turbulent Ekman layers with increasing static stability: Modifications to the bulk structure and second-order statistics. *Journal of Fluid Mechanics*, *760*, 494–539. https://doi.org/10.1017/jfm.2014.597

Smagorinsky, J. (1963). General circulation experiments with the primitive equations: I. The basic experiment. *Monthly Weather Review*, *91*(3), 99–164. https://doi.org/10.1175/1520-0493(1963)091<0099:GCEWTP>2.3.CO;2

Speziale, C. G. (1991). Analytical methods for the development of Reynolds-stress closures in turbulence. *Annual Review of Fluid Mechanics*, *23*(1), 107–157. https://doi.org/10.1146/annurev.fl.23.010191.000543

Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. *Paper presented at Proceedings of the IEEE conference on computer vision and pattern recognition*. https://doi.org/10.1109/CVPR.2014.220

Tennekes, H., Lumley, J. L., & Lumley, J. (1972). *A first course in turbulence*. MIT press.

Voller, V., & Porte-Agel, F. (2002). Moore's law and numerical modeling. *Journal of Computational Physics*, *179*(2), 698–703. https://doi.org/10.1006/jcph.2002.7083

von Kármán, T. (1930). Mechanische ahnlichkeit und turbulenz. In *Paper presented at proceedings of the 3rd international congress on applied mechanics*.

Wray, A. A. (1990). *Minimal storage time advancement schemes for spectral methods*. Report No. MS, 202. NASA Ames Research Center.

Xie, C., Wang, J., & Weinan, E. (2020). Modeling subgrid-scale forces by spatial artificial neural networks in large eddy simulation of turbulence. *Physical Review Fluids*, *5*(5), 054606. https://doi.org/10.1103/PhysRevFluids.5.054606

Yamamoto, Y., & Tsuji, Y. (2018). Numerical evidence of logarithmic regions in channel flow at $Re_\tau = 8000$. *Physical Review Fluids*, *3*(1), 012602. https://doi.org/10.1103/PhysRevFluids.3.012602

Yang, X., Zafar, S., Wang, J.-X., & Xiao, H. (2019). Predictive large-eddy-simulation wall modeling via physics-informed neural networks. *Physical Review Fluids*, *4*(3), 034602. https://doi.org/10.1103/PhysRevFluids.4.034602