

An Investigation of LES Wall Modeling for Rayleigh–Bénard Convection via Interpretable and Physics-Aware Feedforward Neural Networks with DNS

AARON WANG^{}^a, XIANG I. A. YANG^b, AND MIKHAIL OVCHINNIKOV^a

^a *Pacific Northwest National Laboratory, Richland, Washington*

^b *The Pennsylvania State University, University Park, Pennsylvania*

(Manuscript received 22 May 2023, in final form 7 December 2023, accepted 17 December 2023)

ABSTRACT: The traditional approach of using the Monin–Obukhov similarity theory (MOST) to model near-surface processes in large-eddy simulations (LESs) can lead to significant errors in natural convection. In this study, we propose an alternative approach based on feedforward neural networks (FNNs) trained on output from direct numerical simulation (DNS). To evaluate the performance, we conduct both a priori and a posteriori tests. In the a priori (offline) tests, we compare the statistics of the surface shear stress and heat flux, computed from filtered DNS input variables, to the stress and flux obtained from the filtered DNS. Additionally, we investigate the importance of various input features using the Shapley additive explanations value and the conditional average of the filter grid cells. In the a posteriori (online) tests, we implement the trained models in the System for Atmospheric Modeling (SAM) LES and compare the LES-generated surface shear stress and heat flux with those in the DNS. Our findings reveal that vertical velocity, a traditionally overlooked flow quantity, is one of the most important input features for determining the wall fluxes. Increasing the number of input features improves the a priori test results but does not always improve the model performance in the a posteriori tests because of the differences in input variables between the LES and DNS. Last, we show that physics-aware FNN models trained with logarithmic and scaled parameters can well extrapolate to more intense convection scenarios than in the training dataset, whereas those trained with primitive flow quantities cannot.

SIGNIFICANCE STATEMENT: The traditional near-surface turbulence model, based on a shear-dominated boundary layer flow, does not represent near-surface turbulence in natural convection. Using a feedforward neural network (FNN), we can construct a more accurate model that better represents the near-surface turbulence in various flows and reveals previously overlooked controlling factors and process interactions. Our study shows that the FNN-generated models outperform the traditional model and highlight the importance of the near-surface vertical velocity. Furthermore, the physics-aware FNN models exhibit the potential to extrapolate to convective flows of various intensities beyond the range of the training dataset, suggesting their broader applicability for more accurate modeling of near-surface turbulence.

KEYWORDS: Convection; Turbulence; Boundary layer; Large eddy simulations; Subgrid-scale processes; Machine learning

1. Introduction

Wall modeling, an approach to modeling near-surface turbulence, is a crucial component in large-eddy simulations (LESs) of wall-bounded flows at high Reynolds numbers (Piomelli and Balaras 2002; Bose and Park 2018; Yang and Griffin 2021), including flows in the atmospheric boundary layer (Deardorff 1980; Moeng 1984; Bou-Zeid et al. 2005; Salesky and Anderson 2018). Despite its importance, there is currently no one-size-fits-all approach for all types of flow (Slotnick et al. 2014; Wang et al. 2023). Traditional wall models are mostly derived in flows that are quasi steady, horizontally homogeneous in surface and flow properties, and horizontally dominated (i.e., the horizontal components of the near-surface velocity are much larger than the vertical

component), which are not necessarily suitable for unsteady and locally convective flows.

One of the most widely used wall models to account for hydrostatic stability is based on the Monin–Obukhov similarity theory (MOST; Monin and Obukhov 1954), which assumes a horizontally homogeneous, quasi-steady, and predominantly horizontal flow. In a neutral boundary layer, MOST reduces to the law of the wall (Prandtl 1933). Applying the law of the wall for wall modeling is equivalent to the drag coefficient approach that uses a constant roughness length and first-level vertical grid spacing (e.g., Roberts et al. 2020). However, the limitations of MOST have been well known (Stiperski and Calaf 2023). For example, MOST fails to account for horizontal wind variances (Wyngaard and Coté 1974), wall normal velocity variances, and temperature variances (Mahrt 1999). Additionally, its overly idealized assumptions call into question its broad applicability, for example, in Rayleigh–Bénard convection (RBC) and other natural convection flows.

In spite of the long-standing limitations of MOST, without a more reliable alternative, atmospheric models have used MOST for evaluating surface fluxes, even for the flow over complex terrain (e.g., Schumann and 1990) and intense convection

 Denotes content that is immediately available upon publication as open access.

Corresponding author: Aaron Wang, aaron.wang@pnnl.gov

(e.g., Schenkman et al. 2014; Markowski and Bryan 2016; Roberts et al. 2020), which clearly violate the MOST assumptions. Numerical simulations have shown that the vertical velocity induced by horizontal convergence (implying a horizontally heterogeneous velocity field) is sensitive to the use of wall models (Wang et al. 2020, 2023). Without relying on the MOST, the non-equilibrium wall models have also been developed in the engineering community (Balaras et al. 1996; Cabot and Moin 2000; Wang and Moin 2002; Park and Moin 2014), but currently, their applications have only been explored in shear-dominated boundary layers, and therefore further discussion of these approaches is not pursued here.

RBC is an example of a locally and horizontally heterogeneous phenomenon that violates the MOST assumptions. RBC is characterized by local warm updrafts and cold downdrafts, where the local near-surface vertical velocity can be on the same order of magnitude as the horizontal velocity. Previous studies have revealed the deviation from MOST in the updraft and downdraft of convection (e.g., Fodor et al. 2019), but MOST is still often used in the LES of RBC to model near-surface turbulence while accounting for near-surface stability (e.g., Salesky and Anderson 2018; Thomas et al. 2019; Yang et al. 2022).

Efforts to construct wall models using neural networks (NNs) began as early as Milano and Koumoutsakos (2002). Over the last decade, significant advances in machine learning (ML) and high-performance computing have provided greater opportunities to develop innovative wall modeling approaches based on NNs. For instance, LES wall models based on NNs have been constructed for canonical boundary layers (e.g., Yang et al. 2019; Bae and Koumoutsakos 2022; Vadrot et al. 2023), periodic hills (Zhou et al. 2021), rotating channels (Huang et al. 2019; Huang and Yang 2021), and irregular surfaces on aircrafts (Lozano-Durán and Bae 2020, 2023). To the best of our knowledge, however, LES wall models for RBCs using an NN-based approach have yet to be explored. Therefore, the objective of this work is to employ NNs to develop wall models as alternatives to the MOST-based models for LES of RBC and to gain a deeper understanding of the mechanisms governing the surface shear stress and heat flux.

For training and evaluation purposes, we perform direct numerical simulations (DNSs), which resolve all scales of motion and do not require a wall model, and then filter the DNS results to resolutions that are relevant to LES. We employ a feedforward neural network (FNN) to relate the near-wall flow properties to the surface shear stress and heat flux. We also use the same DNS results to evaluate the accuracy of the MOST-based model for the wall fluxes. Compared to the previous ML-wall-model studies, the complexity of the flow considered in this paper is the involvement of thermal stratification and heat transfer. It is not entirely clear what off-wall information should be used when modeling the wall shear stress and wall heat flux. The absence of such knowledge puts us at risk of overfitting or underfitting. To determine the most relevant input features, we evaluate the importance of each input using the Shapley additive explanations values (SHAP value; Shapley 1953; Štrumbelj and Kononenko 2014), which, combined with deep learning, is a form of interpretable ML. To better understand the SHAP

values and gain the physical insight, we also analyze the velocity and temperature profiles within the filter grid cells conditionally selected according to the low and high ranges of the input variables.

One challenge for ML algorithms is to apply the trained model beyond the scope of its training data. Yang et al. (2019) demonstrated that the NN including physical information has the capability to model the boundary layer at a Reynolds number higher than that used to train the NN. On the contrary, an NN trained without incorporating any physical information struggles to extrapolate. For RBC, however, there is no well-known relationship for the near-surface flow quantities like that between the velocity and wall-normal distance in a canonical boundary layer. The NN ability to extrapolate can be improved by using dimensionless parameters within the local and near-surface regions, which allows NN to account for relationships observed in large-scale circulations, such as power laws among the Nusselt number, Reynolds number, and Rayleigh number (Ahlers et al. 2009). As we do not invoke the governing equation in our training, we refer to our approach as the “physics-aware” approach to distinguish it from the physics-informed NN approach (Cai et al. 2021), where one penalizes the solution based on the governing equation. Specifically, the NNs here are called physics-aware feedforward neural networks (PAFNNs).

In this paper, we first evaluate the performance of the NN models (referring to both FNN and PAFNN models hereafter) and the importance of the features using a priori tests. Specifically, we will use DNS-filtered data as input to the models and compare their output surface shear stress and heat flux with the DNS-filtered ones. Later, to evaluate their applicability, we will implement the NN models in LES for the a posteriori tests. To assess the extrapolation capability of the NN models, an additional set of DNS and LES runs with more intense RBC is performed.

The rest of the paper is organized as follows: Section 2 introduces the methods, which cover the DNS configuration, the filtering approach to mimic the LES data, the MOST wall model, the deep learning configuration, the SHAP value, and the a posteriori tests. Section 3 presents the results. Finally, section 4 provides the conclusions. For the readers' reference, we place the flow statistics from the interior of the domain in appendix A and more details of the training process in appendix B.

2. Methods

The training data used in this work are filtered from DNS. Therefore, we first present the DNS model description, simulation setup, and the filtering approach used to obtain data relevant to the LES grid resolutions. Next, we introduce the traditional MOST and several newly trained NN models, explaining how the filtered data are used as inputs. To understand the importance of the inputs for each FNN wall model, we introduce the SHAP value. Finally, we describe how we implement the NN models in LES for a posteriori tests in both the training and the extrapolation scenarios.

a. DNS

The DNS is performed using the System for Atmospheric Modeling (SAM), which solves dimensionalized equations for velocity and moist static energy (see details in [Khairoutdinov and Randall 2003](#), appendix A). The velocity components are solved on a staggered Arakawa C grid and advected using a second-order central scheme. The scalar is solved at the center of each grid cell and advected using a multidimensional positive-definite advection transport algorithm ([Smolarkiewicz and Grabowski 1990](#)). Pressure is diagnosed by solving a Poisson equation. Moisture is not introduced in the simulations of this study.

SAM was originally designed for atmospheric LES but has recently been adapted to simulate cloud chambers, which have a scale similar to the case in this study ([Thomas et al. 2019, 2023](#); [Yang et al. 2022](#); [Wang et al. 2024](#)). For the DNS used in this study, the anelastic approximation is replaced by the Boussinesq approximation. Additionally, the eddy viscosity is substituted with the kinematic viscosity ($\nu = \mu/\rho$, where $\mu = 1.717 \times 10^{-5}$ Pa s is the dynamic viscosity of air, and $\rho = 1.2096$ kg m⁻³ is the density of air), and the wall model is replaced by an explicit no-penetration and no-slip boundary condition. The Prandtl number is 0.72. The simulation setup largely follows the DNS performed by [Chandrakar et al. \(2022\)](#): a 1 m³ cube domain with a cold top surface of 282 K and a warm bottom surface of 294 K (indicated as $\Delta T = 12$ K hereafter), yielding a Rayleigh number of $\sim 10^9$ ([Table 1](#)). Different from [Chandrakar et al. \(2022\)](#), sidewalls are removed in this work, and the doubly periodic lateral boundary conditions are applied for simplicity. The horizontal size of the domain results in a domain aspect ratio of 1, which may not be ideal for studying the structure of RBC ([Grötzbach 1983](#); [Stevens et al. 2018](#)), but it is sufficient for capturing the characteristics of near-surface turbulence within RBC.

According to the observation in the Pi chamber ([Chang et al. 2016](#); [Chandrakar et al. 2016](#)), the maximum magnitude of vertical velocity has an order of 0.1 m s⁻¹, which is taken as the turbulent velocity scale (v). With the length scale (l) of 1 m, the dissipation rate (ϵ) and the Kolmogorov length scale (η) can be derived as

$$\epsilon = \frac{v^3}{l} \sim 10^{-3} \text{ m}^2 \text{ s}^{-3}, \quad (1)$$

$$\eta = \left(\frac{\nu^3}{\epsilon}\right)^{1/4} \sim 10^{-3} \text{ m}. \quad (2)$$

The DNS has 512 grid points in each direction (similar to those applied in [Chandrakar et al. 2022](#)), yielding a homogeneous and constant grid spacing of ~ 2 mm. The time step is adjustable to account for the Courant–Friedrichs–Lewy criterion. The quasi-steady state of mean surface heat flux is reached in roughly 1 min ([Fig. A1](#)), and the result at 1–10 min is used for training. Although a quasi-steady state of domain-averaged quantities is reached, the locations of updraft and downdraft shift horizontally, which can represent the flow's horizontal heterogeneity and unsteadiness. An illustration of

TABLE 1. The Ra and resulting Nu from the DNS runs.

ΔT (K)	Ra	Nu	$\text{Nu Ra}^{-1/3}$
12	1.472×10^9	74.5	6.547×10^{-2}
24	2.945×10^9	93.9	6.549×10^{-2}

the resulting temperature field, turbulent coherent structures, and energy spectrum is shown in [Fig. 1](#). Because the focus of this research is on the wall model rather than the structure of RBC, we leave the statistics of domainwide flow quantities in [appendix A](#) and focus on the near-surface area in the main context hereafter.

The definition of boundary layer depth in this study is complicated by the slightly negatively tilted temperature profile observed within the domain's interior (see [Figs. A2a,b](#)), which is a consequence of the limited horizontal domain with periodic boundaries. If we define the boundary layer by temperature gradient ($\partial T/\partial z$; i.e., as $\partial T/\partial z$ is 99% close to the centerline value, where z is the distance from the surface), there are 12 grid points in the boundary layer, which meets the requirement of roughly 10 points for this Ra ([Stevens et al. 2010](#)). [Table 1](#) shows that the resulting Nusselt numbers align with experimental data from [Niemela et al. \(2000\)](#) and [Chavanne et al. \(2001\)](#), as well as the DNS from [Stevens et al. \(2010\)](#) (see the intercomparison of $\text{Nu Ra}^{-1/3}$ in [Stevens et al. 2010](#), Fig. 1a).

For an a posteriori test, a separate DNS run is carried out with a 276-K cold top surface and a 300-K warm bottom surface (indicated as $\Delta T = 24$ K hereafter). The grid resolution follows the DNS with $\Delta T = 12$ K. If we define the boundary layer by $\partial T/\partial z$ as above, the boundary layer will contain 11 grid points. Concerning the limitation of computational resource, this run is performed for a physical time of 4 min, but the data are enough for analysis ([Fig. A1](#)). A comparison of the energy spectra of two DNS runs is presented in [Fig. 1b](#). It should be noted that this second DNS run is not used for training the NN models but instead is employed to evaluate the extrapolation ability of the NN models after they are implemented in LES.

[Table 1](#) demonstrates that the two DNS runs yield similar values of $\text{Nu Ra}^{-1/3}$, and the values are consistent with the experiments and DNS results presented by [Stevens et al. \(2010\)](#), [Fig. 1a](#). However, [Stevens et al. \(2010\)](#) show that a broader Ra range reveals a decrease in $\text{Nu Ra}^{-1/3}$ with increasing Ra, implying that the true scaling power of Ra for Nu is less than $1/3$ [see [Grossmann and Lohse \(2000, Table 1\)](#) for the scaling power in various experiments].

b. Data preparation

For parameterization development and evaluation, the DNS output is mapped to the LES grid (allowing the overlapping of filter grid cells to increase the amount of data) by applying a top-hat filter (i.e., DNS results are averaged over each LES grid cell). The filter length is chosen to be consistent with the grid spacing used in recent LES studies of cloud chambers, ranging from 3.125 to 6.25 cm ([Thomas et al. 2019, 2023](#); [Yang et al. 2022](#); [Wang et al. 2024](#)). Note that these

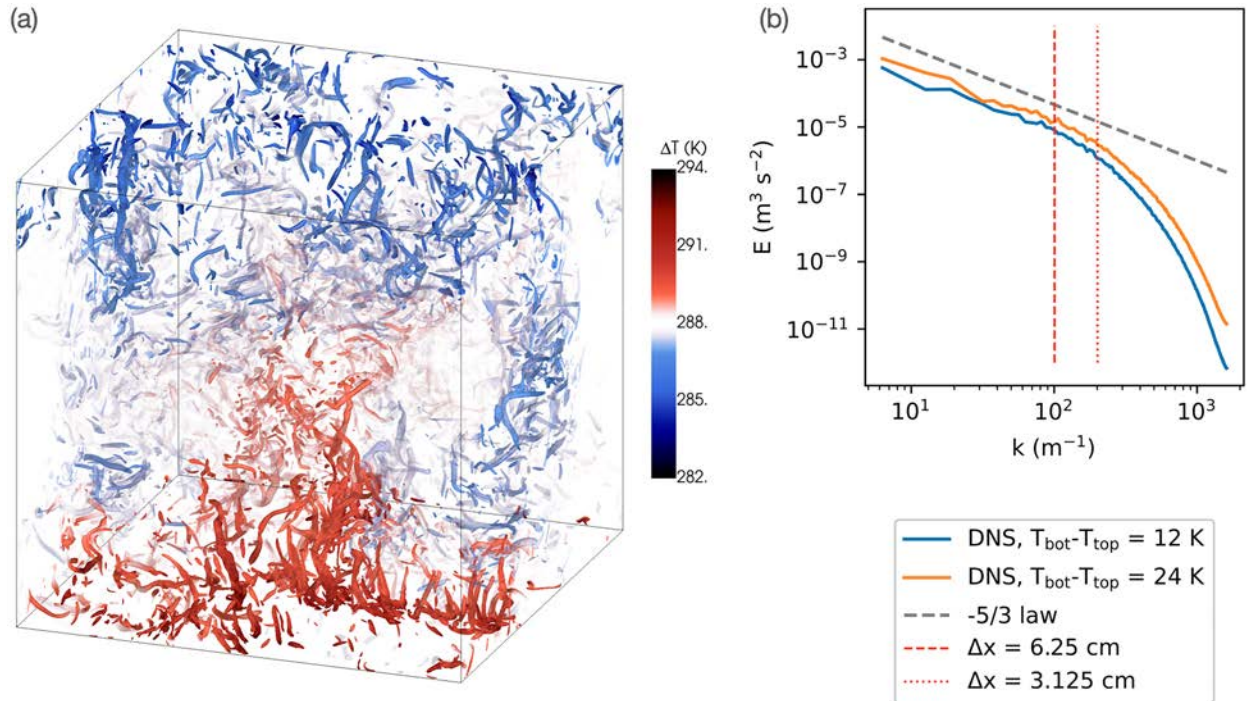


FIG. 1. An illustration of the DNS runs. (a) The turbulent coherent structure is visualized by a snapshot of the DNS run with $T_{\text{bot}} - T_{\text{top}} = 12$ K (where $T_{\text{bot}} - T_{\text{top}}$ indicates the bottom minus top temperature difference) at $t = 5$ min using isosurface of $\lambda_2 = -40 \text{ s}^{-2}$ (colored by temperature with transparency indicated by the white region in the color bar), which represents the local pressure minimum owing to vortical motions and thus visualizes turbulent vortex tubes (Jeong and Hussain 1995). (b) The horizontal energy density spectra from the central plane ($z = 0.5$ m) of two DNS runs are presented, where blue and orange lines represent the spectra at the last snapshots in the DNS runs of $T_{\text{bot}} - T_{\text{top}} = 12$ and 24 K, respectively; gray line represents the slope of the inertial subrange (Kolmogorov 1941); and red dashed and dotted lines represent the range of the filter length.

filter lengths are within the small-scale part of the inertial subrange, as shown by the red dashed and dotted lines in Fig. 1b, confirming the appropriateness of these grid spacings for LES.

The inputs to the MOST and NN models include the distance from the walls (h , which is half of the filter length) and

two filtered variables: magnitude of horizontal wind speed ($|\tilde{U}| = \sqrt{\tilde{u}^2 + \tilde{v}^2}$, where u and v are zonal and meridional wind components) and temperature difference between the surface and the first imaginary LES grid cell ($\Delta\tilde{T}$). Here, the tilde over each variable indicates that it is filtered over the LES grid cell. For the NN models, three more variables

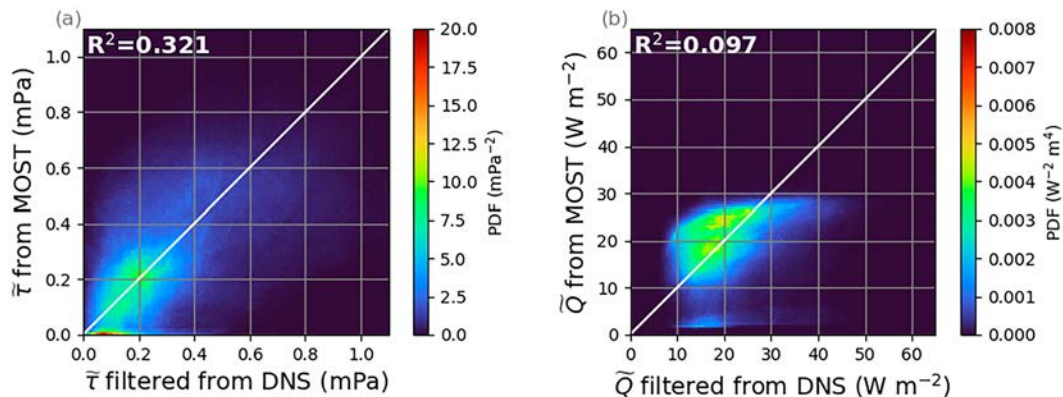


FIG. 2. Filtered (a) surface shear stress and (b) heat flux in DNS compared to those modeled by the traditional MOST. The upper-left corner of each panel shows the coefficient of determination (R^2 , also known as the explained proportion of the total deviance; Di Mari et al. 2023).

are filtered for the inputs: vertical velocity ($\tilde{w}$), magnitude of pressure gradient force ($\rho^{-1}|\nabla\tilde{p}|$), and the angle between horizontal wind and pressure gradient force (θ , measured as $\cos\theta$). The surface shear stress ($\tilde{\tau}$) and the heat flux ($\tilde{Q}$) are filtered over the square area beneath the filter grid cell of the inputs as the target values to train the NN models and evaluate the accuracy of the MOST and NN models. Notice that we use velocity as a model input. Including velocity as input is usually not advisable since velocity is not Galilean invariant. However, since wall modeling assumes a stationary wall, invoking the velocity as an input here does not violate Galilean invariance.

Information in the NN models is fed forward through activation function. Before being fed forward, the features in the input layer are standardized. Similarly, the target values in the output layer are standardized before being used for penalties. The standardization is carried out as

$$\phi_{\text{input/output_layer}} = \frac{\phi - \mu(\phi)}{\sigma(\phi)}, \quad (3)$$

where ϕ represents any input feature or target value; μ and σ indicate mean and standard deviation of the training data, respectively.

For training the PAFNN models, a choice needs to be made about whether to invoke global (as in Zhou et al. 2021) or only local quantities for input. Because a wall model is intended to be applicable to any geometry and flow type, using only local variables is highly preferred (Yang et al. 2017).

The local $|\tilde{U}|$ and $\Delta\tilde{T}$ are combined with h to form the following nondimensional parameters:

$$\text{Re}_{hU} = \frac{|\tilde{U}|h}{\nu}, \quad (4)$$

$$\text{Ra}_h = \frac{g\beta\Delta\tilde{T}h^3}{\nu\alpha}, \quad (5)$$

where $g = 9.81 \text{ m s}^{-2}$ is the gravitational acceleration, $\beta = 3.5 \times 10^{-3} \text{ K}^{-1}$ is the thermal expansion coefficient, and $\alpha = \nu/\text{Pr}$ is the thermal diffusivity; Re_{hU} is a local Reynolds number based on the magnitude of horizontal velocity at a distance h from the surface, and Ra_h is a local Rayleigh number based on the temperature difference between the surface and a distance h . For the vertical velocity scaling, we take

$$\text{Fr}_w = \frac{\tilde{w}}{\sqrt{g\beta\Delta\tilde{T}h}}, \quad (6)$$

which is analogous to the Froude number based on vertical velocity and buoyant acceleration. A different scaling using Reynolds number based on vertical velocity and the ratio of vertical velocity to wind speed has also been explored, but Eq. (6) is found to perform better in this study. Because vertical velocity and buoyancy can affect each other, Fr_w can be viewed as the vertical velocity of large-scale circulation compared to that initiated by hydrostatic stability. If $\text{Fr}_w = 1$, the vertical velocity is purely dominated by hydrostatic stability.

The influence of the pressure gradient force, $\rho^{-1}|\nabla\tilde{p}|$, is accounted for through the Euler number:

$$\text{Eu}_h = \frac{|\nabla\tilde{p}|h}{\rho\tilde{U}^2}. \quad (7)$$

Another input parameter, namely, the direction of the pressure gradient force relative to the horizontal velocity, $\cos\theta$, is already nondimensional. Last, the output parameters $\tilde{\tau}$ and $\tilde{Q}$ are nondimensionalized to form as the local drag coefficient (C_D) and Nusselt number based on the heat transfer to the height of h (Nu_h):

$$C_D = \frac{\tilde{\tau}}{\rho\tilde{U}^2} \quad (8)$$

and

$$\text{Nu}_h = \frac{\tilde{Q}h}{k\Delta\tilde{T}}, \quad (9)$$

where $k = \alpha\rho C_p$ is the thermal conductivity and $C_p = 1004 \text{ J kg}^{-1} \text{ K}^{-1}$ is the specific heat of air.

c. The Monin–Obukhov similarity model

MOST predicts that in a horizontally homogeneous and quasi-steady state turbulent boundary layer, the near-surface velocity and temperature vary with height as

$$\overline{U}(z) = \frac{u_*}{\kappa} \left[\ln\left(\frac{z}{z_0}\right) - \Psi_M(\zeta) \right], \quad (10)$$

$$\theta_0 - \bar{\theta}(z) = \frac{\theta_*}{\kappa} \left[\ln\left(\frac{z}{z_T}\right) - \Psi_H(\zeta) \right], \quad (11)$$

where U represents horizontal velocity; θ is potential temperature (which has negligible difference from temperature, T , in this work); θ_0 is the potential temperature of the wall; u_* and θ_* are the friction velocity and scaling temperature; $\kappa = 0.4$ is the von Kármán constant; z_0 and z_T , known as the roughness lengths for velocity and temperature, are the heights where $\overline{U}(z_0)$ and $\theta_0 - \bar{\theta}(z_T)$ reach zero; $\zeta = z/L$ is the stability parameter, where L is the Obukhov length; Ψ_M and Ψ_H are integral stability functions for velocity and potential temperature. For the details of Obukhov length, stability parameter, and stability functions, the readers are referred to Garratt (1994, chapter 3). The overbar indicates Reynolds-average quantities, which are different from the filtered quantities in LES where the energy-containing eddies are preserved. However, in an equilibrium approach, the filtered velocity and temperature in LES are often used as the Reynolds-average inputs for MOST to derive the local surface shear stress and heat flux:

$$\tilde{\tau} = \rho u_*^2, \quad (12)$$

$$\tilde{Q} = -\rho C_p u_* \theta_*. \quad (13)$$

In the a priori test, h , $|\tilde{U}|$, and $\Delta\tilde{T}$ ($\approx \theta_0 - \bar{\theta}$) derived in section 2(b) are used as inputs for Eqs. (10) and (11). The MOST-predicted

TABLE 2. Introduction of the FNN cases.

Case	Input	Output
FNN_basic	$h, \bar{U} , \Delta \bar{T}$	$\bar{\tau}, \bar{Q}$
FNN_+w	$h, \bar{U} , \Delta \bar{T}, \bar{w}$	$\bar{\tau}, \bar{Q}$
FNN_+w+ ∇p	$h, \bar{U} , \Delta \bar{T}, \bar{w}, \rho^{-1} \nabla \bar{p} , \cos \theta$	$\bar{\tau}, \bar{Q}$
PAFNN_basic	$\ln(\text{Re}_{hU}), \ln(\text{Ra}_h)$	$\ln(C_D), \ln(\text{Nu}_h)$
PAFNN_+w	$\ln(\text{Re}_{hU}), \ln(\text{Ra}_h), \text{Fr}_w$	$\ln(C_D), \ln(\text{Nu}_h)$
PAFNN_+w+ ∇p	$\ln(\text{Re}_{hU}), \ln(\text{Ra}_h), \text{Fr}_w, \ln(\text{Eu}_h), \cos \theta$	$\ln(C_D), \ln(\text{Nu}_h)$

$\bar{\tau}$ and $\bar{Q}$ from Eqs. (12) and (13) are then compared with the $\bar{\tau}$ and $\bar{Q}$ from DNS. To avoid the deviation resulting from the model parameters for different flows, z_0 and z_T in Eqs. (10) and (11), which vary with flow and surface types (Garratt 1994), are tuned so that the mean values of τ and Q given by MOST match those from DNS as $T_{\text{bot}} - T_{\text{top}} = 12$ K. This calibration suggests that $z_0 = 0.98$ mm and $z_T = 0.53z_0$.

d. The neural network models

The training of the NN models is performed using Keras (Chollet et al. 2015) and TensorFlow (Abadi et al. 2015) in Python. The input layer consists of standardized and filtered variables from DNS data and a bias. The output layer contains the standardized values of $\bar{\tau}$ and $\bar{Q}$ for FNN, or C_D and Nu_h for PAFNN. The NN models comprise two hidden layers, containing 12 and 6 neurons, respectively. We determine this configuration through trial and error, aiming to achieve relatively good results with the fewest layers and neurons possible, thereby minimizing the risk of overfitting. Specifically, we start with small numbers of hidden layers and neurons, gradually increasing the hidden layers and neurons until no further reduction in the loss is observed. The resulting total of 18 neurons is on a scale similar to the 8, 16, and 30 total neurons used in Yang et al. (2019). We test the configurations on cases with the most input features in both FNN and PAFNN. For simplicity and given that hyperparameter optimization is beyond the scope of this research, we adopt the same configurations for all NN models, as we do not observe significant changes in results with the different configurations tested for other NN models. The neurons in a layer are weighted and fed into the leaky rectified linear unit (leaky ReLU) function to activate those in the next layer. The optimization of the training process is achieved using the Adam optimizer and mean squared error as the loss function. Among the 4991222 randomly chosen events, 60% are used as the training set, and the remaining 40% are used as the validation set, with a maximum of 5000 epochs and a batch size of 10000. To prevent overfitting, we utilize an early stopping technique with a patience of 20 epochs. This approach terminates the training process when the validation set loss fails to decrease for 20 consecutive epochs, even if the training set loss continues to decrease. After the twentieth epoch, which serves as a warm-up period, the early stopping process begins. The best weights are then restored based on the validation set loss. This ensures that the model generalizes well on unseen data by halting training at the optimal point. Without the early stopping criteria, we could still restore the best weights based

on the validation set. However, saving all weights for restoration would be wasteful in terms of storage, and it would be inefficient in terms of computational resources to finish the specified epochs when the validation set loss is already increasing. The testing set contains randomly chosen 2139094 events different from the validation and training sets. Given the inherent randomness in the training process, each model is trained several times to select the optimal outcome. Specifically, the optimal outcome is determined by achieving the highest R^2 value without evident anomalous pattern [examined subjectively from two-dimensional histogram like panels (a) and (b) in Figs. 3–5 and 7–9] which tend to occur when the number of input features is small. Further details of the training process, including the reduction of loss over epochs and data distribution, are provided in appendix B.

Several LES wall models are trained for comparison purposes, as summarized in Table 2. The first model (FNN_basic) is trained using the basic inputs as those for the MOST model to examine whether the traditional wall model could be improved with the same inputs. The next case (FNN_+w) is trained with the addition of $\bar{w}$ to the input layer to assess the importance of vertical velocity in the wall model. The third case (FNN_+w+ ∇p) is trained to further investigate the influence of pressure gradient force.

To construct PAFNN models, three additional cases are developed using nondimensional variables, as detailed in section 2(b). The variables are represented on a logarithmic scale, except for Fr_w and $\cos \theta$, which can take negative values. The use of logarithmic scales is based on previous research that suggested power-law relationships between the surface heat flux (scaled as Nu_h) and the surface shear stress (scaled as C_D) with the Reynolds number and Rayleigh number (Grossmann and Lohse 2000, 2002, 2011; Stevens et al. 2011, 2013), at least for sufficiently large Reynolds numbers and Rayleigh numbers. Invoking the logarithmic scaling simplifies the regression task in the asymptotic regime. In addition, to extrapolate outside the training conditions, the behavior of the target variable at infinity must conform to the behavior of the activation function at infinity, as per the extrapolation theorem in Bin et al. (2022). By employing the leaky ReLU activation function, the above is only true if we invoke the logarithmic scale. In section 3e, we will evaluate PAFNN performance on extrapolated input parameters.

e. The Shapley additive explanations value

In addition to comparing different NN cases, the importance of each feature in the NN models is further evaluated

by the SHAP values (Shapley 1953). We refer interested readers to Štrumbelj and Kononenko (2014) for the detailed equations. The primary concept of the SHAP value is to quantify how a variation in an input feature contributes to a variation in the output. A SHAP value is assigned to each feature for a particular event, and the distribution of SHAP values provides a measure of the overall importance and correlation (positive or negative) of each feature on a specific output.

The SHAP values and related figures in this work are produced mainly by the SHAP package for Python (Lundberg and Lee 2017). Although the SHAP values are standardized when directly computed from the input and output for NN, those presented in this work are scaled back by the output standard deviation to better quantify their contribution.

f. A posteriori test

To perform a posteriori tests, we implement the trained wall models in SAM LES, which uses the same solver and configuration as SAM DNS with a few modifications. The grid spacing is coarsened to 3.125 cm, and we use an anelastic approximation (as originally adopted in SAM LES for cloud chambers; Thomas et al. 2019, 2023; Wang et al. 2024), a turbulent-kinetic-energy (TKE) subgrid-scale (SGS) model (Deardorff 1980), and either a MOST or NN wall model. The upper boundary adopts the same wall model as the lower surface but with reverse values of $\Delta\tilde{T}$ and $\tilde{w}$. LESs are carried out for the same two bottom-top temperature differences as DNS: the $\Delta T = 12$ K simulations evaluate the performance of the NN models in the same type of flow that is used for the training, and the $\Delta T = 24$ K runs explore the performance of the NN models in the extrapolated regime when applied outside of the training parameter space.

While the evaluations of LES of some flows are based on systematic behavior at certain locations (e.g., the flow over hills; Zhou et al. 2021), this is not the case for RBC in this work. The locations of updraft and downdraft plumes shift with time, making it challenging to compare the LES and DNS results at specific locations. Instead, we compare the results based on the mean and distribution of local $\tilde{\tau}$ and $\tilde{Q}$. To evaluate LES results, $\tilde{\tau}$ and $\tilde{Q}$ in DNS are also filtered with a filter length of 3.125 cm before calculating the distribution. To compare the surface fluxes and near-surface quantities during the quasi-steady states (Figs. A1a,b), the results during $t = 5$ –10 min for the $\Delta T = 12$ K runs and $t = 1$ –4 min for the $\Delta T = 24$ K runs are analyzed.

3. Results

The snapshot of the DNS run reveals a turbulent convective layer with updraft and downdraft plumes (Fig. 1a). The concentration of turbulent vortex tubes is higher along these plumes. Near the lower surface, the turbulent coherent structure is characterized by sparse vortex tubes below the downdraft plume and abundant vortex tubes being drawn into the updraft plume. This structure differs from regularly organized hairpin vortices in a horizontally homogeneous boundary

layer, implying that MOST's assumption of horizontal statistical homogeneity is not satisfied in this case.

The results of the a priori tests in sections 3a to 3c demonstrate the inadequacy of MOST in modeling local surface shear stress and heat flux, the better results obtained by the NN models, and the gradual improvement of the NN models with the inclusion of additional input features. Later in sections 3d and 3e, we present the a posteriori tests and evaluate the ability of the PAFNN models to extrapolate.

a. A priori test: The Monin–Obukhov similarity model

Figure 2 presents a comparison between $\tilde{\tau}$ and $\tilde{Q}$ filtered from DNS and those modeled by MOST. Regarding $\tilde{\tau}$ (Fig. 2a), the main peak occurs at $\tilde{\tau} = 0$ to 0.2 mPa for DNS, whereas for MOST, it is almost always ~ 0 mPa. This discrepancy can be attributed to $|\tilde{U}|$, which can approach zero and result in a negligible $\tilde{\tau}$ according to MOST, while the turbulent motion within the filter grid cell may still produce a noticeable $\tilde{\tau}$. This phenomenon is an example of how RBC differs from a shear-dominated boundary layer flow. When $|\tilde{U}|$ increases, the modeled $\tilde{\tau}$ grows and gradually fits the filtered DNS values. For $\tilde{Q}$ (Fig. 2b), a distribution trend similar to that of $\tilde{\tau}$ is observed (i.e., growing from the bottom and then turning right in the two-dimensional histogram), except that the main peak now lies in the range where MOST overestimates $\tilde{Q}$. The overestimated peak results from the strong $|\tilde{U}|$ owing to Eq. (13) which assumes that the horizontal velocity increases the near-surface mixing, leading to a higher value of $\tilde{Q}$. Additionally, another local peak located at $\tilde{Q} < 5 \text{ W m}^{-2}$ as modeled by MOST can reach $> 30 \text{ W m}^{-2}$ as filtered from DNS, indicating that some grid cells can reach relatively high heat flux even with very low $|\tilde{U}|$. This again suggests the limitation of MOST to represent the near-surface turbulence in a non-shear-dominated flow.

Low values of the coefficient of determination for $\tilde{\tau}$ ($R^2 = 0.34$) and $\tilde{Q}$ ($R^2 < 0.1$) clearly illustrate the inadequacy of MOST in accurately modeling these variables. This inadequacy can be attributed to two reasons: 1) MOST cannot represent the near-surface turbulence in this flow, which is a well-known MOST limitation (see section 1), and 2) the three input features, h , $|\tilde{U}|$, and $\Delta\tilde{T}$, are insufficient to capture the details in the filter grid cell. The first reason can be assessed by comparing the results of case FNN_basic to those of MOST, while the second reason can be explored by comparing the results of cases FNN_+w and FNN_+w+ ∇p with FNN_basic.

b. A priori test: Feedforward neural network trained with primitive flow quantities

Figures 3a and 3b present the results for FNN_basic. For a better comparison, the color-bar scales remain the same for all the similar two-dimensional histograms (Figs. 2–9). Compared to the MOST results shown in Fig. 2, the R^2 values for $\tilde{\tau}$ and, especially, for $\tilde{Q}$ are significantly improved. The influence of individual input variables can be understood from the corresponding SHAP values. Figures 3c–f show that, for both $\tilde{\tau}$ and $\tilde{Q}$, the most important feature is $\Delta\tilde{T}$, followed by $|\tilde{U}|$ and h . $\Delta\tilde{T}$ is positively correlated with $\tilde{\tau}$ and $\tilde{Q}$, which is

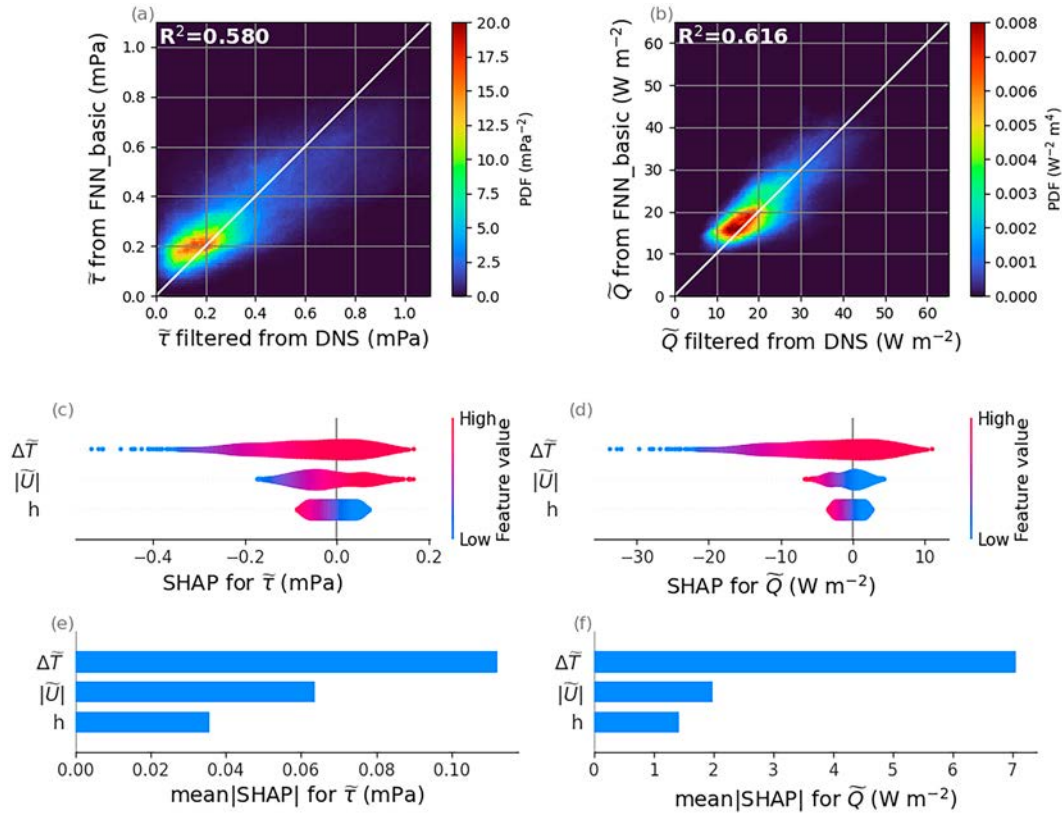


FIG. 3. (a),(b) As in Fig. 2, but for FNN_basic instead of MOST. (c),(d) The distribution of SHAP values, and (e),(f) their average magnitudes (overall importance) for (e) $\tilde{\tau}$ and (f) $\tilde{Q}$. In (c) and (d), the color indicates each feature's relative value (high or low); the width of each feature's bar represents the distribution of the data, and the x axis denotes the feature's contribution to the deviation of the target from the target's mean value.

consistent with MOST. $|\tilde{U}|$, on the other hand, is positively correlated with $\tilde{\tau}$ but negatively correlated with $\tilde{Q}$, so MOST-modeled $\tilde{Q}$ is biased with $|\tilde{U}|$ (Fig. 2b). Last, h is negatively correlated with $\tilde{\tau}$ and $\tilde{Q}$, consistent with MOST. Although the R^2 value yielded by FNN_basic is better than MOST, the two-dimensional histograms reveal its limitation in modeling the lower ranges of the distributions ($\tilde{\tau} \approx 0.1$ mPa and $\tilde{Q} < 10$ W m $^{-2}$). This limitation may be due to the inability of current features to represent the turbulence or the inability of the filtered feature to represent the details of weak turbulence.

With the addition of $\tilde{w}$ to the inputs, FNN_+w can noticeably improve R^2 compared to FNN_+basic (13.4% and 9.7% increase for $\tilde{\tau}$ and $\tilde{Q}$, respectively), although the improvements in the lower range of $\tilde{\tau}$ and $\tilde{Q}$ are minor (cf. Figs. 4a,b to Figs. 3a,b). The SHAP values reveal that $\tilde{w}$ is the most important feature for $\tilde{\tau}$ and the second most important feature for $\tilde{Q}$. Additionally, a downdraft (low feature values of $\tilde{w}$ in Figs. 4c,d) increases both $\tilde{\tau}$ and $\tilde{Q}$. Traditionally, vertical velocity near the surface is not taken as an input of the wall model because of the assumption of horizontal homogeneity (i.e., the horizontal convergence/divergence is negligible). The importance of $\tilde{w}$ again proves the limitation of applying the wall model derived for shear-dominated flows to a convection-dominated flow. Because $\tilde{w}$ and $\Delta\tilde{T}$ are not independent, the

relative importance of $\Delta\tilde{T}$ is reduced compared to FNN_basic (cf. Fig. 4e to Fig. 3e). The negative correlation between $\tilde{w}$ and $\Delta\tilde{T}$ (as indicated by the inverse correlation of $\tilde{w}$ and $\Delta\tilde{T}$ with $\tilde{\tau}$ or $\tilde{Q}$, as shown in Figs. 4c,d) suggests that $\tilde{w}$ is the primary mechanism that influences $\Delta\tilde{T}$ (otherwise, an increase in $\Delta\tilde{T}$ should increase buoyancy and thus strengthen $\tilde{w}$). Consequently, the influence of $\tilde{w}$ cannot be incorporated through $\Delta\tilde{T}$ into the inputs.

Figure 5 shows that the results can be further improved considering the influence of the pressure gradient via two input variables, $\rho^{-1}|\nabla\tilde{p}|$ and $\cos\theta$, in the FNN_+w+ $\nabla\tilde{p}$ case (Table 2). Specifically, R^2 for $\tilde{\tau}$ increases by $\sim 18.7\%$ (Fig. 5a). At the same time, R^2 for $\tilde{Q}$ changes very little. Consistent with these R^2 changes, SHAP value analysis shows the relatively high importance of $\rho^{-1}|\nabla\tilde{p}|$ and $\cos\theta$ for $\tilde{\tau}$ and their negligible importance for $\tilde{Q}$ (Figs. 5c-f). The SHAP values of $\rho^{-1}|\nabla\tilde{p}|$ show its positive correlation with both $\tilde{\tau}$ and $\tilde{Q}$. Although the mean SHAP value of $\rho^{-1}|\nabla\tilde{p}|$ is not the highest for $\tilde{\tau}$, several instances with exceedingly high SHAP values of $\rho^{-1}|\nabla\tilde{p}|$ reveal the contribution of $\rho^{-1}|\nabla\tilde{p}|$ to the extreme range of $\tilde{\tau}$ (Fig. 5c). The SHAP values of $\cos\theta$ show that the acceleration of an air parcel leads to higher $\tilde{\tau}$ and lower $\tilde{Q}$, but the influence on $\tilde{Q}$ is almost negligible (Figs. 5d,f).

To better understand how each input variable affects $\tilde{\tau}$ and $\tilde{Q}$, we examine the vertical profiles of $|\tilde{U}|$ and T within

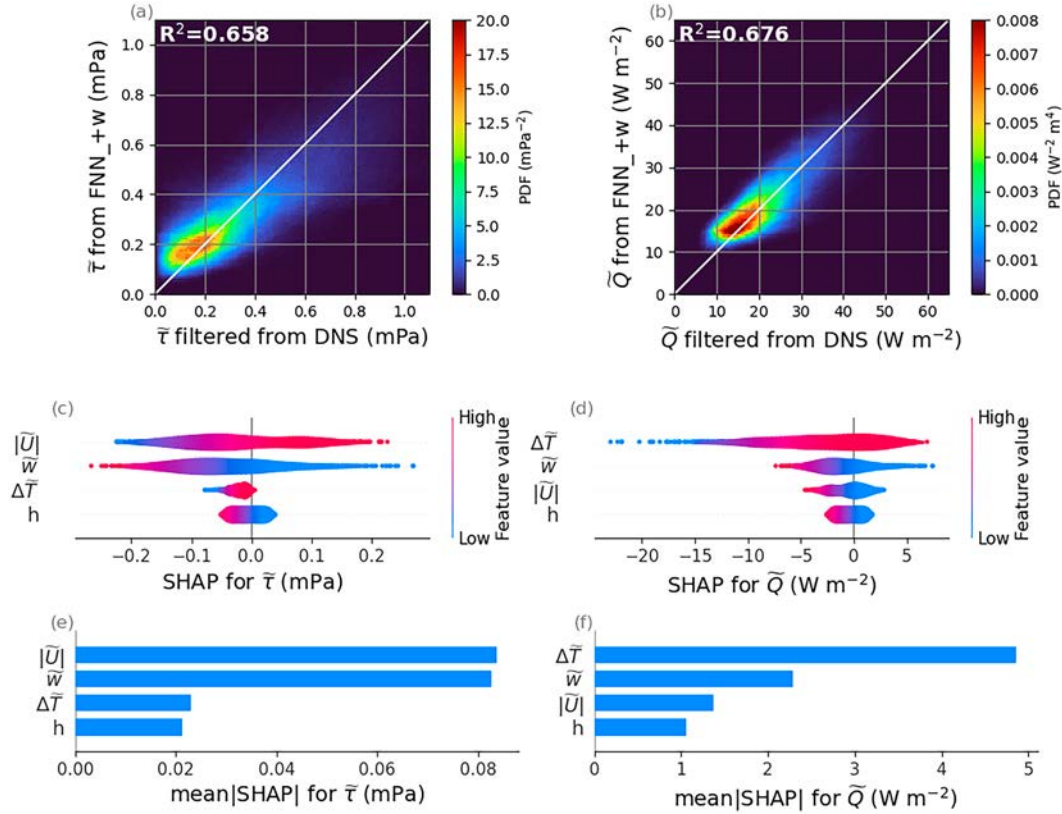


FIG. 4. As in Fig. 3, but for the case FNN+ w.

selected filter grid cells (Fig. 6). The angle brackets in the figure and hereafter indicate the temporal and horizontal mean. The values of $\tilde{\tau}$ and $\tilde{Q}$ are determined by $\partial|U|/\partial z$ and $-\partial T/\partial z$ immediately above the surface (denoted as $\partial|U|_s/\partial z$ and $-\partial T_s/\partial z$, respectively). We choose grid cells based on high (above the 90th percentile) or low (below the 10th percentile) feature values, and then apply the conditional average. We use a filter length of 3.125 cm (i.e., the influence of h is disregarded) to simplify the analysis and better visualize $\partial|U|_s/\partial z$ and $-\partial T_s/\partial z$.

Figure 6 reveals the deviation of the mean $|U|$ and T profiles simulated by DNS from those suggested by MOST (comparing the red solid line to the blue line in each panel). Figure 6a shows that high $|\tilde{U}|$ implies enhanced $\partial|U|_s/\partial z$, which leads to enhanced $\tilde{\tau}$. For the T profile, although $\Delta\tilde{T}$ increases with $|\tilde{U}|$, $-\partial T_s/\partial z$ remains nearly unchanged (Fig. 6f). In other words, with given $\Delta\tilde{T}$, higher $|\tilde{U}|$ implies a lower magnitude of $-\partial T_s/\partial z$ and thus lower $\tilde{Q}$, leading to the negative correlation between $|\tilde{U}|$ and $\tilde{Q}$ (Figs. 3d, 4d, and 5d). Regarding the influence of $\Delta\tilde{T}$, high $\Delta\tilde{T}$ implies sharpened $\partial|U|_s/\partial z$ and $-\partial T_s/\partial z$ (Figs. 6b,g), indicating enhanced $\tilde{\tau}$ and $\tilde{Q}$. The influence of $\tilde{w}$ is illustrated by Figs. 6c and 6h, which show that downdrafts bring high $|U|$ and low T values closer to the surface, leading to higher $\partial|U|_s/\partial z$ and $-\partial T_s/\partial z$. This effect of near-surface w is consistent with previous studies (Brown and Thomas 1977; Raupach 1981; Adrian et al. 2000; Ganapathisubramani et al. 2003; Salesky and Anderson 2018), and leads to enhanced $\tilde{\tau}$ and $\tilde{Q}$. On the contrary, an updraft (ejection) has the

opposite influence. Additionally, because downdrafts result in a high value of $\Delta\tilde{T}$, the low $\tilde{w}$ condition and the high $\Delta\tilde{T}$ condition have qualitatively similar $|U|$ and T profiles near the surface (cf. Figs. 6b,g with Figs. 6c,h). Last, high $\rho^{-1}|\nabla\tilde{p}|$ and an accelerating forcing (i.e., high $\cos\theta$) both increase $\partial|U|_s/\partial z$, while $-\partial T_s/\partial z$ is barely affected (Figs. 6d,e,i,j), consistent with the corresponding SHAP values (Fig. 5).

c. *A priori test: Physics-aware feedforward neural network*

The performance of the NN can be improved by incorporating information about the physical system, e.g., by forming a reduced set of input variables using dimensional analysis (Gunaratnam et al. 2003). Nondimensionalizing FNN_basic inputs using Eqs. (4) and (5) reduces the number of input features from three to two in the PAFNN_basic case (2). However, as shown in Figs. 7a and 7b, the R^2 of PAFNN_basic is inferior to FNN_basic with much narrower ranges of modeled $\tilde{\tau}$ and $\tilde{Q}$ and smaller R^2 values. Specifically, $\tilde{\tau}$ in DNS ranges from 0 to more than 1 mPa, but the values modeled by PAFNN_basic fall almost entirely between 0.1 and 0.6 mPa. Similarly, $\tilde{Q}$ in DNS ranges from approximately 5 to more than 50 W m^{-2} , but the values modeled by PAFNN_basic only have a range of approximately 10–30 W m^{-2} . Note that the PAFNN models are trained using $\ln(C_D)$ and $\ln(\text{Nu}_h)$, but the results are converted back to $\tilde{\tau}$ and $\tilde{Q}$ for a fair

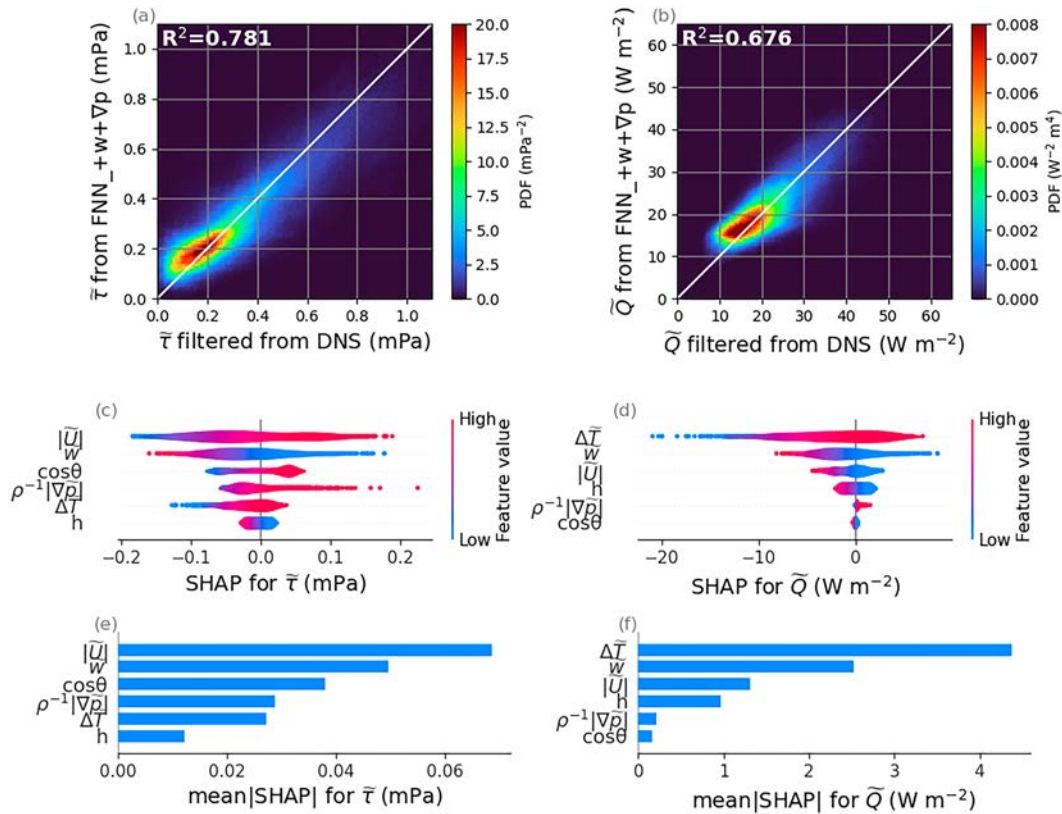


FIG. 5. As in Fig. 3, but for the case FNN₊+w+∇p.

comparison with the previous results and for the subsequent implementation in LES. The least errors measured in the training process in the $\ln(C_D)$ and $\ln(\text{Nu}_h)$ spaces are not necessarily the cases in the $\bar{\tau}$ and $\bar{Q}$ spaces, so the conversion may partly contribute to the poorer performance of the PAFNN models. Ranking the importance of input features using the SHAP values reveals that Re_{hU} is the more influential parameter for C_D and Ra_h is more important for Nu_h (Figs. 7c–f). Low Re_{hU} and high Ra_h contribute to both high C_D and Nu_h .

After considering the influence of vertical velocity, PAFNN₊+w significantly improves the results (cf. Fig. 8 to Fig. 7). Specifically, R^2 for $\bar{\tau}$ and $\bar{Q}$ are improved by 99% and 30%, respectively. These R^2 values are comparable to those obtained by FNN₊+w, even though PAFNN₊+w is trained with one less feature and the target values in the logarithmic nondimensional parameter space. The SHAP values suggest that Fr_w is the second most important feature for both C_D and Nu_h , again demonstrating the importance of accounting for vertical velocity near the surface. However, it should be noted that the comparison of SHAP values for Fr_w and $\cos\theta$ with other variables on a logarithmic scale is less clear, in contrast to the primitive flow quantities depicted in Figs. 4 and 5 where all the features are in a linear scale. Last, Fr_w exhibits a negative correlation with both C_D and Nu_h , supporting the observation made in section 3b.

Accounting for the influence of pressure gradient force further improves the R^2 values for both $\bar{\tau}$ and $\bar{Q}$ (cf. Figs. 9a,b to

Figs. 8a,b). The improvement in R^2 values from PAFNN₊+w to PAFNN₊+w+∇p is more significant for $\bar{\tau}$ (~19.4%) than for $\bar{Q}$ (~0.8%), similar to the improvements of FNN₊+w+∇p over FNN₊+w. Regarding the importance of features to C_D , $\cos\theta$ and Eu_h rank second and fourth. Both features are positively correlated with C_D , similar to the influence of $\cos\theta$ and $\rho^{-1}|\nabla p|$ on $\bar{\tau}$ (as revealed in Fig. 5c). Overall, based on the distribution of SHAP values, only Re_{hU} is significantly more important to C_D than other features (Figs. 9c,e). Regarding the importance of the features in Nu_h , the influence of Ra_h and Fr_w is significantly greater than the other features, again highlighting the importance of vertical velocity. The importance of Eu_h and $\cos\theta$ is the lowest, and the importance of Re_{hU} is in the middle, similar to the relationships among the primitive flow quantities and $\bar{Q}$, as revealed in Fig. 5c.

In summary, when comparing the PAFNN with the FNN models, only PAFNN_{basic} shows noticeably poorer performance than FNN_{basic}. To ensure a fairer comparison between the FNN and PAFNN models, their results are evaluated against the number of features (Fig. 10, which is also known as the elbow method; Thorndike 1953). We see that $1 - R^2$ is reduced significantly as the number of input features increases from 2 to 3, and then declines slowly as the number of features increases further. This suggests that including more than three input features increases the risk of overfitting, especially for $\bar{Q}$.

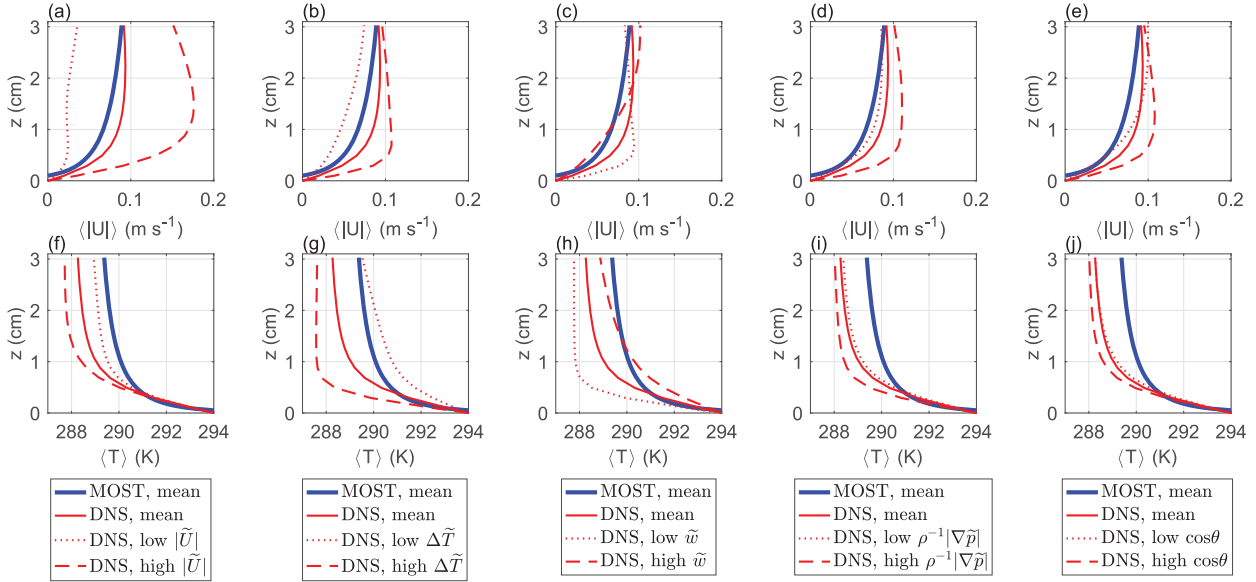


FIG. 6. Vertical profiles of (a)–(e) $\langle |U| \rangle$ and (f)–(j) $\langle T \rangle$ from DNS (red lines), including the mean (solid line) and conditionally averaged profiles with low (i.e., values below the 10th percentile; dotted line) and high (i.e., values above the 90th percentile; dashed line) values of five FNN input variables: (a),(f) $|\tilde{U}|$, (b),(g) $-\Delta\tilde{T}$, (c),(h) $\tilde{w}$, (d),(i) $\rho^{-1}|\nabla\tilde{p}|$, and (e),(j) $\cos\theta$. The MOST profiles of $|U|$ and T that yield the mean $\tilde{\tau}$ and $\tilde{Q}$ in DNS are plotted for comparison (blue lines). Note that the MOST profiles of $|U|$ and T , calculated for given u_* and θ_* using Eqs. (10) and (11), are defined only for $z \geq z_0$ and $z \geq z_T$, respectively.

d. A posteriori baseline tests

In this section, we subject the NN models to a posteriori tests by implementing these models in SAM, performing LES, and comparing the simulated distributions of $\tilde{\tau}$ and $\tilde{Q}$ among these runs, as well as with the original LES (applying the MOST-based model) and DNS simulations. Figures 11a and 11c show that the distribution of $\tilde{\tau}$ modeled by MOST (green lines) has a wider range than that obtained from DNS (black lines). Specifically, the MOST model overestimates the probability of $\tilde{\tau}$ values near 0 mPa, while capturing the higher values slightly better, consistent with the a priori test described earlier (Fig. 2a).

The distribution of $\tilde{\tau}$ modeled by FNN_basic (Figs. 11a,c, solid blue lines) has an overpredicted and unsmooth peak near $\tilde{\tau} \sim 0.35$ mPa with the underestimated probability of $\tilde{\tau}$ smaller than 0.1 mPa, resulting in an overestimated $\langle \tilde{\tau} \rangle$. The unsmooth distribution suggests limitations in the NN models with insufficient input features, and such unsmoothness always appears somewhere in the probability density functions (PDF) after retraining the FNN_basic model, even after testing several different combinations of neurons and hidden layers (not shown). Additionally, the lower ranges of $\tilde{\tau}$ and $\tilde{Q}$ are barely captured, which has been revealed in the a priori study [see panels (a) and (b) in Figs. 3–5 and 7–9]. With the additional input of vertical velocity (FNN_+w), the mean value of $\tilde{\tau}$ is underestimated, the peak position approaches that of the DNS (Fig. 11a, dashed blue line), and the distribution is smoother, though the small value tail deviates further from DNS (Fig. 11c). Adding the ∇p input (FNN_+w+ ∇p) matches the DNS peak well and improves the small value

tails, although the mean value is further underestimated (Figs. 11a,c, dotted blue lines).

For the physics-aware models, PAFNN_basic overestimates the probability for $0.1 < \tilde{\tau} < 0.4$ mPa and underestimates the probability outside of this range (Figs. 11a,c, solid red lines), consistent with the narrow range observed in the a priori study (Fig. 7a). The underprediction of large $\tilde{\tau}$ values also leads to a significantly underestimated mean (0.29 mPa) compared to the DNS (0.37 mPa). The unsmooth distribution around the peak reveals the same concern of insufficient input features as revealed by the FNN_basic model. Including the vertical velocity input (PAFNN_+w) improves both the distribution shape and the mean value, although near the peak some unsmoothness is still observable (Figs. 11a,c, red dashed lines). After accounting for the ∇p input (PAFNN_+w+ ∇p), the distribution slightly shifts to the left (Fig. 11c, dotted red line), leading to a slightly reduced mean value, but the distribution is the smoothest.

Regarding the surface heat flux, Figs. 11b and 11d show that the LES with the MOST model (green lines) predicts a negatively skewed distribution compared to a positively skewed PDF from DNS and overestimates the mean $\langle \tilde{Q} \rangle$. FNN_basic results in a nearly symmetric PDF (Fig. 11b, solid blue lines), with better $\langle \tilde{Q} \rangle$ than that obtained by MOST. Taking into account the vertical velocity (FNN_+w) improves the $\tilde{Q}$ distribution, with reduced probability near $\tilde{Q} \sim 22$ W m⁻² and a better mean value (Figs. 11b,d, dashed blue lines). When the pressure gradient force is taken into account (FNN_+w+ ∇p), the value of $\langle \tilde{Q} \rangle$ slightly decreases and further approaches the DNS value (Fig. 11b, blue dotted line). For the PAFNN models, with the basic inputs (PAFNN_basic), the peak location and

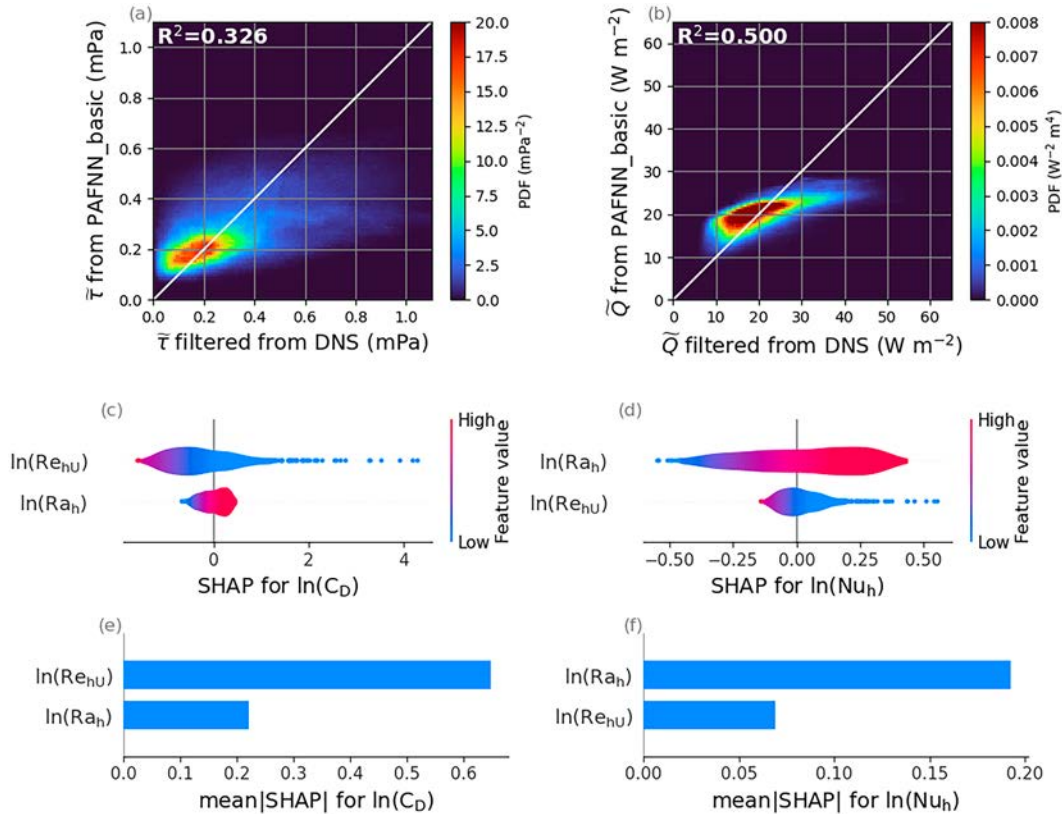


FIG. 7. As in Fig. 3, but for the case PAFNN_basic.

mean value of $\tilde{Q}$ are better than those obtained by MOST, but the distribution is relatively narrow, with very few $\tilde{Q}$ values above 30 W m^{-2} (Figs. 11b,d, solid red lines), as observed in the a priori tests in Fig. 7b. After taking into account the vertical velocity (PAFNN_+w), the resulting $\tilde{Q}$ has a much wider distribution, a peak closer to that of DNS, and a much improved mean value (Figs. 7b,d, red dashed lines). With the influence of pressure gradient force (PAFNN_+w+ ∇p), the right tail slightly deviates from that of DNS (Fig. 7d, red dotted line), and the $\langle \tilde{Q} \rangle$ value decreases while still well matching the DNS value.

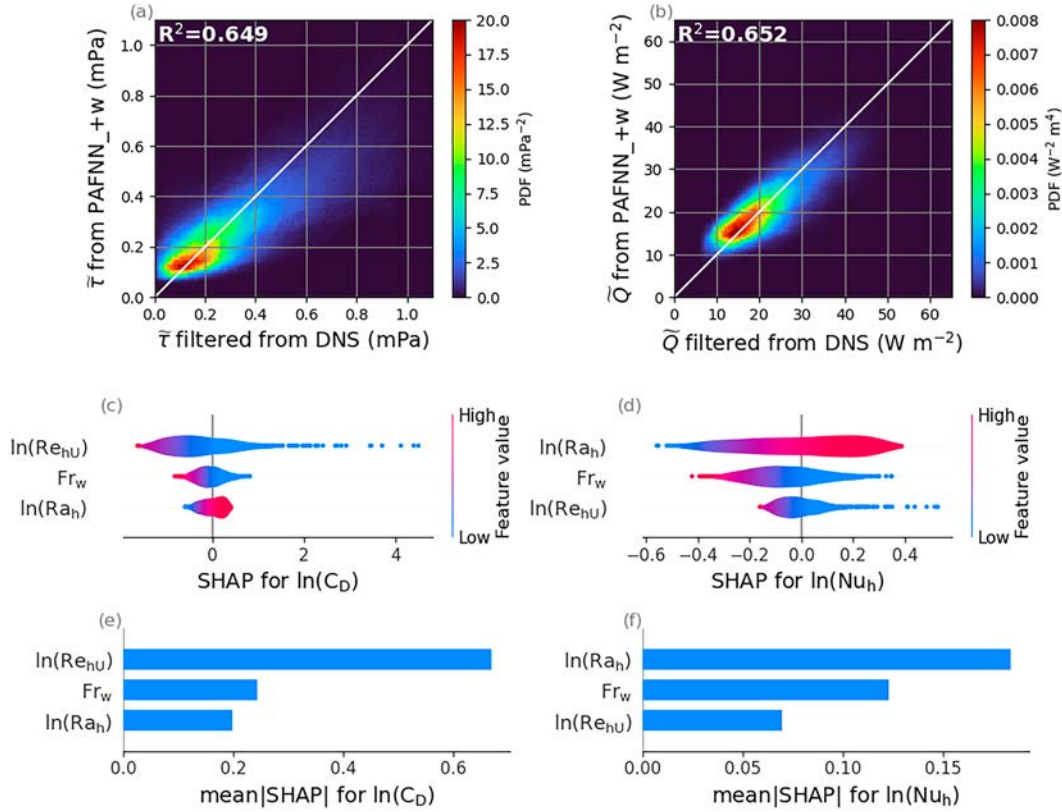
The performance of the wall models heavily depends on the quality of the input variables. During the NN training and in the a priori testing, these variables are taken from a reference DNS and are the same for all wall models, while in the a posteriori (interactive) testing the inputs, provided now by LES, are affected by the employed wall models. Figure 12 shows the primitive flow quantities on the first LES grid level that contribute to the input layer of the NN models. The distributions of $|\tilde{U}|$ in LES runs match that from DNS relatively well (Fig. 12a). However, all LES runs overemphasize the peak of the PDF and underpredict frequency of occurrence of $\Delta \tilde{T}$ below 4.5 K (Fig. 12b). This may worsen the limitation of the NN models in capturing the lower ranges of $\tilde{\tau}$ and $\tilde{Q}$, which are positively correlated with $\Delta \tilde{T}$. The overpredicted $\Delta \tilde{T}$ may result from overly active mixing in the interior of the domain in LES, possibly due to the overestimated dissipative

effect of the SGS model or numerical diffusion in the advection scheme (Pressel et al. 2017; Wang et al. 2021). Although the dissipative and diffusive effects of SGS models and advection schemes are beyond the scope of this study, they can also contribute to the PDF of $\tilde{w}$ being too narrow, (Fig. 12c), and the PDF of $\rho^{-1}|\nabla \tilde{p}|$ being shifted to smaller values relative to that in DNS (Fig. 12d). Because $\rho^{-1}|\nabla \tilde{p}|$ is underestimated in LES, including its influence reduces the mean value of the $\tilde{\tau}$ and $\tilde{Q}$ distributions (Fig. 11), regardless of whether the values are improved or not.

Overall, including the input from vertical velocity into the NN wall models improves the peak and smoothness of the considered PDFs. In terms of the mean value, FNN_+w yields the best result for $\tilde{\tau}$; PAFNN_+w and PAFNN_+w+ ∇p perform similarly well for $\tilde{Q}$. Since the analysis is based on a set of simulations of one specific case, and the performance of the NN models can vary if the models are trained again (because of the randomness inherent in the training process), it is fair to conclude that, for the training scenario ($\Delta T = 12 \text{ K}$), the NN wall models perform well after considering the vertical velocity.

e. A posteriori extrapolation tests

In the final set of tests in this study, we investigate the performance of the NN wall models when they operate on input variables outside the ranges for which they were trained. In

FIG. 8. As in Fig. 3, but for the case PAFNN₊+w.

these extrapolation tests, we conduct simulations of more unstable RBC by doubling the temperature difference between the bottom and top boundaries (from 12 to 24 K), while keeping the rest of the model setup unchanged.

First, we note that the MOST model cannot accurately capture either the mean or the PDF shape of $\langle \bar{\tau} \rangle$ and $\langle \bar{Q} \rangle$ (Fig. 13, green lines), suggesting that z_0 and z_T of the MOST model may need case-specific adjustments (Wang et al. 2024). The FNN models operating on primitive flow quantities perform even worse, overestimating $\langle \bar{\tau} \rangle$ and $\langle \bar{Q} \rangle$ by about a factor of 2 or more (Fig. 13, blue lines). In contrast, the PAFNN models demonstrate good extrapolation abilities (Fig. 13, red lines). PAFNN_{basic} still predicts relatively narrow ranges, and PAFNN₊+w and PAFNN₊+w+ ∇p perform similarly well in reproducing the distributions of $\bar{\tau}$ and $\bar{Q}$ from DNS.

Figure 14 shows the near-surface primitive flow quantities for examining the input features. Here, we focus only on the distribution obtained by the PAFNN models compared to that in the DNS (represented by the red and black lines in Fig. 14, respectively). Similar to the training case as shown in Fig. 12, $\Delta \bar{T}$ is overpredicted, $\bar{w}$ is closer to zero, and $\rho^{-1}|\nabla \bar{p}|$ is underpredicted in the LES runs compared to that obtained by the DNS run. The underpredicted $\rho^{-1}|\nabla \bar{p}|$ still reduces the resulting mean values of $\bar{\tau}$ and $\bar{Q}$ (Fig. 13).

In summary, although comparing the different LES runs with DNS is complicated by the difference in the flow

quantities in each run, we can conclude that PAFNN models possess the capability to extrapolate, at least within the RBC regime considered in this study, and including vertical velocity captures more details. Although including pressure gradient force does not improve the mean surface shear stress and heat flux, the smoothness of PDF is improved, and the local surface shear stress is reproduced more accurately according to the a priori tests. In a pressure-driven flow, including pressure gradient force may be crucial.

4. Conclusions

Many traditional LES wall models rely on MOST, which is derived for a shear-dominated flow and may not be well suited for RBC. One important flow quantity in RBC is the near-surface vertical velocity, which is not considered by MOST. To explore alternative wall models, we perform DNS to train the NN (including the FNN and PAFNN) models with various input features for a better modeling of surface shear stress and heat flux. To understand the influence of the input flow quantities, the SHAP values and the conditional average of the $|U|$ and T profiles are investigated. For the a priori study, we use the DNS-filtered data as inputs for the MOST and NN models, and then compare the models' outputs with the DNS-filtered data. For the a posteriori tests, we implement the NN models in LES and evaluate the distribution of surface shear stress and heat flux against the DNS

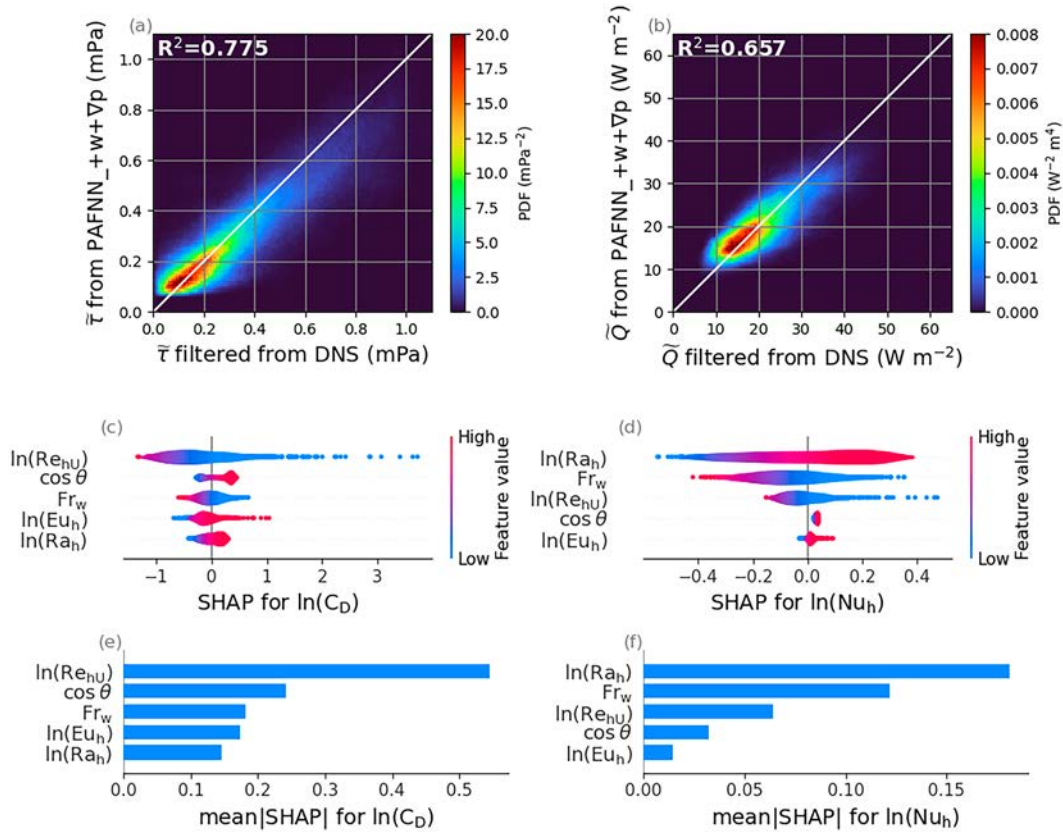


FIG. 9. As in Fig. 3, but for the case PAFNN_+w+ ∇p .

results. For the a posteriori extrapolation tests, additional DNS and LES runs are performed with a doubled temperature difference between the bottom and top boundaries.

A priori tests show that, in the same scenario (temperature difference between the bottom and top boundaries) as in the training process, the NN models are able to capture the local surface shear stress and heat flux. Incorporating vertical velocity into the NN models enhances both the modeled surface shear stress and heat flux. Meanwhile, the inclusion of the horizontal pressure gradient force yields a noticeable improvement in surface shear stress but only marginally

enhances the heat flux. An analysis of the input features shows that a strong horizontal velocity magnitude is positively correlated with surface shear stress (analogous to the MOST model) but is negatively correlated with heat flux (contrary to the MOST model). A pronounced vertical temperature difference is positively correlated with both surface shear stress and heat flux. Negative near-surface vertical velocities, or down-drafts, drive the high- $|U|$ and low- T air toward the surface. This enhances the vertical wind shear and temperature gradient near the surface, subsequently amplifying both surface shear stress and heat flux. A significant horizontal pressure

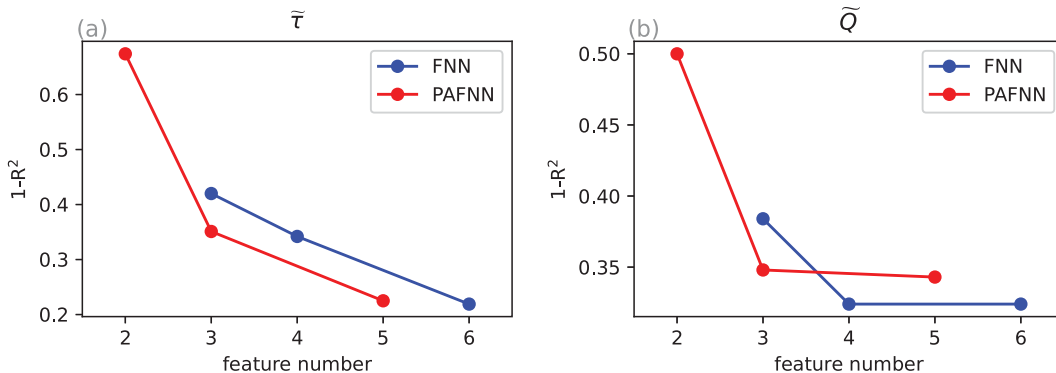


FIG. 10. The feature number vs. the unexplained variation (i.e., $1 - R^2$) for (a) $\bar{\tau}$ and (b) $\bar{Q}$.

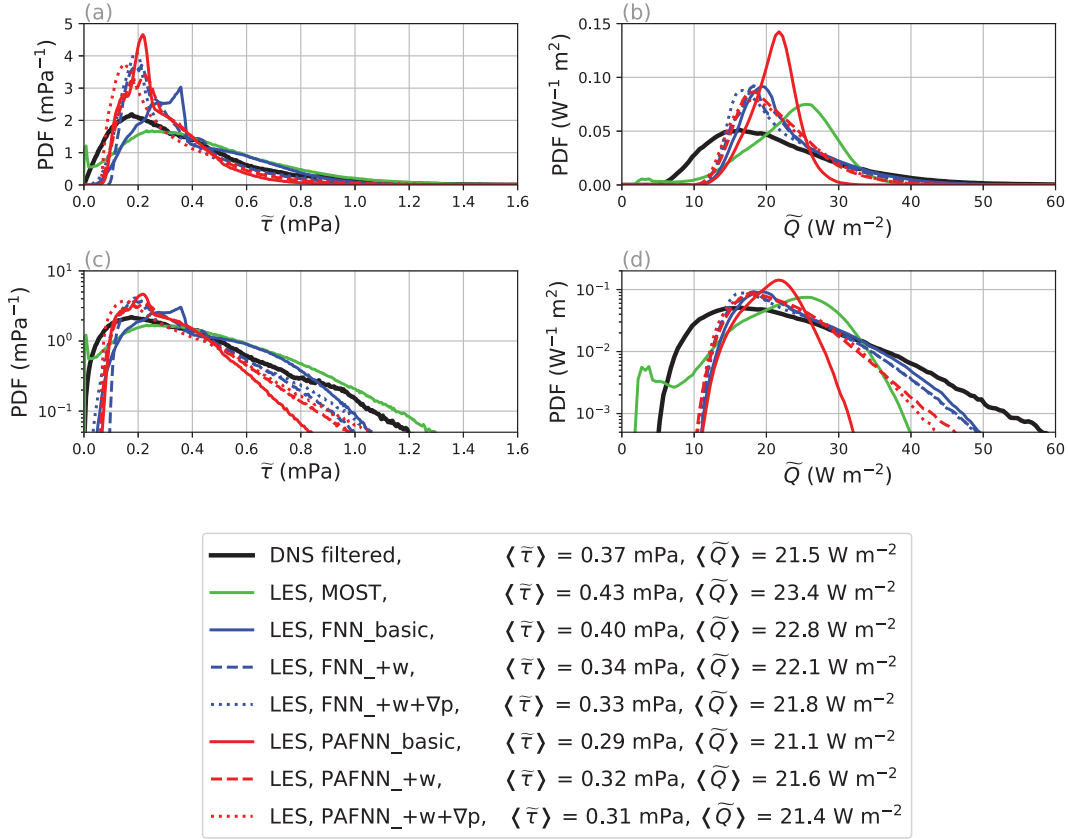
A posteriori tests, $T_{\text{bot}} - T_{\text{top}} = 12 \text{ K}$ 

FIG. 11. The distribution of (a),(c) $\tilde{\tau}$ and (b),(d) $\tilde{Q}$ in the LES runs with the MOST and newly implemented NN models compared with that filtered from DNS. (c) and (d) use a logarithmic y axis to emphasize the small probabilities in the tails. The mean values of $\tilde{\tau}$ and $\tilde{Q}$ are presented in the legend.

gradient force and its alignment with the horizontal velocity both augment the near-surface vertical gradient of $|U|$, thereby increasing surface shear stress. However, they exert only a minimal impact on T and, consequently, the heat flux.

When trained with nondimensional parameters as data and with more than three input features, the PAFNN models perform comparably well to the FNN models in the scenarios of the training process. Finally, all NN models exhibit shortcomings

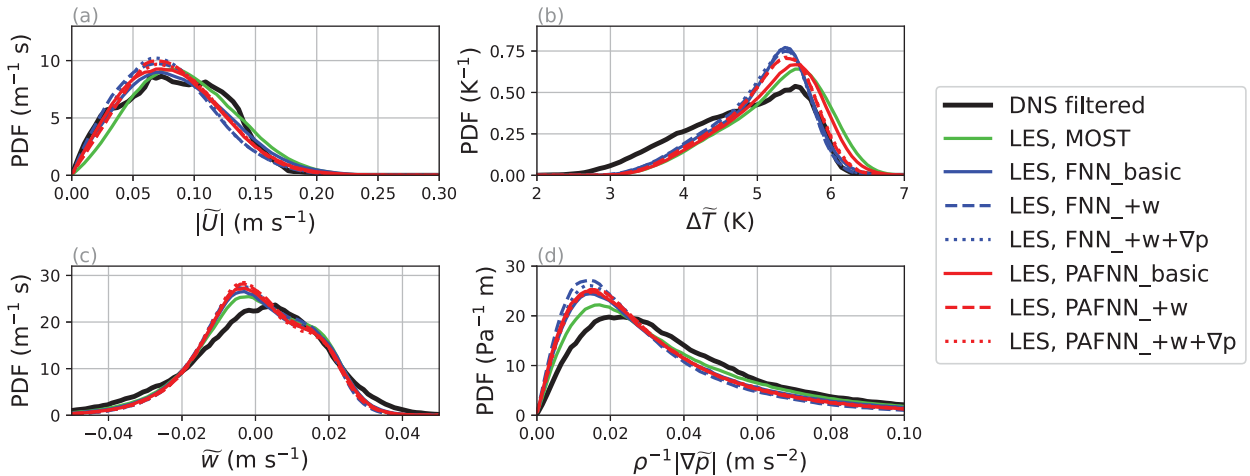
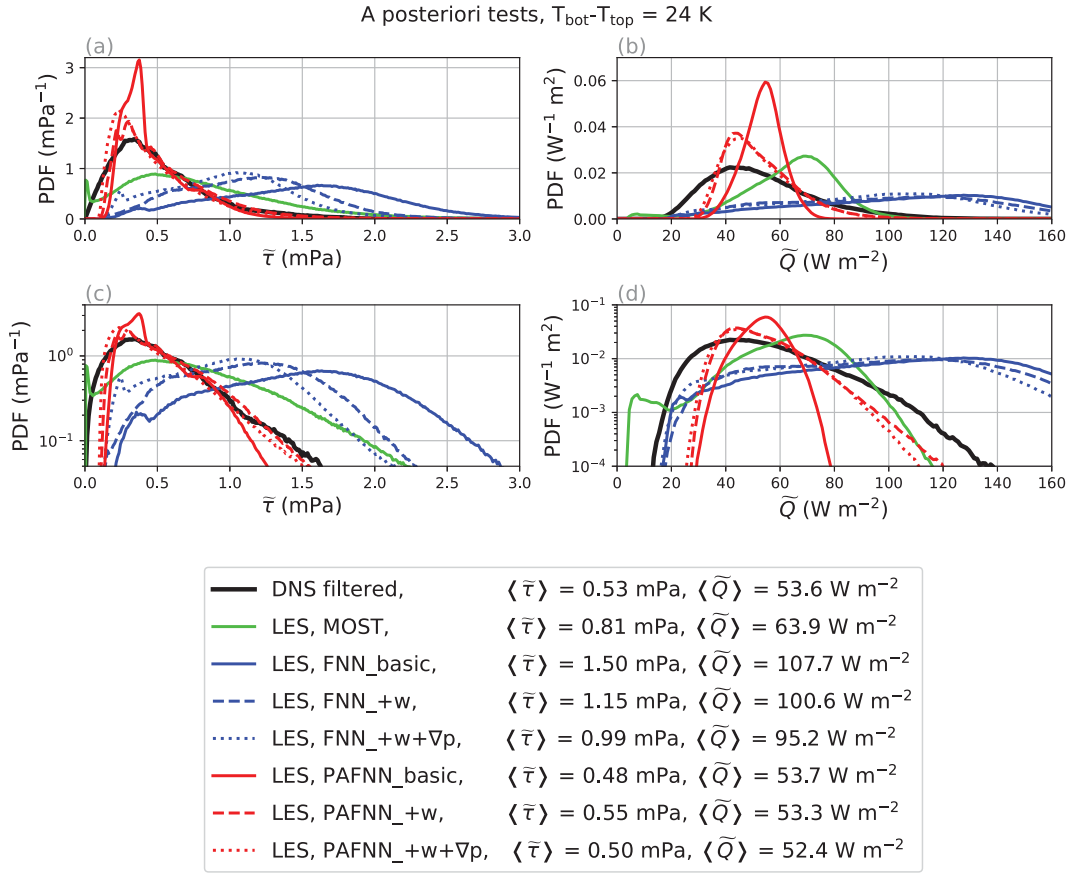
A posteriori tests, $T_{\text{bot}} - T_{\text{top}} = 12 \text{ K}$ 

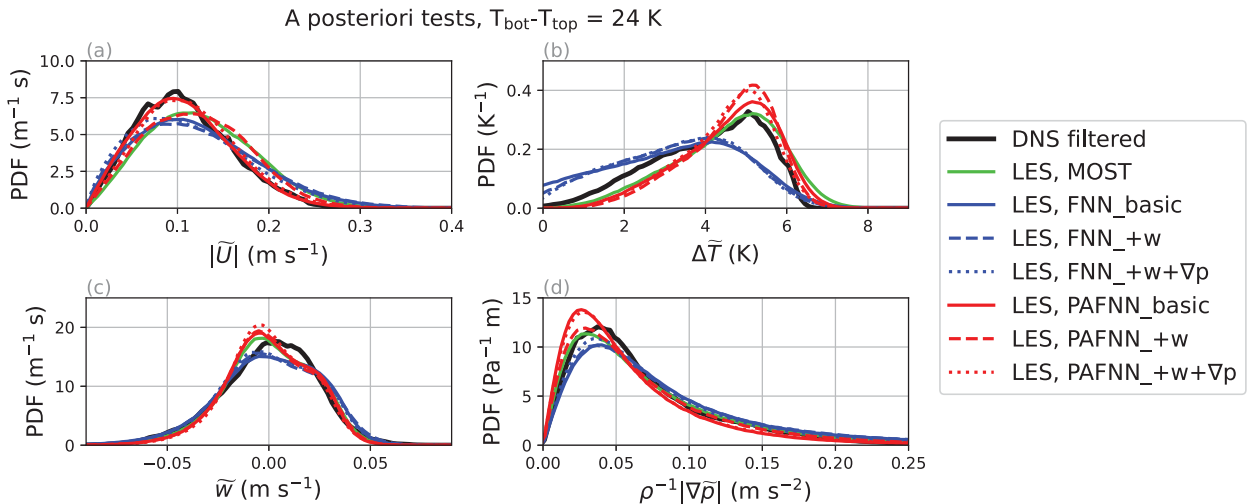
FIG. 12. The distributions of the first-grid-level primitive flow quantities that contribute to the inputs of the NN models obtained by each LES run and filtered from DNS.

FIG. 13. As in Fig. 11, but for the extrapolation runs ($\Delta T = 24 \text{ K}$).

when modeling the lower ranges of surface shear stress and heat flux.

In the a posteriori tests simulating the same RBC scenario used for the training, LES runs utilizing the newly implemented NN models can achieve reasonable mean values of

surface shear stress and heat flux. Nevertheless, when including only the same input as the MOST model, the NN models exhibit unsmooth distributions near the peaks, revealing the limitation owing to insufficient input features. The details in the distributions of surface shear stress and heat flux are

FIG. 14. As in Fig. 12, but for the extrapolation runs ($\Delta T = 24 \text{ K}$).

improved after including the influence of vertical velocity. Including the influence of pressure gradient force results in the smoothest distribution of PDF, although the mean surface shear stress and heat flux are underestimated. However, the input features provided in LES are different from DNS, possibly because of a greater dissipation from the SGS models and advection schemes. The quality of the input features can affect the performance of the wall models. For example, the limitations of the NN models in modeling the lower ranges can be worsened by the overestimated near-surface temperature difference in LES, and the lack of improvement after including pressure gradient force may result from the underestimated pressure gradient force in LES. Last, the a posteriori tests with a more intense RBC scenario show that the FNN models are incapable of extrapolating, whereas the PAFNN models demonstrate the ability to do so.

This work contributes to advancing the understanding of near-surface turbulence and applying ML techniques to the development of NN-based wall models. Our analysis of feature importance highlights the negative correlation of the near-surface vertical velocity with surface shear stress and heat flux. The vertical velocity's influence should exist and be considered in the wall model of any flow where vertical velocity is nonnegligible. Furthermore, although vertical velocity and near-surface temperature difference are not completely independent in this study, they can be independent in flows such as a tornado, where vertical velocity is driven by the strong perturbation pressure gradient force (e.g., Wang et al. 2020, 2023). In terms of application, the PAFNN models have successfully extrapolated to a more unstable flow, suggesting their potential for use in other atmospheric conditions.

From the presented results, several directions for follow-up research can be identified. To enhance the NN-based wall models, a more systematic exploration of hyperparameters for training is recommended, and testing different filter types for DNS data, such as the Gaussian filter, is worth considering. Additionally, it remains to be explored whether the means and standard deviations derived from the training data are suitable for various flows. Regarding limitations, the NN models do not currently account for surface roughness like the MOST-based model. Also, the current PAFNN models are not yet applicable to a stable boundary layer, because the Rayleigh number is restricted to positive values when used in logarithmic scale as an input. For application to atmospheric flows, the temperature difference may need to be parameterized for different hydrostatic regimes like MOST, and further experiments are required to confirm its reliability. Finally, for atmospheric applications, the proposed approach needs to be expanded to include the development and testing of NN representations for fluxes of moisture and other scalars.

Acknowledgments. This research is supported by the U.S. Department of Energy Office of Science Atmospheric System Research (ASR) project at PNNL. PNNL is operated for the Department of Energy by Battelle Memorial Institute under Contract DE-AC05-76 RL01830. This research used resources of the National Energy Research Scientific Computing

Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract DE-AC02-05CH11231 using NERSC Awards BER-ERCAP0021333 and BER-ERCAP0024023. Yang acknowledges AFOSR for financial support. Figure 1a is produced using Python PyVista package (Sullivan and Kaszynski 2019). Figure 1b is produced by the code modified from Python wrapper for the John Hopkins University Turbulence Database Cluster library (Li et al. 2008; Perlman et al. 2007). For transparency and following the position statement of the Committee on Publication Ethics (COPE), we acknowledge the assistance of ChatGPT and Bing Chat in examining the possible reasons for errors in the data processing code, generating the first draft of the data processing code, checking spelling, correcting grammar, along with some refinements for the context.

Data availability statement. The SAM model was kindly provided by Prof. Marat Khairoutdinov and publicly available at <http://rossby.msrc.sunysb.edu/~marat/SAM.html>. The output from the SAM simulations, including the DNS and LES runs, and the Python scripts for training NN are stored on NERSC HPSS storage system at <https://portal.nersc.gov/archive/home/w/wang1202/www/Wang2023JAS>. TensorFlow is an open source available at <https://www.tensorflow.org/>. Keras is also an open source available at <https://keras.io/>. The SHAP value package for Python is available at <https://github.com/slundberg/shap>.

APPENDIX A

Flow Statistics

As introduced in section 2a, the focus of this study is on near-surface turbulence. The domain aspect ratio with periodic lateral boundaries is not ideal for examining the interior structure of RBC. Keeping this in mind, the flow statistics of the DNS runs and LES runs with two different wall models (MOST and PAFNN_ $+w+\nabla p$) are provided for the readers' information.

Figure A1 displays the time series of both the DNS and LES runs. For the LES runs, only two representative wall models are chosen to maintain clarity in the figures. The surface sensible heat flux can represent the steadiness of near-surface turbulence, while the resolved TKE can represent the steadiness of the interior of the domain. In each case, the quasi-steady states are reached in approximately 1 min. The LES runs achieve the quasi-steady states slightly faster than the DNS runs, due to the assumption of complete mixing within the LES grid cells.

Figure A2 depicts horizontally averaged vertical profiles of flow statistics for the two a posteriori tests, with ΔT of 12 and 24 K. Figures A2a and A2b reveal that the temperature profile in the interior of the flow is slightly tilted, complicating the definition of boundary layer depth. The tilt in the temperature profile appears to be related to periodic lateral boundaries, because an additional LES run with no-penetration lateral boundaries shows no tilt (not shown). If

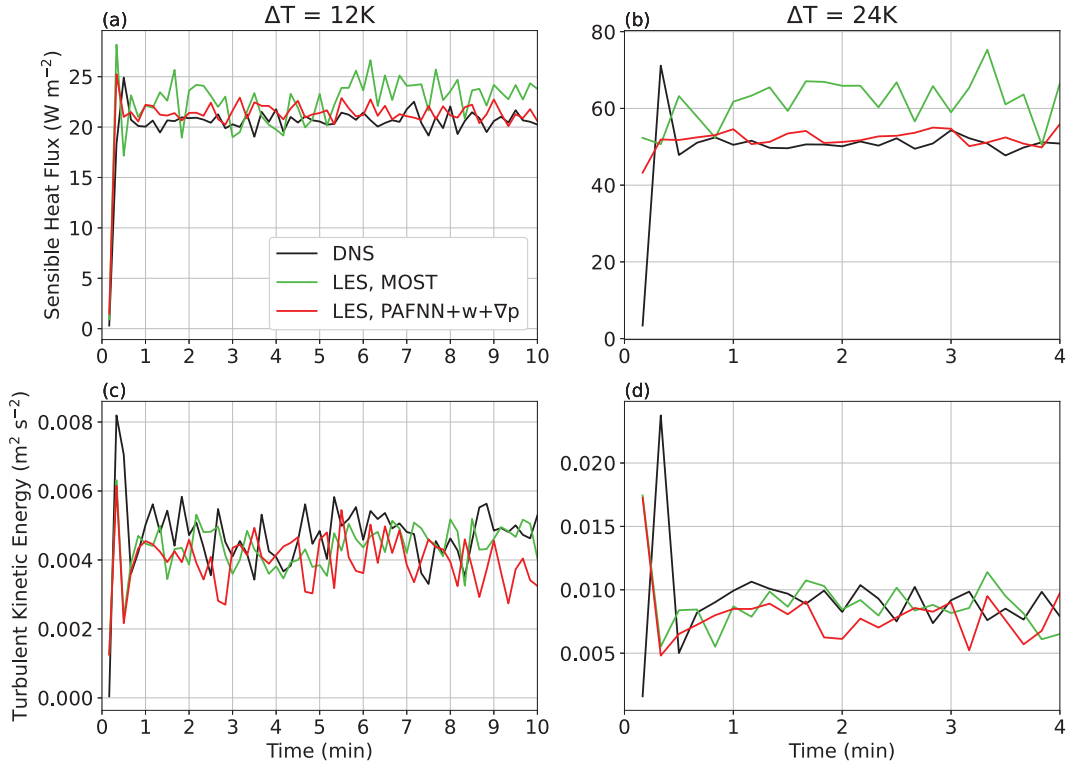


FIG. A1. Time series of (a),(b) domain-mean surface sensible heat flux and (c),(d) domain-mean resolved TKE in the (a),(c) $\Delta T = 12 \text{ K}$ and (b),(d) $\Delta T = 24 \text{ K}$ cases. The black lines represent the DNS runs, green lines represent the LES runs with the MOST wall model, and the red lines represent the LES runs with the PAFNN_+w wall model.

the boundary layer is examined by the temperature gradient, before the temperature gradient reaches 99% close to the centerline value, the LES runs consist of only one grid point in the boundary layer, while the DNS runs have 12 and 11 grid points for the $\Delta T = 12 \text{ K}$ and $\Delta T = 24 \text{ K}$ cases, respectively. Figures A2c and A2d present the temperature variance. DNS resolves a much higher temperature variance near the surface. In the LES, the MOST model produces a slightly larger temperature variance than the PAFNN model. Figures A2e and A2f illustrate that the DNS runs contain

more than 10 grid points below the near-surface peaks of TKE, whereas in the LES runs the local maxima of TKE are at levels closest to the walls. It is important to note that the resolved TKE in LES should be less than the total TKE represented by DNS. Therefore, LES with the MOST wall model that yields resolved TKE similar to, or greater than, that of DNS is indicative of an overpredicted sensible heat flux (i.e., excessive energy input). This is further confirmed in Figs. A2g and A2h, where the resolved vertical velocity variance in LES with the MOST wall model exceeds that predicted by DNS.

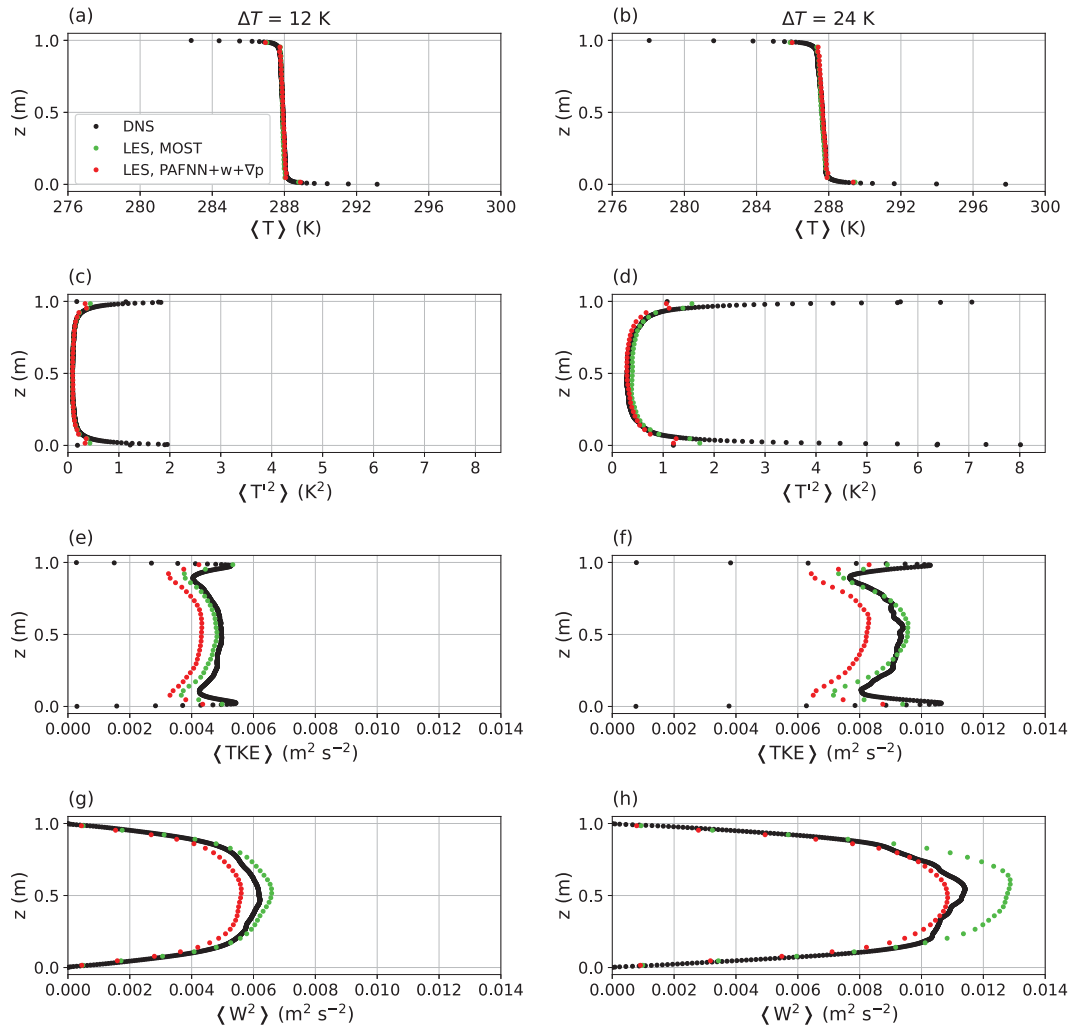


FIG. A2. Temporally (after $t = 2$ min) and horizontally averaged vertical profiles of (a),(b) temperature, (c),(d) temperature variance, (e),(f) resolved TKE, and (g),(h) resolved vertical velocity variance in the (a),(c),(e),(g) $\Delta T = 12$ K and (b),(d),(f),(h) $\Delta T = 24$ K cases. The black markers represent the DNS runs, green markers represent the LES runs with the MOST wall model, and the red markers represent the LES runs with the PAFNN₊w wall model. Note that in (a) to (c), many red markers overlap with the green ones.

APPENDIX B

Training Details

Figure B1 illustrates the reduction in loss over epochs. The loss is determined by the mean squared error (MSE) for each batch, which is defined as

$$\text{MSE}_{\text{batch}} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (\text{B1})$$

where n is the batch size, y_i is the true target value, and $\hat{y}_i$ is the prediction. Within an epoch, all batches are processed, so the MSE at each epoch in Fig. B1 represents the

weighted average of $\text{MSE}_{\text{batch}}$, with the batch sizes acting as weights. The last batch may have a different size if the entire dataset size is not completely divisible by the batch size.

Using the early stopping criterion, the training for each case halts when the loss stops decreasing for 20 consecutive epochs. As the number of features increases, the loss can reach a lower minimum, and the required number of epochs increases.

Figures B2 and B3 display the distributions of features versus targets used in the training of both the FNN and PAFNN models. Generally, features that are more correlated with targets result in higher feature importance, as seen in Figs. 3–5 and 7–9.

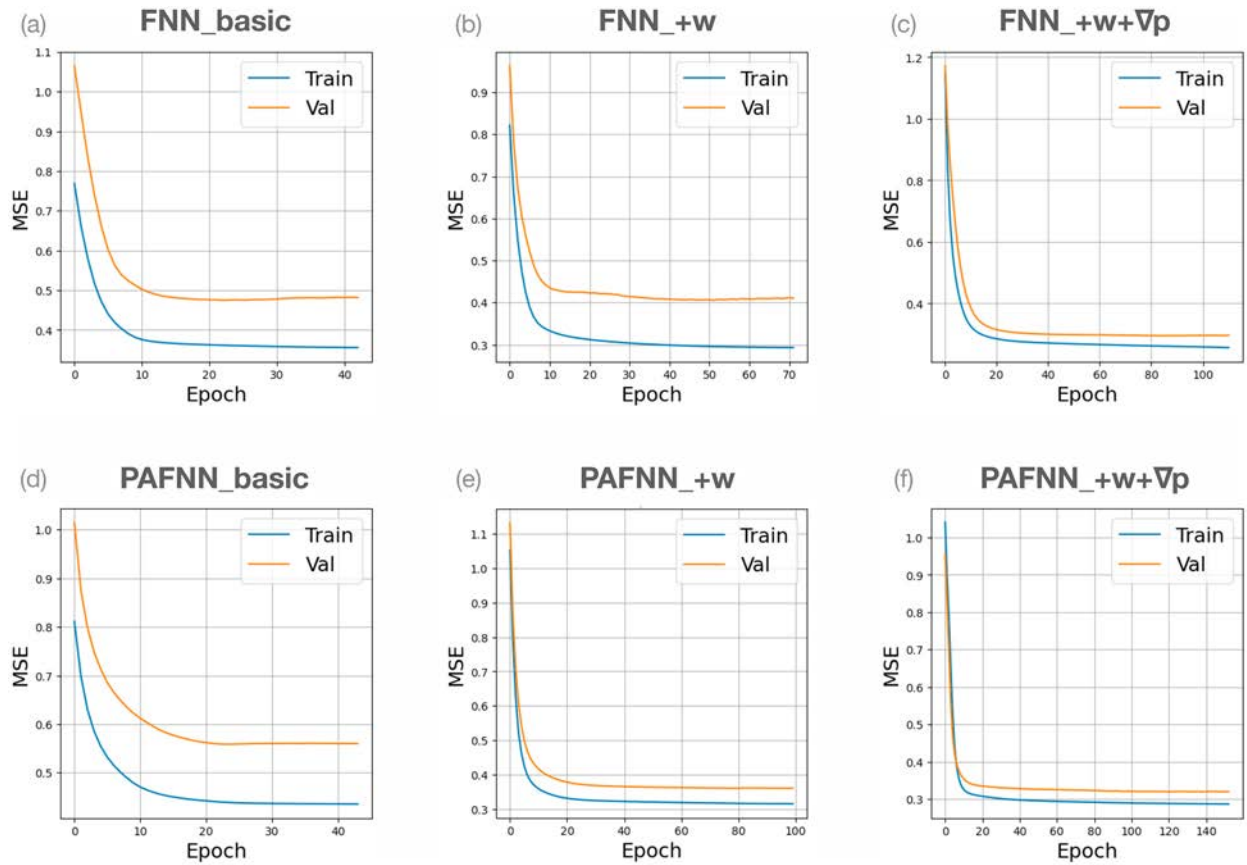


FIG. B1. The decrease in loss (MSE) across epochs for each case. Blue lines represent the loss for the training data, while orange lines indicate the loss for the validation data.

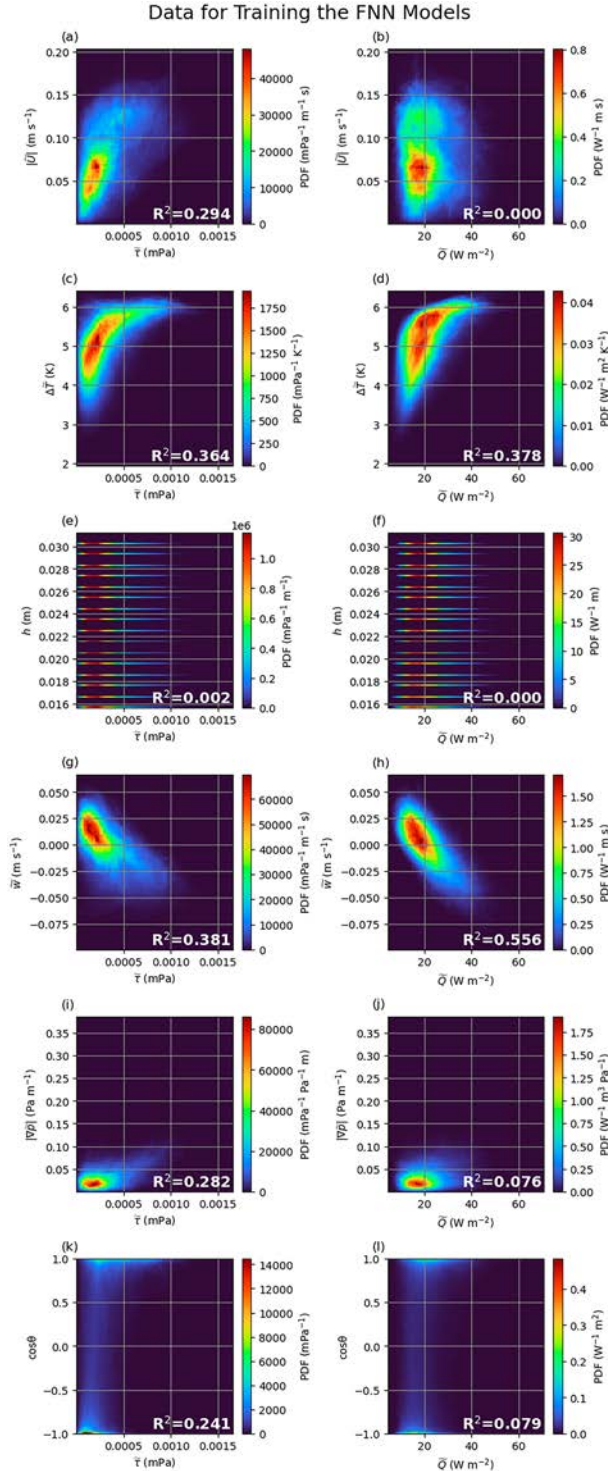


FIG. B2. The two-dimensional histogram showing the data distribution of features vs targets for the FNN models.

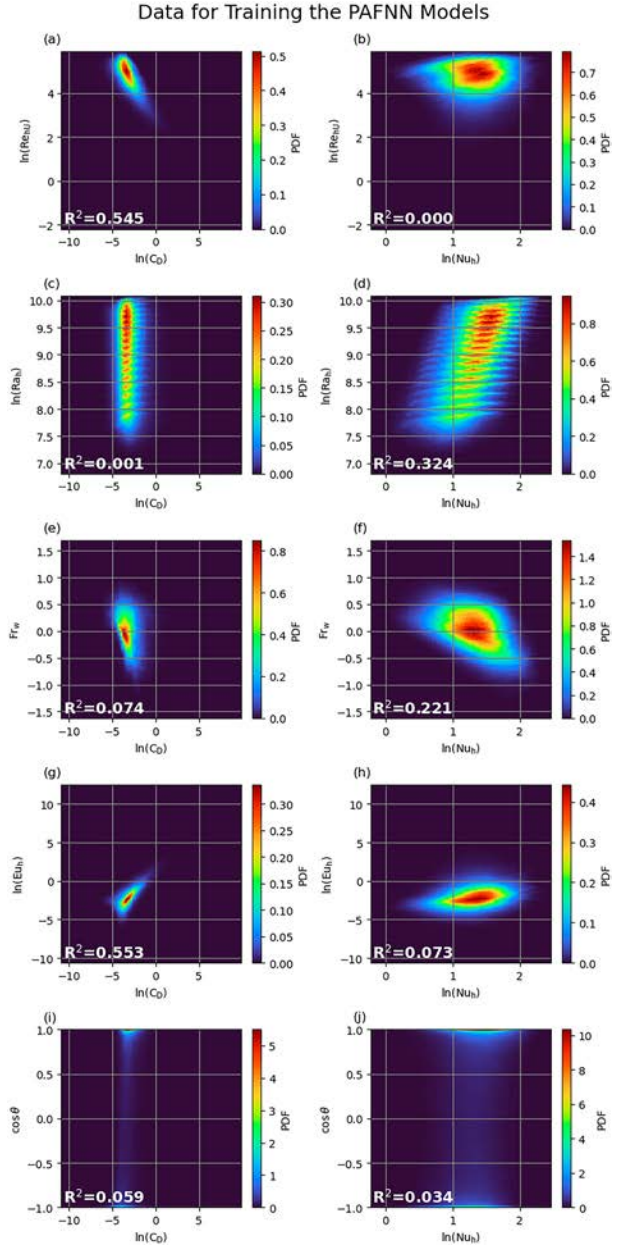


FIG. B3. As in Fig. B2, but for the PAFNN models.

REFERENCES

- Abadi, M., and Coauthors, 2015: TensorFlow: Large-scale machine learning on heterogeneous systems. TensorFlow, <https://www.tensorflow.org/>.
- Adrian, R. J., C. D. Meinhart, and C. D. Tomkins, 2000: Vortex organization in the outer region of the turbulent boundary layer. *J. Fluid Mech.*, **422**, 1–54, <https://doi.org/10.1017/S0022112000001580>.
- Ahlers, G., S. Grossmann, and D. Lohse, 2009: Heat transfer and large scale dynamics in turbulent Rayleigh-Bénard convection. *Rev. Mod. Phys.*, **81**, 503–537, <https://doi.org/10.1103/RevModPhys.81.503>.

- Bae, H. J., and P. Koumoutsakos, 2022: Scientific multi-agent reinforcement learning for wall-models of turbulent flows. *Nat. Commun.*, **13**, 1443, <https://doi.org/10.1038/s41467-022-28957-7>.
- Balaras, E., C. Benocci, and U. Piomelli, 1996: Two-layer approximate boundary conditions for large-eddy simulations. *AIAA J.*, **34**, 1111–1119, <https://doi.org/10.2514/3.13200>.
- Bin, Y., L. Chen, G. Huang, and X. I. A. Yang, 2022: Progressive, extrapolative machine learning for near-wall turbulence modeling. *Phys. Rev. Fluids*, **7**, 084610, <https://doi.org/10.1103/PhysRevFluids.7.084610>.
- Bose, S. T., and G. I. Park, 2018: Wall-modeled large-eddy simulation for complex turbulent flows. *Annu. Rev. Fluid Mech.*, **50**, 535–561, <https://doi.org/10.1146/annurev-fluid-122316-045241>.
- Bou-Zeid, E., C. Meneveau, and M. Parlange, 2005: A scale-dependent Lagrangian dynamic model for large eddy simulation of complex turbulent flows. *Phys. Fluids*, **17**, 025105, <https://doi.org/10.1063/1.1839152>.
- Brown, G. L., and A. S. W. Thomas, 1977: Large structure in a turbulent boundary layer. *Phys. Fluids*, **20** (10), S243–S252, <https://doi.org/10.1063/1.861737>.
- Cabot, W., and P. Moin, 2000: Approximate wall boundary conditions in the large-eddy simulation of high Reynolds number flow. *Flow Turbul. Combust.*, **63**, 269–291, <https://doi.org/10.1023/A:1009958917113>.
- Cai, S., Z. Mao, Z. Wang, M. Yin, and G. E. Karniadakis, 2021: Physics-informed neural networks (PINNs) for fluid mechanics: A review. *Acta Mech. Sin.*, **37**, 1727–1738, <https://doi.org/10.1007/s10409-021-01148-1>.
- Chandrakar, K. K., W. Cantrell, K. Chang, D. Ciochetto, D. Niedermeier, M. Ovchinnikov, R. A. Shaw, and F. Yang, 2016: Aerosol indirect effect from turbulence-induced broadening of cloud-droplet size distributions. *Proc. Natl. Acad. Sci. USA*, **113**, 14243–14248, <https://doi.org/10.1073/pnas.1612686113>.
- , H. Morrison, W. W. Grabowski, G. H. Bryan, and R. A. Shaw, 2022: Supersaturation variability from scalar mixing: Evaluation of a new subgrid-scale model using direct numerical simulations of turbulent Rayleigh–Bénard convection. *J. Atmos. Sci.*, **79**, 1191–1210, <https://doi.org/10.1175/JAS-D-21-0250.1>.
- Chang, K., and Coauthors, 2016: A laboratory facility to study gas–aerosol–cloud interactions in a turbulent environment: The II Chamber. *Bull. Amer. Meteor. Soc.*, **97**, 2343–2358, <https://doi.org/10.1175/BAMS-D-15-00203.1>.
- Chavanne, X., F. Chilla, B. Chabaud, B. Castaing, and B. Hebral, 2001: Turbulent Rayleigh–Bénard convection in gaseous and liquid He. *Phys. Fluids*, **13**, 1300–1320, <https://doi.org/10.1063/1.1355683>.
- Chollet, F., and Coauthors, 2015: Keras. GitHub, <https://github.com/fchollet/keras>.
- Deardorff, J. W., 1980: Stratocumulus-capped mixed layers derived from a three-dimensional model. *Bound.-Layer Meteor.*, **18**, 495–527, <https://doi.org/10.1007/BF00119502>.
- Di Mari, R., S. Ingrassia, and A. Punzo, 2023: Local and overall deviance r -squared measures for mixtures of generalized linear models. *J. Classif.*, **40**, 233–266, <https://doi.org/10.1007/s00357-023-09432-4>.
- Fodor, K., J. P. Mellado, and M. Wilczek, 2019: On the role of large-scale updrafts and downdrafts in deviations from Monin–Obukhov similarity theory in free convection. *Bound.-Layer Meteor.*, **172**, 371–396, <https://doi.org/10.1007/s10546-019-00454-3>.
- Ganapathisubramani, B., E. K. Longmire, and I. Marusic, 2003: Characteristics of vortex packets in turbulent boundary layers. *J. Fluid Mech.*, **478**, 35–46, <https://doi.org/10.1017/S002212002003270>.
- Garratt, J. R., 1994: *The Atmospheric Boundary Layer*. Cambridge University Press, 316 pp.
- Grossmann, S., and D. Lohse, 2000: Scaling in thermal convection: A unifying theory. *J. Fluid Mech.*, **407**, 27–56, <https://doi.org/10.1017/S002212099007545>.
- , and —, 2002: Prandtl and Rayleigh number dependence of the Reynolds number in turbulent thermal convection. *Phys. Rev. E*, **66**, 016305, <https://doi.org/10.1103/PhysRevE.66.016305>.
- , and —, 2011: Multiple scaling in the ultimate regime of thermal convection. *Phys. Fluids*, **23**, 045108, <https://doi.org/10.1063/1.3582362>.
- Grötzbach, G., 1983: Spatial resolution requirements for direct numerical simulation of the Rayleigh–Bénard convection. *J. Comput. Phys.*, **49**, 241–264, [https://doi.org/10.1016/0021-9991\(83\)90125-0](https://doi.org/10.1016/0021-9991(83)90125-0).
- Gunaratnam, D. J., T. Degroff, and J. S. Gero, 2003: Improving neural network models of physical systems through dimensional analysis. *Appl. Soft Comput.*, **2**, 283–296, [https://doi.org/10.1016/S1568-4946\(02\)00061-3](https://doi.org/10.1016/S1568-4946(02)00061-3).
- Huang, X. L. D., and X. I. A. Yang, 2021: A Bayesian approach to the mean flow in a channel with small but arbitrarily directional system rotation. *Phys. Fluids*, **33**, 015103, <https://doi.org/10.1063/5.0035552>.
- , —, and R. F. Kunz, 2019: Wall-modeled large-eddy simulations of spanwise rotating turbulent channels—Comparing a physics-based approach and a data-based approach. *Phys. Fluids*, **31**, 125105, <https://doi.org/10.1063/1.5129178>.
- Jeong, J., and F. Hussain, 1995: On the identification of a vortex. *J. Fluid Mech.*, **285**, 69–94, <https://doi.org/10.1017/S002212095000462>.
- Khairoutdinov, M. F., and D. A. Randall, 2003: Cloud resolving modeling of the ARM summer 1997 IOP: Model formulation, results, uncertainties, and sensitivities. *J. Atmos. Sci.*, **60**, 607–625, [https://doi.org/10.1175/1520-0469\(2003\)060<0607:CRMOTA>2.0.CO;2](https://doi.org/10.1175/1520-0469(2003)060<0607:CRMOTA>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**, 301–305.
- Li, Y., and Coauthors, 2008: A public turbulence database cluster and applications to study Lagrangian evolution of velocity increments in turbulence. *J. Turbul.*, **9**, N31, <https://doi.org/10.1080/14685240802376389>.
- Lozano-Durán, A., and H. J. Bae, 2020: Self-critical machine-learning wall-modeled LES for external aerodynamics. *arXiv*, 2012.10005v2, <https://doi.org/10.48550/arXiv.2012.10005>.
- , and —, 2023: Machine learning building-block-flow wall model for large-eddy simulation. *J. Fluid Mech.*, **963**, A35, <https://doi.org/10.1017/jfm.2023.331>.
- Lundberg, S. M., and S.-I. Lee, 2017: A unified approach to interpreting model predictions. *Proc. 31st Int. Conf. on Neural Information Processing Systems*, Long Beach, CA, NIPS, 4768–4777, <https://dl.acm.org/doi/10.5555/3295222.3295230>.
- Mahrt, L., 1999: Stratified atmospheric boundary layers. *Bound.-Layer Meteor.*, **90**, 375–396, <https://doi.org/10.1023/A:1001765727956>.
- Markowski, P. M., and G. H. Bryan, 2016: LES of laminar flow in the PBL: A potential problem for convective storm simulations. *Mon. Wea. Rev.*, **144**, 1841–1850, <https://doi.org/10.1175/MWR-D-15-0439.1>.

- Milano, M., and P. Koumoutsakos, 2002: Neural network modeling for near wall turbulent flow. *J. Comput. Phys.*, **182**, 1–26, <https://doi.org/10.1006/jcph.2002.7146>.
- Moeng, C.-H., 1984: A large-eddy-simulation model for the study of planetary boundary-layer turbulence. *J. Atmos. Sci.*, **41**, 2052–2062, [https://doi.org/10.1175/1520-0469\(1984\)041<2052:ALESMF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1984)041<2052:ALESMF>2.0.CO;2).
- Monin, A. S., and A. M. Obukhov, 1954: Basic laws of turbulent mixing in the atmosphere near the ground. *Tr. Geofiz. Inst., Akad. Nauk SSSR*, **24**, 163–187.
- Niemela, J. J., L. Skrbek, K. R. Sreenivasan, and R. J. Donnelly, 2000: Turbulent convection at very high Rayleigh numbers. *Nature*, **404**, 837–840, <https://doi.org/10.1038/35009036>.
- Park, G. I., and P. Moin, 2014: An improved dynamic non-equilibrium wall-model for large eddy simulation. *Phys. Fluids*, **26**, 015108, <https://doi.org/10.1063/1.4861069>.
- Perlman, E., R. Burns, Y. Li, and C. Meneveau, 2007: Data exploration of turbulence simulations using a database cluster. *Proc. 2007 ACM/IEEE Conf. on Supercomputing*, Reno, NV, Association for Computing Machinery, 23, <https://doi.org/10.1145/1362622.1362654>.
- Piomelli, U., and E. Balaras, 2002: Wall-layer models for large-eddy simulations. *Annu. Rev. Fluid Mech.*, **34**, 349–374, <https://doi.org/10.1146/annurev.fluid.34.082901.144919>.
- Prandtl, L., 1933: Recent results of turbulence research. NACA Tech. Memo. 720, 31 pp., <https://ntrs.nasa.gov/api/citations/19930094697/downloads/19930094697.pdf>.
- Pressel, K. G., S. Mishra, T. Schneider, C. M. Kaul, and Z. Tan, 2017: Numerics and subgrid-scale modeling in large eddy simulations of stratocumulus clouds. *J. Adv. Model. Earth Syst.*, **9**, 1342–1365, <https://doi.org/10.1002/2016MS000778>.
- Raupach, M., 1981: Conditional statistics of Reynolds stress in rough-wall and smooth-wall turbulent boundary layers. *J. Fluid Mech.*, **108**, 363–382, <https://doi.org/10.1017/S0022112081002164>.
- Roberts, B., M. Xue, and D. T. Dawson II, 2020: The effect of surface drag strength on mesocyclone intensification and tornado genesis in idealized supercell simulations. *J. Atmos. Sci.*, **77**, 1699–1721, <https://doi.org/10.1175/JAS-D-19-0109.1>.
- Salesky, S. T., and W. Anderson, 2018: Buoyancy effects on large-scale motions in convective atmospheric boundary layers: Implications for modulation of near-wall processes. *J. Fluid Mech.*, **856**, 135–168, <https://doi.org/10.1017/jfm.2018.711>.
- Schenkman, A. D., M. Xue, and M. Hu, 2014: Tornado genesis in a high-resolution simulation of the 8 May 2003 Oklahoma City supercell. *J. Atmos. Sci.*, **71**, 130–154, <https://doi.org/10.1175/JAS-D-13-073.1>.
- Schumann, U., 1990: Large-eddy simulation of the up-slope boundary layer. *Quart. J. Roy. Meteor. Soc.*, **116**, 637–670, <https://doi.org/10.1002/qj.49711649307>.
- Shapley, L. S., 1953: A value for n-person games. *Contributions to the Theory of Games*, Vol. II, H. W. Kuhn, and A. W. Tucker, Eds., Princeton University Press, 307–317.
- Slotnick, J. P., A. Khodadoust, J. Alonso, D. Darmofal, W. Gropp, E. Lurie, and D. J. Mavriplis, 2014: CFD vision 2030 study: A path to revolutionary computational aerosciences. NASA Tech. Rep. NASA/CR–2014-218178, 58 pp., <https://ntrs.nasa.gov/api/citations/20140003093/downloads/20140003093.pdf>.
- Smolarkiewicz, P. K., and W. W. Grabowski, 1990: The multidimensional positive definite advection transport algorithm: Nonoscillatory option. *J. Comput. Phys.*, **86**, 355–375, [https://doi.org/10.1016/0021-9991\(90\)90105-A](https://doi.org/10.1016/0021-9991(90)90105-A).
- Stevens, R. J., R. Verzicco, and D. Lohse, 2010: Radial boundary layer structure and Nusselt number in Rayleigh–Bénard convection. *J. Fluid Mech.*, **643**, 495–507, <https://doi.org/10.1017/S0022112009992461>.
- , D. Lohse, and R. Verzicco, 2011: Prandtl and Rayleigh number dependence of heat transport in high Rayleigh number thermal convection. *J. Fluid Mech.*, **688**, 31–43, <https://doi.org/10.1017/jfm.2011.354>.
- , E. P. van der Poel, S. Grossmann, and D. Lohse, 2013: The unifying theory of scaling in thermal convection: The updated prefactors. *J. Fluid Mech.*, **730**, 295–308, <https://doi.org/10.1017/jfm.2013.298>.
- , A. Blass, X. Zhu, R. Verzicco, and D. Lohse, 2018: Turbulent thermal superstructures in Rayleigh–Bénard convection. *Phys. Rev. Fluids*, **3**, 041501, <https://doi.org/10.1103/PhysRevFluids.3.041501>.
- Stiperski, I., and M. Calaf, 2023: Generalizing Monin-Obukhov similarity theory (1954) for complex atmospheric turbulence. *Phys. Rev. Lett.*, **130**, 124001, <https://doi.org/10.1103/PhysRevLett.130.124001>.
- Štrumbelj, E., and I. Kononenko, 2014: Explaining prediction models and individual predictions with feature contributions. *Knowl. Inf. Syst.*, **41**, 647–665, <https://doi.org/10.1007/s10115-013-0679-x>.
- Sullivan, C. B., and A. A. Kaszynski, 2019: PyVista: 3D plotting and mesh analysis through a streamlined interface for the Visualization Toolkit (VTK). *J. Open Source Software*, **4**, 1450, <https://doi.org/10.21105/joss.01450>.
- Thomas, S., M. Ovchinnikov, F. Yang, D. van der Voort, W. Cantrell, S. K. Krueger, and R. A. Shaw, 2019: Scaling of an atmospheric model to simulate turbulence and cloud microphysics in the Pi Chamber. *J. Adv. Model. Earth Syst.*, **11**, 1981–1994, <https://doi.org/10.1029/2019MS001670>.
- , F. Yang, M. Ovchinnikov, W. H. Cantrell, and R. A. Shaw, 2023: Scaling of turbulence and microphysics in a convection–cloud chamber of varying height. *J. Adv. Model. Earth Syst.*, **15**, e2022MS003304, <https://doi.org/10.1029/2022MS003304>.
- Thorndike, R. L., 1953: Who belongs in the family? *Psychometrika*, **18**, 267–276, <https://doi.org/10.1007/BF02289263>.
- Vadrot, A., X. I. A. Yang, H. J. Bae, and M. Abkar, 2023: Log-law recovery through reinforcement-learning wall model for large eddy simulation. *Phys. Fluids*, **35**, 055122, <https://doi.org/10.1063/5.0147570>.
- Wang, A., Y. Pan, and P. M. Markowski, 2020: The influence of turbulence memory on idealized tornado simulations. *Mon. Wea. Rev.*, **148**, 4875–4892, <https://doi.org/10.1175/MWR-D-20-0031.1>.
- , —, and —, 2021: The influence of WENO schemes on large-eddy simulations of a neutral atmospheric boundary layer. *J. Atmos. Sci.*, **78**, 3613–3628, <https://doi.org/10.1175/JAS-D-21-0033.1>.
- , —, G. H. Bryan, and P. M. Markowski, 2023: Modeling near-surface turbulence in large-eddy simulations of a tornado: An application of thin boundary layer equations. *Mon. Wea. Rev.*, **151**, 1587–1607, <https://doi.org/10.1175/MWR-D-22-0060.1>.
- , M. Ovchinnikov, F. Yang, S. Schmalfluss, and R. A. Shaw, 2024: Designing a convection–cloud chamber for collision-coalescence using large-eddy simulation with bin microphysics. *J. Adv. Model. Earth Syst.*, **16**, e2023MS003734, <https://doi.org/10.1029/2023MS003734>.
- Wang, M., and P. Moin, 2002: Dynamic wall modeling for large-eddy simulation of complex turbulent flows. *Phys. Fluids*, **14**, 2043–2051, <https://doi.org/10.1063/1.1476668>.

- Wyngaard, J. C., and O. Coté, 1974: The evolution of a convective planetary boundary layer—A higher-order-closure model study. *Bound.-Layer Meteor.*, **7**, 289–308, <https://doi.org/10.1007/BF00240833>.
- Yang, F., M. Ovchinnikov, S. Thomas, A. Khain, R. McGraw, R. A. Shaw, and A. M. Vogelmann, 2022: Large-eddy simulations of a convection cloud chamber: Sensitivity to bin microphysics and advection. *J. Adv. Model. Earth Syst.*, **14**, e2021MS002895, <https://doi.org/10.1029/2021MS002895>.
- Yang, X. I. A., and K. P. Griffin, 2021: Grid-point and time-step requirements for direct numerical simulation and large-eddy simulation. *Phys. Fluids*, **33**, 015108, <https://doi.org/10.1063/5.0036515>.
- , G. I. Park, and P. Moin, 2017: Log-layer mismatch and modeling of the fluctuating wall stress in wall-modeled large-eddy simulations. *Phys. Rev. Fluids*, **2**, 104601, <https://doi.org/10.1103/PhysRevFluids.2.104601>.
- , S. Zafar, J.-X. Wang, and H. Xiao, 2019: Predictive large-eddy-simulation wall modeling via physics-informed neural networks. *Phys. Rev. Fluids*, **4**, 034602, <https://doi.org/10.1103/PhysRevFluids.4.034602>.
- Zhou, Z., G. He, and X. Yang, 2021: Wall model based on neural networks for LES of turbulent flows over periodic hills. *Phys. Rev. Fluids*, **6**, 054610, <https://doi.org/10.1103/PhysRevFluids.6.054610>.