AGU ADVANCING EARTH AND SPACE SCIENCES

Geophysical Research Letters

RESEARCH LETTER
10.1029/2025GL120100

Gray-Zone Simulation of the Dry Convective Boundary Layer Using Fourier Neural Operator

Tengfei Luo[1,2], Zhijie Li[3], Jianchun Wang[4], Pak Wai Chan[5], Naigeng Wu[6], and Xiaoming Shi[2,7]

[1]Center for Ocean Research in Hong Kong and Macau, Hong Kong University of Science and Technology, Hong Kong, China, [2]Division of Environment and Sustainability, Hong Kong University of Science and Technology, Hong Kong, China, [3]Department of Biomedical Engineering, National University of Singapore, Singapore, [4]Department of Mechanics and Aerospace Engineering, Southern University of Science and Technology, Shenzhen, China, [5]Hong Kong Observatory, Hong Kong, China, [6]Guangdong Meteorological Observatory, Marine Weather Forecast Center of South China Sea, Guangzhou, China, [7]IAS Center for AI for Scientific Discoveries, Hong Kong University of Science and Technology, Hong Kong, China

Key Points:

- Three-dimensional Fourier neural operator (FNO) models are applied to simulate the dry convective boundary layer gray zone at 800-m grid spacing
- The statistics and instantaneous structures predicted by the FNO models are notably better than those by gray-zone numerical simulations
- The FNO models exhibit a good generalization performance for different surface heat fluxes prescribed in the simulations

Supporting Information:

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

Correspondence to:

X. Shi,
shixm@ust.hk

Citation:

Luo, T., Li, Z., Wang, J., Chan, P. W., Wu, N., & Shi, X. (2026). Gray-zone simulation of the dry convective boundary layer using Fourier neural operator. *Geophysical Research Letters*, *53*, e2025GL120100. https://doi.org/10.1029/2025GL120100

Received 24 OCT 2025
Accepted 18 JUL 2026

Abstract

Previous machine learning (ML) weather prediction models primarily focus on global mesoscale forecasting. This study develops three-dimensional Fourier neural operator (FNO) models for simulating the dry convective boundary layer (CBL) at 800-m grid spacing, which is a resolution in the gray zone. The filtered large-eddy simulation data of the CBL are used for training the FNO models. The FNO models outperform traditional gray-zone simulations in predicting a variety of flow statistics and instantaneous structures, including the scale of updrafts and downdrafts, the vertical profiles of statistics, and the energy spectra. The FNO models, trained using data for a few surface heat fluxes $Q_s = 0.14 - 0.26\ \mathrm{K\ m\ s^{-1}}$, demonstrate certain generalization capability for other Q_s within and out of this range $\left(Q_s = 0.1 - 0.3\ \mathrm{K\ m\ s^{-1}}\right)$. The FNO model is a promising ML method for fast and accurate weather prediction in the gray zone.

Plain Language Summary

Data-driven machine learning (ML) weather prediction models have been developed rapidly and have shown state-of-the-art performance over the past few years. However, global weather forecasting models cannot well resolve the characteristics of convection and dominant turbulence in the gray zone. This paper develops three-dimensional Fourier neural operator (FNO) models in the idealized convective boundary layer gray zone, which can effectively reconstruct turbulent eddies and convection. The three-dimensional FNO model is a potential approach for developing regional ML weather prediction models.

1. Introduction

With increasing computational resources, kilometer-scale resolutions have become possible in numerical weather prediction (NWP) (Chow et al., 2019; Honnert et al., 2020; Lean et al., 2024). The kilometer mesh moves atmospheric modeling into the gray zone (or terra incognita) (Honnert et al., 2011; Wang et al., 2025; Wyngaard, 2004), where the effective resolution (filter length scale) is comparable to the scale of the dominant turbulent eddies in the atmospheric boundary layer. Numerical simulations require parameterization of turbulence and convection, but this is quite challenging in the gray zone. Parameterization methods limit the accuracy of high-resolution NWP. The newly developed purely data-driven machine learning (ML) methods would be a better choice, as they can avoid the challenges of parameterization.

In recent years, many machine learning weather prediction (MLWP) models have been developed and have shown an outstanding performance in both accuracy and efficiency (de Burgh-Day & Leeuwenburg, 2023; Schultz et al., 2021). MLWP models consist of two categories. One category is purely data-driven ML models, including FourCastNet (Pathak et al., 2022), PanguWeather (Bi et al., 2023), GraphCast (Lam et al., 2023), and FuXi (Chen et al., 2023). The other category is hybrid models with ML-based physics parameterizations, such as NeuralGCM (Kochkov et al., 2024). These MLWP models, trained using ERA5 reanalysis data, can achieve global weather forecasting but only reach 0.25° resolution. Their coarse resolution cannot resolve the characteristics of convection or dominant turbulence in the planetary boundary layer (Ben Bouallegue et al., 2024; Bonavita, 2024). The error of the global ML models increases with increasing wavenumber and with increasing forecast time, possibly due to difficulties in representing fundamental dynamical processes, such as the geostrophic effect and vertical motions (Bonavita, 2024). Studies on ML weather models focusing on higher horizontal resolution are still limited and largely unexplored.

Fourier neural operator (FNO) is a data-driven method for solving partial differential equations proposed by Li et al. (2020). Based on two-dimensional (2D) FNO or Spherical FNO (Bonev et al., 2023), FourCastNet and ACE (AI2 Climate Emulator) have demonstrated a strong performance in global weather forecasting and climate simulation (Bonev et al., 2025; Pathak et al., 2022; Watt-Meyer et al., 2023, 2025). Bi et al. (2023) pointed out that 3D models can capture the relationship between atmospheric states at different heights and thus yield significant accuracy gains compared to 2D models. The 3D FNO models have been developed and shown to perform well in predicting 3D turbulence (Li et al., 2022; Luo et al., 2024; Wang et al., 2024) and urban microclimate (Peng et al., 2024). Thus, we aim to explore the potential to apply 3D FNO for predictions in the gray zone, as a test for future development of a regional ML weather model. In the current study, we trained 3D FNO models using filtered large-eddy simulation data, achieving accurate and efficient predictions in the dry convective boundary layer (CBL) gray zone.

2. Experiment Design and Methods

2.1. Experiment Design

Following the setup of Sullivan and Patton (2011), a daytime dry convective atmospheric boundary layer driven by constant surface buoyancy fluxes Q_s, ranging from 0.10 to 0.30 K m s^{-1}, and weak geostrophic winds $(U_g, V_g) = (1, 0)$ m s^{-1} is studied in a simulation domain $\left(L_x, L_y, L_z\right) = (25.6, 25.6, 2.56)$ km. The Coriolis parameter $f = 1 \times 10^{-4}$ s^{-1}, initial inversion height $z_i = 1024$ m, and the initial virtual potential temperature θ_0 has a three-layer structure:

$$\theta_0(z) = \begin{cases} 300\ \mathrm{K} & : 0 < z < 974\ \mathrm{m}, \\ 300\ \mathrm{K} + (z - 974\ \mathrm{m}) \times 0.08\ \mathrm{K\,m^{-1}} & : 974\ \mathrm{m} < z < 1074\ \mathrm{m}, \\ 308\ \mathrm{K} + (z - 1074\ \mathrm{m}) \times 0.003\ \mathrm{K\,m^{-1}} & : z > 1074\ \mathrm{m}. \end{cases} \tag{1}$$

2.2. Cloud Model 1

Cloud Model 1 (CM1, release 21.1) is a state-of-the-art atmospheric model, which is designed to solve three-dimensional, non-hydrostatic, compressible equations of the moist atmosphere (Bryan & Fritsch, 2002). We used the CM1 to produce the benchmark simulation of the dry CBL with periodic boundary conditions in the horizontal (x, y) directions and a no-slip top boundary condition in the vertical z direction. A damping zone is used at $z \geq 2$ km. The Monin–Obukhov similarity theory (MOST) for the surface layer is applied based on a drag rule with the surface roughness $z_0 = 0.1$ m. The TKE turbulence closure scheme is used to model the subgrid-scale turbulence.

For large-eddy simulations (LES), the horizontal grid spacing $\Delta_x = \Delta_y$ is 50 m, and the vertical grid spacing Δ_z is 20 m. There are a total of $512^2 \times 128$ grid points. Although the grid setup does not fully satisfy the requirement of $z_i/\Delta > 60$ in CBL (Sullivan & Patton, 2011), where $\Delta^3 = \Delta_x \Delta_y \Delta_z$, the current LES still provides well resolved results (Zhou et al., 2014) and is sufficient for our study. The simulations are started from small random temperature perturbations at $z < 100$ m. To ensure that similar perturbations are applied in the gray-zone simulations, the range of non-dimensional wavenumber of the initial perturbation is limited to $k_x, k_y \leq 16$.

The gray-zone simulations are performed with grid meshes of $32^2 \times 64$, corresponding to 800 m horizontal spacings and 40 m vertical spacing. The simulation methods are the same as those used in LES simulations.

2.3. Database

We performed 35 LES simulations using CM1 for surface heat fluxes $Q_s = 0.14$, 0.16, 0.18, 0.20, 0.22, 0.24, 0.26 K m s^{-1}, each of which contains five different initial temperature perturbations. The LES was run for 6 hr, with instantaneous flow field data saved every 5 min. Therefore, we have a total of $N_t = 73$ time slices. In the dry CBL, the potential temperature θ and velocity components u_i are the most important physical quantities.

The LES data with $512^2 \times 128$ grids $(50\ \text{m} \times 50\ \text{m} \times 20\ \text{m})$ are filtered and downsampled to the coarse grid with $32^2 \times 32$ grids $(800\ \text{m} \times 800\ \text{m} \times 80\ \text{m})$ as the benchmark and the training data of the FNO models. We utilized a 2D sharp spectral filter in the $x - y$ planes at filter width $\Delta_f = 16\Delta_x$ (Tong et al., 1998; Zhou et al., 2014) and a top-hat filter in the z direction at filter width $\Delta_f = 4\Delta_z$. The coarse grid flow field is referred to as filtered LES (fLES).

To increase the data amount, we downsampled two fLES data sets from a single LES by applying a grid shift of $\Delta_f/2 = 400$ m in the x and y directions for the second data set. Thus, we have $N_c \times N_t = 70 \times 73$ sets of data. Each set of data has a size of $[N_x \times N_y \times N_z \times N_v]$, where $N_x, N_y, N_z = 32$ represent the number of grid points in three spatial dimensions, and $N_v = 5$ is the number of physical variables $\mathbf{f}$, including $\theta - \theta_0$, $u_i - U_i$, and Q_j, where θ_0 is the initial potential temperature profile, and $\mathbf{U} = (U_g, V_g, 0)$ is the geostrophic wind. Q_j is designed to represent different heat fluxes, given by

$$Q_j(x,y,z) = A\,\frac{Q_{s,j}}{Q_n} \times \exp\left(-B\,\frac{Q_n}{Q_{s,i}}\,\frac{z-40}{2560}\right), \tag{2}$$

where $j = 1, \ldots, N_c$ represents the different cases, $Q_{s,j} = 0.14, 0.16, 0.18, 0.20, 0.22, 0.24, 0.26\ \text{K m s}^{-1}$, $Q_n = \sum_{j=1}^{N_c} Q_{s,j}/N_c = 0.2\ \text{K m s}^{-1}$ is the average value of all the heat fluxes, $A = 1.0$ is a parameter that controls the amplitude, and $B = 1.0$ is a parameter that controls the decay rate. The boundary condition mainly affects the turbulence at the bottom of the boundary layer, and its influence diminishes with height. Therefore, we believe that Q_j should decay with height. Q_j as an input can enhance the generalization ability of the FNO models for different surface heat fluxes, which will be discussed in Section 3.2 and in Text S2 in Supporting Information S1. It is important to note that although the model does generate an output for $Q_j(x,y,z)$ at each prediction step, the internal prediction is discarded and replaced by the values described in Equation 2. The correction is crucial for achieving accurate long-term prediction.

2.4. Fourier Neural Operator

Different from many neural networks, which are capable of approximating mappings between finite-dimensional Euclidean spaces, the Fourier neural operator learns the mappings of the high-dimensional data in Fourier spaces by training on a finite set of input-output pairs (Li et al., 2020). The FNO exhibits strong adaptability and generalization ability under different flow configurations.

The data-driven FNO model learns the approximating nonlinear mappings between input a and output u. The architecture of FNO is briefly described as follows (Li et al., 2020, 2022bib_li_et_al_2022; Luo et al., 2024):

1. The input a is projected to a higher dimensional representation v_0 by a shallow fully connected neural network P, denoted as $v_0 = P(a)$.
2. The high-dimensional v_0 is updated through N_f iterations by

$$v_{t+1} = \sigma(Wv_t + \mathcal{K}v_t), \tag{3}$$

where W is a linear transformation, $\mathcal{K}$ is the Fourier integral operator, and σ is a nonlinear activation function.

3. The output u is obtained by the projection of v_{N_f}, denoted as $u = Q(v_{N_f})$, where Q is a fully connected layer.

The training process of the FNO model requires learning the P, W, $\mathcal{K}$, and Q. We provide a more detailed introduction of the FNO model in Text S1 in Supporting Information S1.

The inputs of the FNO model are generally the known states of the previous time slices, and the outputs are the flow fields of the next time slices. We effectively utilized 70×73 sets of data during the training of the FNO model. 80% of the data are used for training and 20% for validation. To avoid reliance on traditional numerical simulations, we trained two FNO models: FNO-I1O1 and FNO-I5O1. The FNO-I1O1 model uses the flow field of the previous time slice $\mathbf{f}_t$ as input, while the FNO-I5O1 model uses the flow fields of the previous five time slices $\mathbf{f}_{t-4}, \mathbf{f}_{t-3}, \mathbf{f}_{t-2}, \mathbf{f}_{t-1}, \mathbf{f}_t$. For both the FNO-I1O1 and FNO-I5O1 models, the flow fields of the next time slice $\mathbf{f}_{t+1}$ are taken as output. The initial five time slices of the FNO-I5O1 model can be generated by the FNO-I1O1 model.

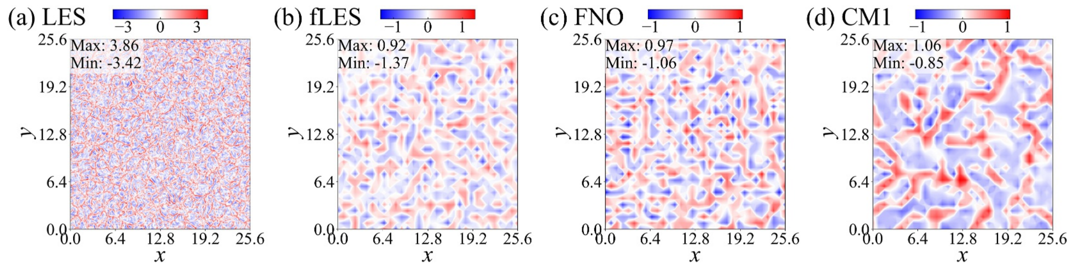

Figure 1. Horizontal contours of vertical velocity w (m/s) in the CBL from: (a) the large-eddy simulations (LES) with 50 m horizontal resolution using CM1, (b) the LES result filtered onto a 800 m grid, (c) the three-dimensional FNO models with 800 m horizontal resolution, and (d) the gray-zone CM1 simulations with 800 m horizontal resolution. Contours are from the horizontal plane at $z = 120$ m and at $t = 4$ h. Red is the updrafts, and blue is the downdrafts.

The stability and generalization capability of the FNO-I5O1 model are better than those of the FNO-I1O1 model (Figure S5 in Supporting Information S1), because the FNO-I5O1 model utilizes the differences between the previous time slices.

A series of parameters (see Text S1 in Supporting Information S1) need to be optimized during the training of the FNO models. The truncated Fourier modes $k_{x_i,\max}$ of the Fourier integral operator $\mathcal{K}$ are chosen to be approximately 2/3 of the modes based on the minimum grid numbers (i.e., $k_{x_i,\max} \approx N_{x_i}/3$) at the LES scales, in order to effectively reconstruct small-scale structures and reduce effects of high-frequency noise (Li et al., 2022; Luo et al., 2024). In the gray zone, the scales of the dominant eddies are comparable to z_i. Therefore, $k_{z,\max} = 6$ (corresponding to approximately 200 m grids) is sufficient. If $k_{z,\max}$ is larger, high-frequency fluctuations may appear in the vertical direction. In the horizontal directions, the horizontal spacings are close to the scales of the dominant eddies. Thus, $k_{x,\max} = k_{y,\max} = 16$ (corresponding to 800 m grids) allows for better reconstruction of the cell structures. Additionally, four Fourier layers are used $\left(N_f = 4\right)$, the channel width d_v is 96, the batch size is 4, and the initial learning rate is 10^{-3} which decays by half after every eight epochs. The FNO models are trained for 30 epochs. The adaptive moment optimization (Adam) algorithm and GELU function are chosen as the optimizer and activation function, respectively (Kingma & Ba, 2014; Li et al., 2020). The loss function is

$$\text{Loss} = \frac{\|\mathbf{f} - \mathbf{f}_{\text{pre}}\|_2}{\|\mathbf{f}\|_2}, \text{ where } \|\mathbf{A}\|_2 = \frac{1}{\sqrt{n}}\sqrt{\sum_{i=1}^{n} A_i^2}, \tag{4}$$

where $\mathbf{f}$ and $\mathbf{f}_{\text{pre}}$ represent the filtered flow field and the predicted flow field, respectively.

3. Results

3.1. Comparison Between FNO Models and Traditional Simulations

The a posteriori tests of the FNO models and the gray-zone CM1 simulations with 800 m horizontal spacing are performed for $Q_s = 0.24$ K m s^{-1}. Although the case of $Q_s = 0.24$ K m s^{-1} is included in the training data set, the test uses a different initial condition. We assess the performance of FNO models in predicting the flow statistics and spatial structures. The FNO-I1O1 model predicts the flow field at $0 - 0.5$ h, and the FNO-I5O1 model predicts the flow field at $0.5 - 6$ h, using the flow fields predicted by FNO-I1O1 as the initial five time slices.

3.1.1. Flow Visualization

Figure 1 presents the horizontal contours of vertical velocity at $z = 120$ m $\sim 0.1\, z_i$ and at $t = 4$ h. We observe the classic organized cell structures in the benchmark LES simulation. In the lower CBL, the warm air near the surface rises, presenting narrow spoke-like updrafts, while the downdrafts appear in broader block-like patterns, as shown in Figure 1a. The small updrafts merge into large-scale plumes with increasing height (Hunt et al., 1988; Lenschow & Stephens, 1980). Filtering at the gray zone scale averages out the small-scale updrafts, making the scale of the updrafts comparable to that of the downdrafts. Both updrafts and downdrafts are uniformly distributed

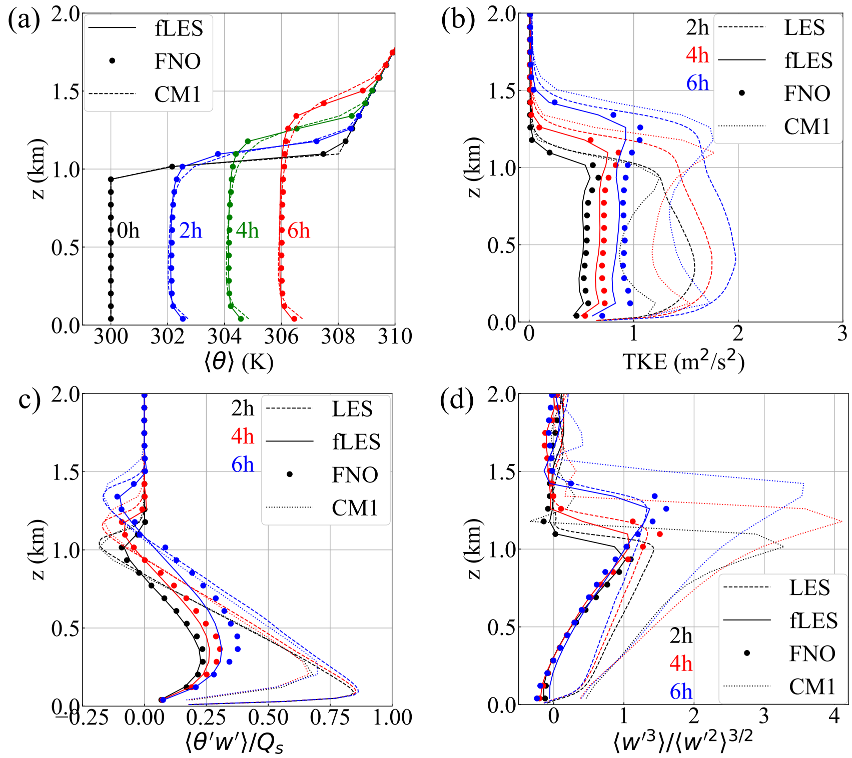

Figure 2. Vertical profiles of (a) potential temperature, (b) turbulent kinetic energy, (c) potential temperature flux, and (d) skewness of the vertical velocity.

in the domain, as illustrated in Figure 1b. Additionally, we observed that the filtered plume structures do not significantly change with height.

As shown in Figure 1d, the gray-zone CM1 simulations predict narrow updrafts and broad downdrafts (Zhou et al., 2014). The scale of the strong updrafts is close to that of the filtered updrafts, while the scale of the downdrafts is much larger than the fLES results. The cell structures resemble enlarged LES results, showing significant differences from fLES. In contrast, the FNO is a purely data-driven ML model that successfully learns the flow characteristics from the filtered flow field. As a result, the structures predicted by the FNO models (Figure 1c) are essentially the same as those of fLES at all heights (Figure S6 in Supporting Information S1).

3.1.2. Flow Statistics

Driven by a surface heat flux, the boundary layer height z_i increases, and the potential temperature increases with time. Figure 2a presents the vertical profiles of potential temperature $\langle\theta\rangle$, where $\langle\cdot\rangle$ denotes the horizontal average value, computed without density weighting. Due to stronger turbulence fluctuations near the surface and a much weaker inversion near z_i (Sullivan & Patton, 2011), the average temperature of gray-zone CM1 simulations reaches higher values near the surface and near z_i. However, the FNO models accurately predict $\langle\theta\rangle$, owing to their ability to effectively reconstruct the instantaneous flow structures.

The fluctuation of a variable ψ is $\psi' = \psi - \langle\psi\rangle$. The turbulent kinetic energy TKE $= \langle u_i' u_i' \rangle/2$ is one of the most important statistics characterizing turbulence intensity. In the gray zone, the scales of the dominant turbulence structures are comparable to the horizontal grid spacings, leading to partially resolved turbulence. The ratio of resolved TKE to total TKE is around 0.5 at 800 m filter scale. Figure 2b shows the vertical profiles of resolved TKE. The profiles of resolved TKE predicted by FNO models are in good agreement with the fLES results. The gray-zone CM1 simulations overestimate the prediction of resolved TKE, which is more comparable to the total resolved values in the LES than the filtered LES, because it underestimates the SGS component.

The vertical profiles of normalized potential temperature flux $\langle\theta' w'\rangle/Q_s$ are presented in Figure 2c. The resolved potential temperature flux in the LES or the gray-zone CM1 simulations is positive in the bottom half of the CBL, and is negative near z_i, which is consistent with previous simulations (Zhou et al., 2014). The resolved temperature flux of the gray-zone CM1 simulations is much larger than the fLES results, especially near the surface, indicating that gray-zone CM1 simulations produce more energetic resolved convection compared to the true solution. This is because the gray zone CM1 simulations over-attribute the total heat flux to the resolved component and underestimate the SGS heat flux (Sullivan & Patton, 2011; Zhou et al., 2014). In contrast, the FNO models have a closer agreement with the fLES results, due to their effective prediction of temperature fields and turbulence fluctuations.

The skewness of the vertical velocity $\langle w'^3\rangle/\langle w'^2\rangle^{3/2}$ is an indicator of the updrafts and downdrafts distribution (Moeng & Rotunno, 1990; Weil, 1990). Figure 2d illustrates the profiles of $\langle w'^3\rangle/\langle w'^2\rangle^{3/2}$. In the high-resolution LES (Sullivan & Patton, 2011; Zhou et al., 2014), the vertical velocity skewness is mainly positive and increases with height at $z/z_i < 0.9$, which is associated with the narrow strong updrafts and broad zones of compensating downdrafts (Hunt et al., 1988), as shown in Figure 1a. However, filtering at the gray zone scales causes the scales of updrafts and downdrafts to become comparable (Figure 1b). Thus, $\langle w'^3\rangle/\langle w'^2\rangle^{3/2}$ of fLES is very small and even negative near the surface. Due to the fact that the FNO models learn the flow characteristics from the filtered flow field, the FNO models perfectly predict the profiles of $\langle w'^3\rangle/\langle w'^2\rangle^{3/2}$. In contrast, gray-zone CM1 simulations exhibit significant deviations from fLES in predicting vertical velocity skewness, with predicted values even being substantially much higher than the LES results.

In addition, we verify the performance of the FNO models in predicting other statistical quantities, including the mean, root mean square (RMS, $\psi_{RMS} = \sqrt{\langle\psi'^2\rangle}$), and skewness values of temperature and velocity components. All results are significantly better than those from gray-zone CM1 simulations, except for the mean values of vertical velocity $\langle w\rangle$. Since $\langle w\rangle$ is slightly greater than 0, and much smaller than the mean values of horizontal velocities and the RMS values, a small error in the vertical velocity can cause $\langle w\rangle$ to fluctuate around 0.

3.1.3. Turbulence Spectra

The two-dimensional energy spectra of vertical velocity fluctuations w', horizontal velocity fluctuations u', and potential temperature fluctuations θ' are shown in Figure 3. The non-dimensional horizontal wavenumber is $k_h = \sqrt{k_x^2 + k_y^2}$. Spectral peaks represent the length scales of the dominant eddies. For the vertical velocity, energy accumulates at gray zone scales, and the energy spectra E_w decrease rapidly as the wavenumber decreases at $k_h < 10$ (corresponding to approximately 2 km horizontal grids). The spectral peaks of horizontal velocity fluctuations also appear in the gray zone near the surface and in the mid-CBL, but the corresponding scales are smaller than those of E_w (Sullivan & Patton, 2011). Additionally, the decay of E_u at low wavenumbers is not as pronounced as that of E_w.

We observed that the gray-zone CM1 simulations fail to accurately predict the energy spectra. For the energy spectra of two velocity fluctuations, the CM1 simulations overestimate the energy spectra at large scales and underestimate the energy spectra at small scales. Additionally, the predicted energy spectra of potential temperature fluctuations are larger than those of fLES at all scales.

The FNO models provide excellent predictions for all three energy spectra at all heights and times. It is important to note that the cutoff wavenumber in the horizontal direction $k_{x,\max}, k_{y,\max}$ for the FNO models is 16. We tested smaller cutoff wavenumbers and found that the horizontal cutoff wavenumber significantly affects the predictions of E_w. If the wavenumber is set to 14, 12, or 10, the prediction of the FNO models at low wavenumbers is not as good as in the current results (Figure S7 in Supporting Information S1). Due to the fact that the TKE is primarily distributed in the gray zone, the horizontal cutoff wavenumber is crucial for capturing the flow characteristics.

3.1.4. Computational Efficiency

We compare the computational efficiency for different methods. Each epoch takes approximately 160 s to train on an NVIDIA RTX A6000 GPU. The entire training process for the FNO models spans around 1.4 h. The prediction of the FNO models only takes about 3 s on a GPU. The gray-zone CM1 simulations take about 90 s on the CPU of Intel Xeon Platinum 8592+ with 2 cores. The full LES takes approximately 16 h on the CPU with 32 cores. The computational efficiency of the FNO models is higher than that of the CM1 simulations.

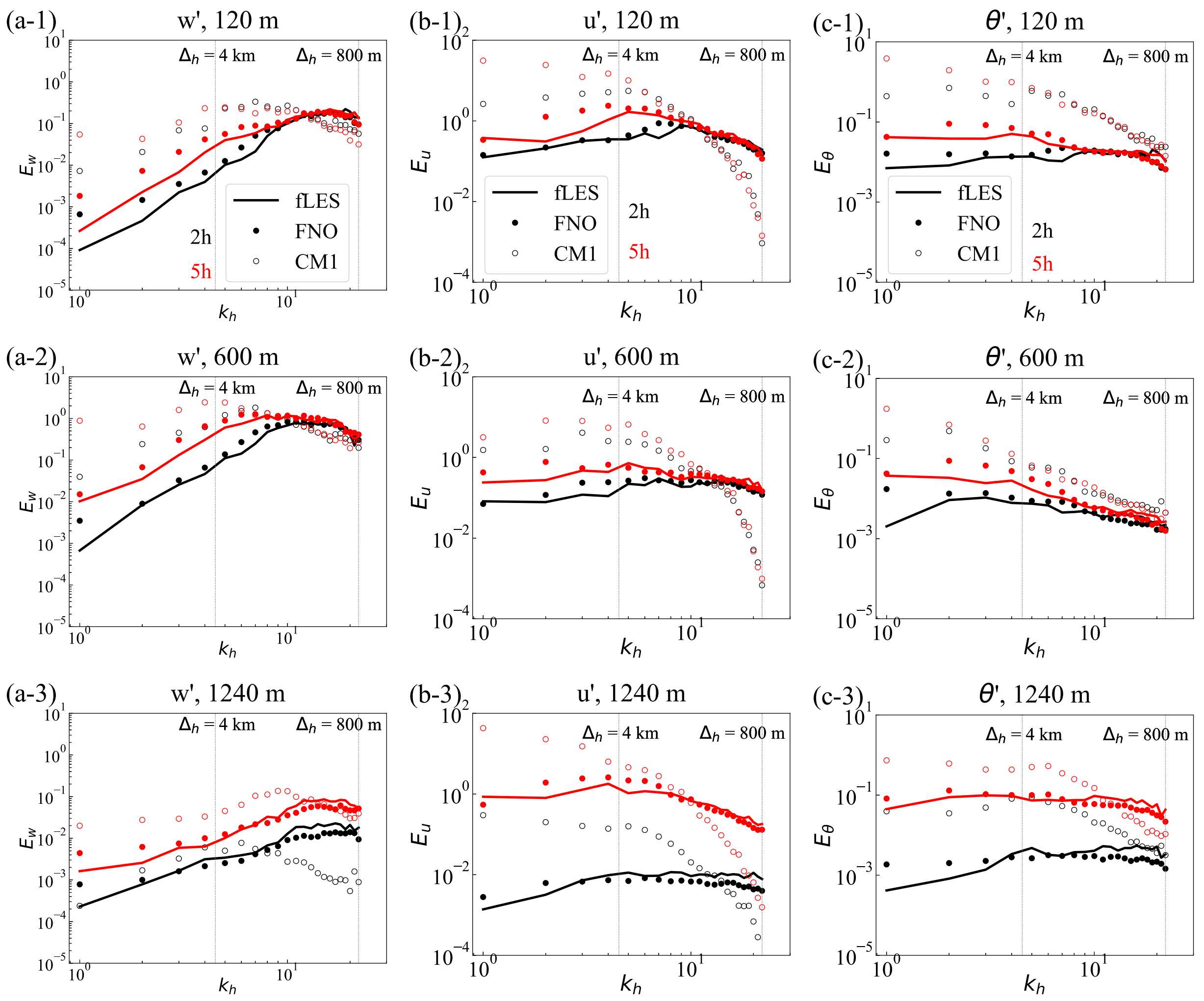

Figure 3. Two-dimensional energy spectra of (a) vertical velocity fluctuations w', (b) horizontal velocity fluctuations u', and (c) potential temperature fluctuations θ' at $t = 2, 5$ h and at $z = 120$ m (top), 600 m (middle), and 1240 m (bottom). Two dotted lines represent wavenumbers corresponding to the 800 m and 4 km horizontal grids, respectively.

3.2. Generalization Ability

In this section, we test the generalization ability of the FNO models for surface heat fluxes Q_s differing from the training data set. The training of the FNO models employed data for $Q_s = 0.14, 0.16, 0.18, 0.20, 0.22, 0.24, 0.26$ K m s^{-1}. We choose $Q_s = 0.10,\ 0.12,\ 0.15,\ 0.2,\ 0.25,\ 0.28,\ 0.3$ K m s^{-1} in the tests. Here, $Q_s = 0.20$ K m s^{-1} is included in the training data set, but a different initial condition is used. The values $Q_s = 0.15, 0.25$ K m s^{-1} represent interpolated values, while $Q_s = 0.10, 0.12, 0.28, 0.30$ K m s^{-1} indicate extrapolated values. The absolute error and relative error of a statistical variable are defined as

$$E_A = \frac{1}{N_z}\sum_{i=1}^{N_z}\sqrt{\left(\psi_{\mathrm{FNO}} - \psi\right)^2}, \quad E_R = \frac{1}{N_z}\sum_{i=1}^{N_z}\frac{\sqrt{\left(\psi_{\mathrm{FNO}} - \psi\right)^2}}{\sqrt{\psi^2}}, \tag{5}$$

where ψ and ψ_{FNO} represent the horizontal statistical variables of the fLES and predicted by the FNO models, respectively.

Figure 4 shows the absolute errors and the profiles of the average potential temperature and the RMS values of the vertical velocity for different surface heat fluxes Q_s. The impact of Q_s on the average potential temperature is the most significant, making the prediction of $\langle\theta\rangle$ the most challenging. The FNO models provide good predictions of

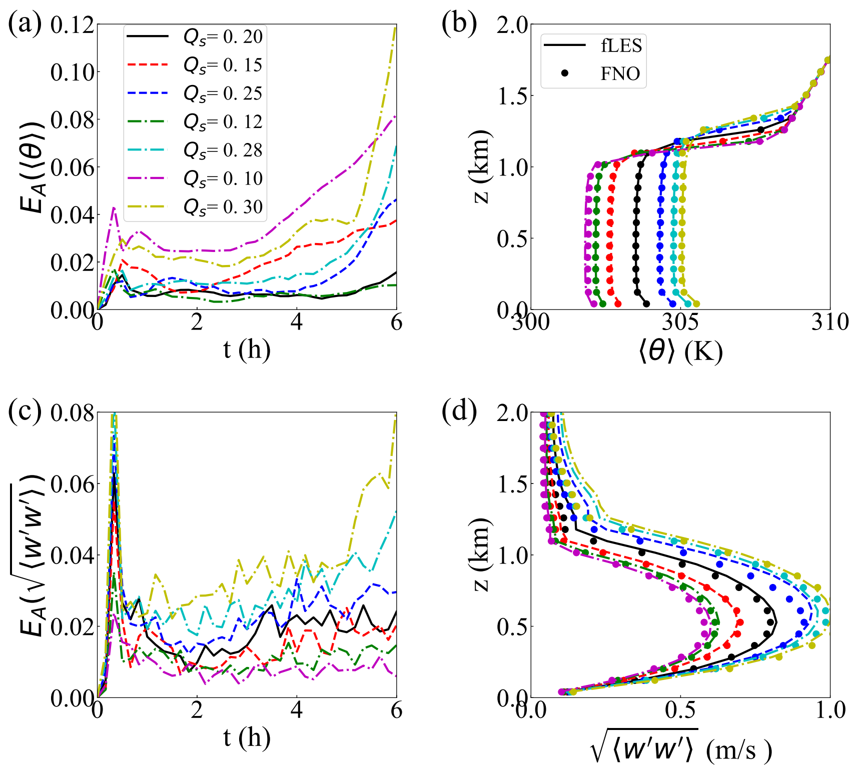

Figure 4. Statistics for different surface heat fluxes Q_s: (a) absolute error E_A and (b) profiles at $t = 4$ h of average potential temperature, (c) absolute error E_A and (d) profiles at $t = 4$ h of RMS values of the vertical velocity. The lines represent the results of fLES, and the filled circles represent the results of the FNO models.

$\langle\theta\rangle$ for all Q_s, with absolute errors less than 0.1 K. For extrapolated Q_s, although the predictions by the FNO models are better at $0 - 5$ h, there are some deviations at $t > 5$ h, with the predicted average temperature being too high for $Q_s = 0.10$ K m s^{-1} and too low near z_i for $Q_s = 0.30$ K m s^{-1} (Figure S3 in Supporting Information S1). The FNO models demonstrate a good generalization for predicting high-order statistics (Figures 4c and 4d and Figure S8 in Supporting Information S1). The absolute errors of RMS values of the vertical velocity $\sqrt{\langle w'w'\rangle}$ are less than 0.08 m s^{-1}, and the relative errors are about 0.2 at $t \geq 0.5$ h for all Q_s. Additionally, we note that the FNO models underestimate $\sqrt{\langle w'w'\rangle}$ near z_i for large Q_s.

The generalization ability of the FNO models benefits from the input of 3D heat flux Q_j and the differences of flow fields between the first five steps. If we only use the temperature and velocity fields as training data, the generalization is poor due to the small differences in the velocity fields for different Q_s (Figure S2a in Supporting Information S1). The input of 3D heat flux during training significantly enhances the generalization ability because it provides a good constraint on the temperature field in Fourier space. Furthermore, as detailed in Text S2 in Supporting Information S1, we found that the FNO models' performance is largely insensitive to the form of the Q_j. In training the 3D FNO models, we need to select appropriate input variables based on physical and mathematical knowledge to enhance their generalization ability. Additionally, we need to design a reasonable 3D representation of the 2D effects that act on the boundaries.

We tested the FNO model's generalization on wind configurations (U_g, V_g) in Text S3 in Supporting Information S1. The FNO models demonstrate better generalization capability for predicting θ, u, and w at $U_g = 0.5 - 2.0$ m s^{-1}. However, the mean values of v are not accurate. This is because the velocity v is generated by the Coriolis force, and the training only uses data of wind speed $U_g = 1.0$ m s^{-1}. Thus, the FNO models did not learn the effects of the Coriolis force. To achieve more accurate predictions, we need to use data with different wind configurations to learn the effects of the Coriolis force.

4. Conclusions

In this study, we apply the three-dimensional Fourier neural operator method to predict the dry convective boundary layer in the gray zone. The spatially filtered data from 35 LES simulations for surface heat fluxes $Q_s = 0.14 - 0.26$ K m s^{-1} are used to train the FNO models with a horizontal spacing of 800 m and a vertical spacing of 80 m, which only allows partial resolving of the energetic eddies in the convective boundary layer.

In the a posteriori tests, the FNO models can accurately predict instantaneous spatial structures of convective eddies and statistical characteristics of the boundary layer. The structures of vertical velocity near the surface predicted by the FNO models are consistent with those of the filtered LES, overcoming the issue of overly large cell structures predicted by traditional numerical simulations. The FNO models perform better than the gray-zone CM1 simulations in predicting profiles of flow statistics, including mean values, RMS values, and skewness of temperature and velocity. Furthermore, the temperature and velocity spectra predicted by the FNO models are close to the results of filtered LES.

The generalization ability of the FNO models on different surface heat fluxes Q_s is evaluated. The FNO models exhibit certain generalization ability in predicting various statistical quantities for interpolated $Q_s = 0.15, 0.25$ K m s^{-1} and extrapolated $Q_s = 0.10, 0.12, 0.28, 0.30$ K m s^{-1}. The input of the heat flux Q_j enhances the generalization ability of the FNO models. The choice and design of input variables are critical for the performance of the FNO models.

The 3D FNO models perform well in the idealized CBL, effectively reconstructing the dominant turbulent eddies and convection characteristics in the gray zone. As a purely data-driven machine learning model, its stability and generalization ability still need further improvement. Our future work aims to improve the stability and generalization ability and test the performance of the FNO model for more complex atmospheric flows.

Conflict of Interest

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

Availability Statement

The CM1 code can be downloaded from https://github.com/NCAR/CM1. The namelist files used to run LES and gray-zone simulations, the FNO code and models, and the related data used in this study can be found in Luo (2026).

Acknowledgments

The authors thank three anonymous reviewers for valuable comments. The authors thank three anonymous reviewers for valuable comments. The work was substantially supported by a grant from the Research Grants Council (RGC) of the Hong Kong Special Administrative Region, China (Project Reference Number: AoE/P-601/23-N), and the Center for Ocean Research in Hong Kong and Macau (CORE), a joint research center between the Laoshan Laboratory and the Hong Kong University of Science and Technology (HKUST). XS is also supported by the RGC Grant HKUST-16301322. XS and NW are also supported by the Youth Innovation Team of China Meteorological Administration (CMA2024QN01). The authors thank HKUST Fok Ying Tung Research Institute and the National Supercomputing Center in Guangzhou Nansha sub-center for providing high-performance computational resources.

References

Ben Bouallegue, Z., Clare, M. C., Magnusson, L., Gascon, E., Maier-Gerber, M., Janoušek, M., et al. (2024). The rise of data-driven weather forecasting: A first statistical assessment of machine learning–based weather forecasts in an operational-like context. *Bulletin of the American Meteorological Society*, *105*(6), E864–E883. https://doi.org/10.1175/bams-d-23-0162.1

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023). Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, *619*(7970), 533–538. https://doi.org/10.1038/s41586-023-06185-3

Bonavita, M. (2024). On some limitations of current machine learning weather prediction models. *Geophysical Research Letters*, *51*(12), e2023GL107377. https://doi.org/10.1029/2023gl107377

Bonev, B., Kurth, T., Hundt, C., Pathak, J., Baust, M., Kashinath, K., & Anandkumar, A. (2023). Spherical Fourier neural operators: Learning stable dynamics on the sphere. In *International Conference on Machine Learning* (pp. 2806–2823).

Bonev, B., Kurth, T., Mahesh, A., Bisson, M., Kossaifi, J., Kashinath, K., et al. (2025). Fourcastnet 3: A geometric approach to probabilistic machine-learning weather forecasting at scale. *arXiv preprint arXiv:2507.12144*.

Bryan, G. H., & Fritsch, J. M. (2002). A benchmark simulation for moist nonhydrostatic numerical models. *Monthly Weather Review*, *130*(12), 2917–2928. https://doi.org/10.1175/1520-0493(2002)130<2917:absfmn>2.0.co;2

Chen, L., Zhong, X., Zhang, F., Cheng, Y., Xu, Y., Qi, Y., & Li, H. (2023). Fuxi: A cascade machine learning forecasting system for 15-day global weather forecast. *npj Climate and Atmospheric Science*, *6*(1), 190. https://doi.org/10.1038/s41612-023-00512-1

Chow, F. K., Schär, C., Ban, N., Lundquist, K. A., Schlemmer, L., & Shi, X. (2019). Crossing multiple gray zones in the transition from mesoscale to microscale simulation over complex terrain. *Atmosphere*, *10*(5), 274. https://doi.org/10.3390/atmos10050274

de Burgh-Day, C. O., & Leeuwenburg, T. (2023). Machine learning for numerical weather and climate modelling: A review. *Geoscientific Model Development*, *16*(22), 6433–6477. https://doi.org/10.5194/gmd-16-6433-2023

Honnert, R., Efstathiou, G. A., Beare, R. J., Ito, J., Lock, A., Neggers, R., et al. (2020). The atmospheric boundary layer and the "gray zone" of turbulence: A critical review. *Journal of Geophysical Research: Atmospheres*, *125*(13), e2019JD030317. https://doi.org/10.1029/2019jd030317

Honnert, R., Masson, V., & Couvreux, F. (2011). A diagnostic for evaluating the representation of turbulence in atmospheric models at the kilometric scale. *Journal of the Atmospheric Sciences*, *68*(12), 3112–3131. https://doi.org/10.1175/jas-d-11-061.1

10.1029/2025GL120100

Hunt, J., Kaimal, J., & Gaynor, J. (1988). Eddy structure in the convective boundary layer—New measurements and new concepts. *Quarterly Journal of the Royal Meteorological Society*, *114*(482), 827–858. https://doi.org/10.1002/qj.49711448202

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

Kochkov, D., Yuval, J., Langmore, I., Norgaard, P., Smith, J., Mooers, G., et al. (2024). Neural general circulation models for weather and climate. *Nature*, *632*(8027), 1060–1066. https://doi.org/10.1038/s41586-024-07744-y

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., et al. (2023). Learning skillful medium-range global weather forecasting. *Science*, *382*(6677), 1416–1421. https://doi.org/10.1126/science.adi2336

Lean, H. W., Theeuwes, N. E., Baldauf, M., Barkmeijer, J., Bessardon, G., Blunn, L., et al. (2024). The hectometric modelling challenge: Gaps in the current state of the art and ways forward towards the implementation of 100-m scale weather and climate models. *Quarterly Journal of the Royal Meteorological Society*, *150*(765), 4671–4708. https://doi.org/10.1002/qj.4858

Lenschow, D., & Stephens, P. (1980). The role of thermals in the convective boundary layer. *Boundary-Layer Meteorology*, *19*(4), 509–532. https://doi.org/10.1007/bf00122351

Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2020). Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*.

Li, Z., Peng, W., Yuan, Z., & Wang, J. (2022). Fourier neural operator approach to large eddy simulation of three-dimensional turbulence. *Theoretical and Applied Mechanics Letters*, *12*(6), 100389. https://doi.org/10.1016/j.taml.2022.100389

Luo, T. (2026). The dataset, the code, the models, and the namelist files for "Gray-zone simulation of the dry convective boundary layer using Fourier neural operator" [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.18297902

Luo, T., Li, Z., Yuan, Z., Peng, W., Liu, T., Wang, L. L., & Wang, J. (2024). Fourier neural operator for large eddy simulation of compressible Rayleigh–Taylor turbulence. *Physics of Fluids*, *36*(7), 075165. https://doi.org/10.1063/5.0213412

Moeng, C.-H., & Rotunno, R. (1990). Vertical-velocity skewness in the buoyancy-driven boundary layer. *Journal of the Atmospheric Sciences*, *47*(9), 1149–1162. https://doi.org/10.1175/1520-0469(1990)047<1149:vvsitb>2.0.co;2

Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., et al. (2022). Fourcastnet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. *arXiv preprint arXiv:2202.11214*.

Peng, W., Qin, S., Yang, S., Wang, J., Liu, X., & Wang, L. L. (2024). Fourier neural operator for real-time simulation of 3d dynamic urban microclimate. *Building and Environment*, *248*, 111063. https://doi.org/10.1016/j.buildenv.2023.111063

Schultz, M. G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L. H., et al. (2021). Can deep learning beat numerical weather prediction? *Philosophical Transactions of the Royal Society A*, *379*(2194), 20200097. https://doi.org/10.1098/rsta.2020.0097

Sullivan, P. P., & Patton, E. G. (2011). The effect of mesh resolution on convective boundary layer statistics and structures generated by large-eddy simulation. *Journal of the Atmospheric Sciences*, *68*(10), 2395–2415. https://doi.org/10.1175/jas-d-10-05010.1

Tong, C., Wyngaard, J. C., Khanna, S., & Brasseur, J. G. (1998). Resolvable-and subgrid-scale measurement in the atmospheric surface layer: Technique and issues. *Journal of the Atmospheric Sciences*, *55*(20), 3114–3126. https://doi.org/10.1175/1520-0469(1998)055<3114:rassmi>2.0.co;2

Wang, Y., Li, Z., Yuan, Z., Peng, W., Liu, T., & Wang, J. (2024). Prediction of turbulent channel flow using Fourier neural operator-based machine-learning strategy. *Physical Review Fluids*, *9*(8), 084604. https://doi.org/10.1103/physrevfluids.9.084604

Wang, Y., Xu, H., Xu, X., Liu, S., Zhang, C., & Huo, Y. (2025). Evaluating scale-aware boundary layer similarity functions and their mechanisms in tropical cyclone modeling using idealized large-eddy simulations. *Geophysical Research Letters*, *52*(8), e2024GL113337. https://doi.org/10.1029/2024gl113337

Watt-Meyer, O., Dresdner, G., McGibbon, J., Clark, S. K., Henn, B., Duncan, J., et al. (2023). Ace: A fast, skillful learned global atmospheric model for climate prediction. *arXiv preprint arXiv:2310.02074*.

Watt-Meyer, O., Henn, B., McGibbon, J., Clark, S. K., Kwa, A., Perkins, W. A., et al. (2025). ACE2: Accurately learning subseasonal to decadal atmospheric variability and forced responses. *npj Climate and Atmospheric Science*, *8*(1), 205. https://doi.org/10.1038/s41612-025-01090-0

Weil, J. C. (1990). A diagnosis of the asymmetry in top-down and bottom-up diffusion using a Lagrangian stochastic model. *Journal of the Atmospheric Sciences*, *47*(4), 501–515. https://doi.org/10.1175/1520-0469(1990)047<0501:adotai>2.0.co;2

Wyngaard, J. C. (2004). Toward numerical modeling in the "terra incognita". *Journal of the Atmospheric Sciences*, *61*(14), 1816–1826. https://doi.org/10.1175/1520-0469(2004)061<1816:tnmitt>2.0.co;2

Zhou, B., Simon, J. S., & Chow, F. K. (2014). The convective boundary layer in the terra incognita. *Journal of the Atmospheric Sciences*, *71*(7), 2545–2563. https://doi.org/10.1175/jas-d-13-0356.1