A Statistical Evaluation of WRF-LES Trace Gas Dispersion Using Project Prairie Grass Measurements

ALEX RYBCHUK,[a] CAROLINE B. ALDEN,[b,c] JULIE K. LUNDQUIST,[d,e] AND GREGORY B. RIEKER[a]

[a] *Department of Mechanical Engineering, University of Colorado Boulder, Boulder, Colorado*
[b] *Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, Colorado*
[c] *National Oceanic and Atmospheric Administration, Boulder, Colorado*
[d] *Department of Atmospheric and Oceanic Sciences, University of Colorado Boulder, Boulder, Colorado*
[e] *National Renewable Energy Laboratory, Golden, Colorado*

(Manuscript received 17 July 2020, in final form 3 March 2021)

ABSTRACT: In recent years, new measurement systems have been deployed to monitor and quantify methane emissions from the natural gas sector. Large-eddy simulation (LES) has complemented measurement campaigns by serving as a controlled environment in which to study plume dynamics and sampling strategies. However, with few comparisons with controlled-release experiments, the accuracy of LES for modeling natural gas emissions is poorly characterized. In this paper, we evaluate LES from the Weather Research and Forecasting (WRF) Model against Project Prairie Grass campaign measurements and surface layer similarity theory. Using WRF-LES, we simulate continuous emissions from 30 near-surface trace gas sources in two stability regimes: strong convection and weak convection. We examine the impact of grid resolutions ranging from 6.25 to 52 m in the horizontal dimension on model results. We evaluate performance in a statistical framework, calculating fractional bias and conducting Welch's *t* tests. WRF-LES accurately simulates observed surface concentrations at 100 m and beyond under strong convection; simulated concentrations pass *t* tests in this region irrespective of grid resolution. However, in weakly convective conditions with strong winds, WRF-LES substantially overpredicts concentrations—the magnitude of fractional bias often exceeds 30%, and all but one *t* test fails. The good performance of WRF-LES under strong convection correlates with agreement with local free convection theory and a minimal amount of parameterized turbulent kinetic energy. The poor performance under weak convection corresponds to misalignment with Monin–Obukhov similarity theory and a significant amount of parameterized turbulent kinetic energy.

KEYWORDS: Surface layer; Dispersion; Large-eddy simulations; Tracers

1. Introduction

Natural gas production within the United States has surged in the past decade, increasing by more than 50% since 2010 (EIA 2020). Significant emissions from routine operations (e.g., Thorpe et al. 2020), malfunctioning equipment (e.g., Conley et al. 2016), and abandoned wells (e.g., Kang et al. 2016) have spurred the development of new methane emission monitoring instruments and platforms, including satellites, piloted aircraft, unmanned aircraft, open-path lasers, and ground-based point sensors (Fox et al. 2019). Source estimation techniques (SETs) are used to interpret source characteristics (e.g., emission rate) from the trace gas concentration measurements collected via these systems (Harper et al. 2011). Operational source estimation techniques (OSETs) are computationally low-cost and simple to use, and they vary from instrument to instrument. Satellites and remote sensing aircraft often use the integrated mass enhancement technique (Frankenberg et al. 2016; Varon et al. 2018; Jongaramrungruang et al. 2019). In situ aircraft measurements often use mass balance techniques (Karion et al. 2013; Conley et al. 2017). Many ground-based sensors employ techniques that rely on a transport and dispersion model, such as the Gaussian Plume Model (Pasquill 1972; U.S. EPA 2014; Coburn et al. 2018).

To build trust, OSETs are often tested and calibrated against measurements in the field. Of all the common OSETs used to quantify natural gas emissions, approaches based on the Gaussian Plume Model have been the most extensively tested against measurements. The Gaussian Plume Model has been evaluated and calibrated against hundreds of controlled releases through studies such as Project Prairie Grass (PPG) (Barad 1958) and the EPA OTM 33A evaluation study (U.S. EPA 2014). These studies have yielded better understanding of the accuracy and limitations of the Gaussian Plume Model for studying emissions from the natural gas sector. However, OSET evaluation studies that are based on measurements come with limitations, as they quantify performance in the specific conditions that are encountered in the field (e.g., atmospheric stability, terrain). For example, the OTM 33A evaluation study characterized performance in relatively flat terrain, but the technique has since been applied in hilly terrain (e.g., Caulton et al. 2019). Additionally, OSETs that rely on measurements from aircraft and satellites have been evaluated against fewer measurements. These techniques are newer, and it can be more expensive and logistically complicated to make measurements of controlled releases with these instruments. As a result, aircraft- and satellite-based OSETs have relied more heavily on synthetic observations from models, namely large-eddy simulation (LES). Overall, as new methods are developed to quantify methane emissions from the natural gas sector, it is critical to ensure that their corresponding OSETs are accurate.

Corresponding author: Alex Rybchuk, alex.rybchuk@colorado.edu

DOI: 10.1175/MWR-D-20-0233.1

Interest has recently grown in using LES as a tool for studying natural gas emissions. LES is a computational approach that numerically solves the volume-averaged Navier–Stokes equations for flow at large scales and parameterizes small-scale flow with subgrid-scale models. It has been extensively applied in studies of the atmospheric boundary layer (ABL) (Deardorff 1972; Moeng 1984; Mason 1994; Beare et al. 2006; Stoll et al. 2020). LES has been used as part of emission quantification studies to improve measurement strategies (Conley et al. 2017), evaluate and improve OSETs (Taylor et al. 2016; Varon et al. 2018), test new OSETs and their assumptions (Conley et al. 2017; Jongaramrungruang et al. 2019), generate realistic synthetic measurements of methane (Saide et al. 2018), and act as a transport model for field campaign observations (Caulton et al. 2018). LES is computationally expensive but offers several advantages over simpler transport and dispersion models. LES models the dynamic behavior of plumes as driven by time-varying winds, thereby circumventing the need to assume time-averaged fields or steady-state behavior, two assumptions employed in many simpler models. LES provides meteorological and concentration fields at all time steps and locations within a domain, whereas observations provide only a subset of these fields. LES can be used to study plume dynamics under desired atmospheric forcing, and furthermore, LES can simulate complicated physics encountered at real-world natural gas facilities, such as complex terrain (Lundquist et al. 2012; Xue et al. 2018) and time-varying emissions (Saide et al. 2018); therefore, in principle, LES could be used to accurately test OSETs or measurement strategies under a wide variety of environmental conditions.

Unfortunately, LES of atmospheric trace gas dispersion has been statistically evaluated against relatively few experimental measurements (Steinfeld et al. 2008; Ardeshiri et al. 2020), and thus its accuracy for emission quantification studies is not extensively characterized. The most well-known comparison studies focus on the strongly convective ABL in flat terrain. Convective tank studies first done by Willis and Deardorff (1976) and improved upon by Weil et al. (2002) provided a controlled environment to study tracer dispersion in strong convection. In addition, the Convective Diffusion Observed by Remote Sensors (CONDORS) study (Eberhard et al. 1988) released tracers into a real convective ABL. Subsequent LES studies have found good agreement with both sets of measurements in the mixed layer (Lamb 1978; Nieuwstadt and de Valk 1987; Weil et al. 2004, 2012; Nottrott et al. 2014).

LES evaluation studies that examine atmospheric dispersion in the surface layer (less than approximately 100 m above ground level) have often found worse performance. For example, Weil et al. (2012) compared surface concentrations in the atmospheric surface layer from an LES-driven Lagrangian particle dispersion model with observations from the PPG field campaign. The study found good agreement between the two beyond approximately 500 m downwind of the source, but 50 m downwind LES underpredicted concentrations by as much as a factor of 2. Other studies suggest that LES dispersion fails to produce expected behavior when forced by conditions other than strong convection. In one neutral boundary layer, LES underpredicted horizontal trace gas dispersion (Nottrott et al. 2014). In a neutrally stratified field campaign study with multiple controlled releases, LES tended to overpredict emissions (Caulton et al. 2018). Taken together, these two studies suggest that LES of the neutral ABL tends to produce weaker near-surface horizontal winds than would be expected, leading to larger concentrations. As many methane monitoring technologies measure within the atmospheric surface layer and in a range of atmospheric stabilities, it is key to understand the performance of LES in these scenarios.

In this paper, we evaluate the performance of LES from the Weather Research and Forecasting Model (WRF-LES) in the atmospheric surface layer under two types of forcing: strong convection and weak convection. We compare simulated surface concentrations from WRF-LES with data from the PPG field campaign, 50–800 m downwind of a passive tracer source. We assess the impact of LES grid resolution on plumes. In addition, we compare with surface layer similarity (SLS), a transport and dispersion model that is often employed in ground-based OSETs. Recognizing the importance of stochastic uncertainty caused by turbulence (Rao 2005), we evaluate performance in a statistical framework (Chang and Hanna 2004) and simulate a 30-member ensemble of plumes. In doing so, we aim to better understand the accuracy of WRF-LES under simple but realistic methane emission scenarios.

In section 2, we describe the WRF-LES dispersion simulations, the PPG field campaign, surface layer similarity theory, and the statistical metrics used in this study. In section 3, we evaluate the performance of WRF-LES in a strongly convective boundary layer, and we find good agreement with both measurements as well as SLS theory. In section 4, we find that WRF-LES performance suffers in a weakly convective boundary layer. In section 5, we explore factors that impact performance in the atmospheric surface layer, turning to Monin–Obukhov similarity theory (MOST), local-free-convection (LFC) theory, and parameterizations of turbulent kinetic energy. In section 6, we offer conclusions that are based on the study findings.

2. Methods

a. *WRF-LES simulations*

We evaluate the performance of the LES code from the Advanced Research version of WRF (WRF-ARW, version 4.1.2) (Skamarock et al. 2019). WRF-ARW is a numerical weather prediction code that uses the finite difference method to solve the compressible, nonhydrostatic Euler equations on a mass-based grid. It is a popular community-driven code with more than 36 000 registered users, and it serves as the foundation for several additional codes with applications ranging from fire modeling (WRF-FIRE) to renewable energy modeling (Powers et al. 2017).

To evaluate the performance of WRF-LES, we simulate dispersion in the atmospheric boundary layer with eight different configurations (Table 1). We model two types of convection—a strongly convective boundary layer (SCBL) and a weakly convective boundary layer (WCBL). We simulate each

TABLE 1. Key input parameters and observed values for each simulation.

Case	SCBL_C	SCBL_M	SCBL_FS	SCBL_FE	WCBL_C	WCBL_M	WCBL_FS	WCBL_FE
Qualitative description	Coarse	Moderate	Fine, surface	Fine, elevated	Coarse	Moderate	Fine, surface	Fine, elevated
Domain size (Lx, Ly, Lz) (km)	(5, 5, 2)	(5, 5, 2)	(5, 5, 2)	(5, 5, 2)	(3, 3, 1)	(3, 3, 1)	(3, 3, 1)	(3, 3, 1)
Cell count (Nx, Ny, Nz)	(96, 96, 96)	(192, 192, 192)	(500, 500, 200)	(500, 500, 200)	(96, 96, 96)	(192, 192, 192)	(500, 500, 160)	(500, 500, 160)
Horizontal resolution (m)	52	26	10	10	31.25	15.625	6.25	6.25
First cell height (m)	20.8	10.4	3	3	10.4	5.2	3	3
Emission height (m)	10.4	5.2	1.5	10.1	5.2	2.6	1.5	7.3
Geostrophic wind Ug, Vg (m s^{-1})	3.6	3.6	3.6	3.6	10	10	10	10
Surface heating (km s^{-1})	0.24	0.24	0.24	0.24	0.1	0.1	0.1	0.1
Obukhov length (m)	−5.7	−5.9	−6.3	−6.3	−20.4	−20.8	−19.5	−19.5
Friction velocity (m s^{-1})	0.29	0.29	0.29	0.29	0.39	0.40	0.39	0.39
Bottom of capping inversion (m)	1050	1050	1050	1050	525	525	525	525

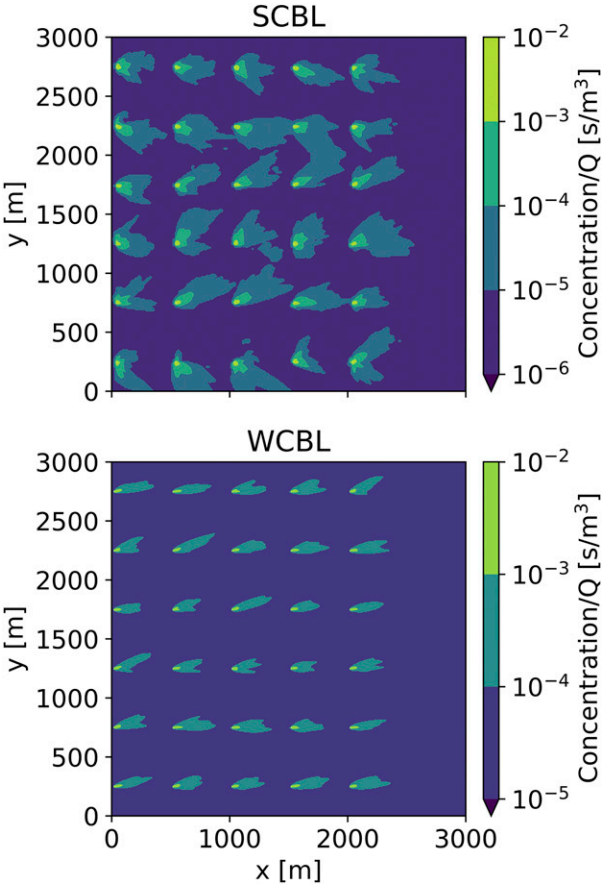

FIG. 1. Grid of normalized 10-min-averaged plume concentrations at 1.5 m within the fine (top) SCBL and (bottom) WCBL.

with a coarse-, a moderate-, and two fine-resolution grids. All cases incorporate flat terrain, cyclic boundary conditions for meteorological fields, a surface roughness of $z_0 = 0.008$ m, and homogeneous surface heating. The specification of z_0 is critical, as in a set of precursor simulations, we found that setting z_0 to 0.1 m produced modeled winds that were substantially weaker than observed winds. Simulations are run without moisture, radiation, microphysics, or other parameterizations commonly employed in mesoscale WRF runs. The simulations in this study use third-order Runge–Kutta to step forward in time, as well as fifth-order horizontal advection and third-order vertical advection. The nonlinear backscatter and anisotropy (NBA) turbulence model captures subgrid effects (Kosović 1997; Mirocha et al. 2010), and MOST provides the lower boundary condition via the MM5 surface layer model (Jiménez et al. 2012).

Both the SCBL and WCBL spin up for two model hours, after which WRF begins to save the fields of interest. The SCBL is forced with constant 3.6 m s^{-1} geostrophic winds, 0.24 W K^{-1} m^{-1} surface heat flux, a 1×10^{-4} s^{-1} Coriolis parameter, and a 0.1-s time step. The SCBL horizontal grid resolutions are $\Delta x = 52$, 26, and 10 m for the coarse, moderate, and fine simulations, respectively. These forcings and the coarse grid resolution are consistent with Weil et al. (2012).

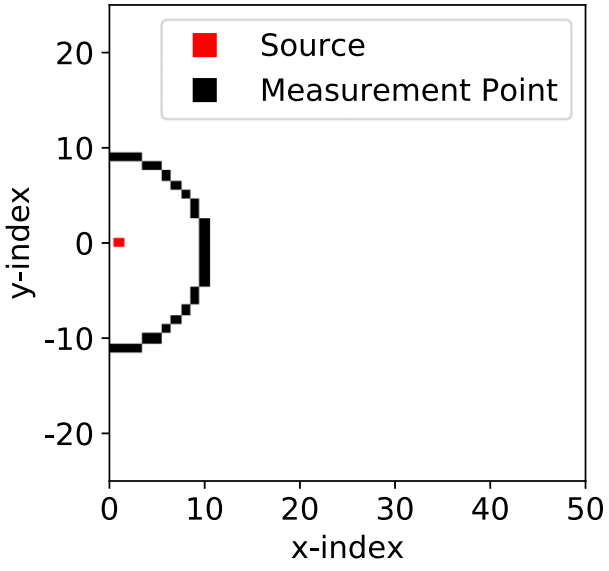

FIG. 2. Measurement points in the fine SCBL ($\Delta x = \Delta y = 10$ m) for a radius of 100 m.

The WCBL is forced with constant 10 m s^{-1} geostrophic winds, 0.1 W K^{-1} m^{-1} surface heat flux, a 1×10^{-4} s^{-1} Coriolis parameter, and a 0.05-s time step. The WCBL horizontal grid resolutions are 31.25, 15.625, and 6.25 m for the coarse, moderate, and fine simulations, respectively. All coarse and moderate simulations use constant vertical grid spacing, respectively 21 and 10.5 m in the SCBL and 10.5 and 5.25 m in the WCBL. The fine-resolution simulations use vertical grid resolutions that change. In the fine SCBL and WCBL simulations, the height of the first grid cell is $z_1 = 3$ m, and concentrations are output at midcell height at 1.5 m. The near-surface grid cells stretch at a rate of 3% until $\Delta z = 10$ or 6.25 m is reached for the fine SCBL and WCBL, respectively. Cells stretch again above the capping inversion at 3%, enabling higher resolution in the region area of interest at reduced computational expense.

To address the highly stochastic nature of dispersion in the turbulent ABL, continuous emissions are simulated from 30 different surface sources in a grid with 500-m spacing, as in Weil et al. (2012) (Fig. 1). Each source experiences different local winds, so that each plume evolves somewhat independently, circumventing the need for an ensemble of simulations for a single set of conditions. Each plume is tagged so that concentrations from one source are distinguishable from the other sources. Plumes are emitted at the midpoint of the lowest grid cells as in Nunalee et al. (2014) in the coarse (''_C''), moderate (''_M''), and fine-surface (''_FS'') simulations. Dispersion is modeled in an Eulerian framework, and, as a result, the height of the emission source decreases as grid resolution is increased. To study this effect, the fine-elevated simulations (''_FE'') emit plumes at the same height as the coarse simulations—approximately 10 m in the SCBL and 5 m in the WCBL. To nullify the impact of recirculating plumes resulting from periodic boundary conditions, we include a trace gas absorbing plane 500 m upwind of each source. Plume mass was confirmed to be conserved in the region between plume source and plume sink.

TABLE 2. PPG observations used in this study.

SCBL			WCBL		
Run	u_* (m s^{-1})	L (m)	Run	u_* (m s^{-1})	L (m)
15	0.22	−6.6	5	0.37	−29
16	0.23	−3.3	9	0.43	−34
25	0.19	−5.4	19	0.37	−25
47	0.22	−5.3	26	0.41	−28
48S	0.21	−5.2	27	0.39	−32
			43	0.33	−13
			44	0.38	−25
			49	0.43	−25
			50	0.42	−31
			51	0.43	−33
			61	0.49	−33
			62	0.32	−27

After a 2-h spinup, we sample trace gas fields and winds every second during a 10-min period, matching the PPG measurement period. To focus on average plume behavior and minimize errors that stem from using point-concentration measurements, we compare crosswind integrated concentrations (CWIC) (Fig. 2), as is common in studies that employ PPG measurements (e.g., van Ulden 1978; Weil et al. 2012). CWIC is calculated at a given radius as

$$\mathrm{CWIC} = \Delta s\left(\sum_i C_i\right), \tag{1}$$

where C_i is the concentration at a cell i and Δs is the arclength between cells. To account for the different release rates used in PPG, CWIC calculations throughout this study are normalized by emission rate Q, and this quantity is referred to as ''concentration'' although in strict terms it is a ''normalized crosswind integrated concentration.'' To compare the medium and coarse simulations to the PPG horizontal array measurements collected at a height of 1.5 m, concentration profiles are extrapolated downward using a fifth-order polynomial fit to concentrations in the lowest 100 m. For each simulated emission source, we calculate 10-min-averaged CWIC at 50, 100, 200, 400, and 800 m downwind.

b. *Project Prairie Grass*

The PPG field campaign was conducted in 1956 in Kansas to study the near-surface behavior of passive tracer plumes during various meteorological conditions (Barad 1958). This campaign serves as a cornerstone for atmospheric dispersion models, informing key parameters in the Pasquill–Gifford stability classes for the Gaussian Plume Model (Venkatram 1996) and acting as a validation dataset for many regulatory dispersion models such as AERMOD (Cimorelli et al. 2005). Seventy controlled releases of SO_2 were carried out: six at 1.5 m above ground level and the remainder at 0.48 m. For each controlled release, 10-min average concentration measurements were collected at an array of 599 individual sampling points. Measurements were conducted in concentric arcs 50,

TABLE 3. Average performance of SLS theory relative to PPG observations.

	SCBL			WCBL		
	FAC2 (%)	FB (%)	NMSE (%)	FAC2 (%)	FB (%)	NMSE (%)
50 m	100	3	2	95	7	1
100 m	100	6	5	100	30	10
200 m	100	3	10	100	18	6
400 m	100	−2	7	92	20	7
800 m	60	18	3	83	23	14

100, 200, 400, and 800 m downwind of the release source. Along each arc, a horizontal array of point measurements was gathered at a height of 1.5 m, spaced 1° apart at 800 m and 2° at all other downwind distances. A vertical array of measurements was also collected 100 m downwind at heights 0.5, 1.0, 1.5, 2.5, 4.5, 7.5, 10.5, 13.5, and 17.5 m using six towers. The overall concentration uncertainties were reported as 1%–2%. The roughness length of the site was estimated to be $z_0 = 0.008$ m (Sawford 2001). The winds employed in this study were measured with a cup anemometer 25 m west of the release source at a height of 2 m during a 10-min period. Obukhov lengths L and friction velocities u_* were not directly measured during the campaign but were estimated from tower measurements in subsequent analysis (Horst et al. 1979). Normalized CWIC for the horizontal array is taken from Horst et al. (1979), and normalized CWIC for the vertical array is calculated using digitized data (http://www.harmo.org/jsirwin).

Measurements from a number of runs are either excluded in this analysis or are not available. The runs used here are listed in Table 2. Concentration measurements not reported for run 63 and run 64 because of "extremely light and variable winds." Vertical tower measurements were gathered only for run 13 and beyond and were additionally not reported for runs 23, 28, 35, 53, 63, and 64. Furthermore, profile measurements were excluded if wind direction was misaligned with the vertical array and fewer than three towers captured measurements; thus, fewer vertical profiles are available for comparison. Winds speeds were not reported for run 3 and run 6, so those runs are excluded from this analysis.

We aim to compare as many observations with WRF-LES concentration simulations as possible. In principle, this comparison would best be achieved by running one simulation for each controlled release, because each release occurs in the presence of a different L and u_*; however, running one high-resolution simulation for each observation would be prohibitively expensive. As an alternative, we assess the performance of WRF-LES by binning PPG runs with similar atmospheric conditions into strongly convective and weakly convective categories. One common method to bin data in atmospheric dispersion studies is the Pasquill–Gifford stability classes (De Visscher 2013). These classes are traditionally delineated using wind speeds and solar radiation, but they can be alternatively delineated using a roughness length and Obukhov length (Golder 1972). Class A corresponds to $0 \geq L \geq -7$ m for the PPG roughness length. This range is used to bin PPG data for comparison with the SCBL LES runs, which have L between −6.3 and −5.7 m. The WCBL LES runs have L between −20.4 and −19.5 m, which falls on the border between class B stability ($-7 \geq L \geq -15$ m) and class C stability ($-15 \geq L \geq -50$ m). Accordingly, we use intermediate values of the PPG runs, $-10 \geq L \geq -35$ m, for the LES WCBL comparison bin. To more closely resemble the WCBL LES, we additionally require $u_* \geq 0.3\ \mathrm{m\ s^{-1}}$.

c. Surface layer similarity theory

SLS theory (van Ulden 1978) is used to complement the PPG observations. Each observation has a different pair of u_* and L values, and none of these pairs precisely match the conditions in the LES; however, SLS theory can be used to calculate approximate CWIC under any desired u_* and L conditions. Normalized CWIC at a height z is calculated for the PPG runs as

$$\frac{\mathrm{CWIC}(z)}{Q} = \frac{0.73}{\overline{u}\,\overline{z}} \exp\left[-\left(\frac{0.66z}{\overline{z}}\right)^{1.5}\right], \tag{2}$$

where Q is the emission rate, $\overline{z}$ is the plume centerline height, and $\overline{u}$ is the wind speed at the plume centerline. The values of $\overline{z}$ and $\overline{u}$ are numerically computed on the basis of MOST, and downwind distance x is implicitly a function of these variables. SLS theory is strictly valid for releases at a height of 0 m, but it generally agrees well with the observations in this study (Tables 3–5). In the SCBL, SLS-modeled concentrations typically have a fractional bias magnitude (section 2d) of less than 30%, although fractional bias can be as large as 115%. In the WCBL, the majority of modeled concentrations also fall within a fractional bias magnitude of 30%, although fractional bias may be as large as 100%. Considering the overall good agreement between SLS theory and observations, we use SLS theory as an imperfect proxy for hypothetical observations, with u_* and L that match those of the LES. To calculate SLS concentration fields that correspond to LES of the SCBL and the WCBL, we use u_* and L values from the fine resolution simulations.

TABLE 4. Fractional bias between SLS theory and individual PPG observations in the SCBL.

Run	50 m	100 m	200 m	400 m	800 m
15	−0.08	−0.19	−0.35	−0.28	−0.38
16	−0.14	−0.14	−0.05	0.22	0.67
25	−0.04	0.14	0.27	−0.05	0.20
47	0.10	0.10	0.04	−0.13	0.19
71	0.32	0.42	0.43	0.48	1.15

TABLE 5. Fractional bias between SLS theory and individual PPG observations in the WCBL.

Run	50 m	100 m	200 m	400 m	800 m
5	0.13	0.41	0.34	0.40	0.49
9	0.07	0.29	0.19	0.17	0.24
19	0.10	0.40	0.50	0.67	1.00
26	0.12	0.30	0.20	0.22	0.27
27	0.08	0.33	0.20	0.20	0.10
43	0.03	0.15	−0.02	−0.04	−0.09
44	0.07	0.33	0.23	0.15	0.20
49	−0.01	0.20	0.05	0.02	−0.02
50	0.07	0.32	0.38	0.31	0.47
51	−0.07	0.22	0.28	0.35	0.77
61	0.09	0.27	0.02	−0.12	−0.20
62	0.15	0.37	−0.12	0.10	−0.08

d. Statistical metrics

Chang and Hanna (2004) summarize metrics for evaluating dispersion models by comparing an observation C_o with a model prediction C_p. Although there is no one optimal metric, they conclude that "good performing models" have predictions that fall within a factor of 2 of observations (FAC2) at least 50% of the time, that the relative mean bias [here, fractional bias (FB)] is less than 30%, and that the relative scatter [here, normalized mean square error (NMSE)] is less than approximately a factor of 2. FAC2 is calculated as the fraction of data within $0.5 \le C_p/C_o \le 2.0$. Fractional bias is calculated as

$$\mathrm{FB} = \frac{2(\overline{C}_o - \overline{C}_p)}{\overline{C}_o + \overline{C}_p}, \tag{3}$$

where averages are taken over the set of measurements or simulations. NMSE is calculated as

$$\mathrm{NMSE} = \frac{\overline{(C_o - C_p)^2}}{\overline{C}_o \overline{C}_p}. \tag{4}$$

Here, only observations for the horizontal array are used for quantitative comparison, because the vertical array stability bins have relatively few observations.

SLS model performance is compared with observations using FAC2, FB, and NMSE (Tables 3–5). In contrast, FAC2 and NMSE are not calculated for LES because these metrics require each observation to be paired with a model prediction. Nonetheless, Chang and Hanna (2004) emphasize the importance of assessing multiple metrics when comparing models to observations. As such, we use the Welch's t test to compare the LES distribution and the observed distribution of concentration. Both distributions are assumed to be Gaussian at each downwind location. The null hypothesis is that the mean concentrations for these distributions are identical, and the test is conducted at the 95% confidence interval with a two-sided tail. Mean LES concentrations are also evaluated using FB. However, for the FB comparison, SLS theory serves as comparison—instead of observations—to minimize error stemming from differences in L and u_*. Seeking to characterize the accuracy of only surface emission sources, we omit statistical comparisons for the LES runs that emit above the lowest grid height.

3. Evaluation of LES in the strongly convective boundary layer

a. Horizontal surface concentrations

As LES grid resolution is refined under strongly convective conditions, mean surface concentrations approximately converge in the far field but diverge closer to the source (Fig. 3). Downwind of 200 m, mean surface concentrations from the surface-release simulations (SCBL_C, SCBL_M, SCBL_FS) as well as the elevated-release simulation (SCBL_FE) nearly collapse onto the same line. Closer to the source, 50-m surface concentrations increase as the grid is refined. To investigate the mechanism that causes this discrepancy, we examine the impact of source height.

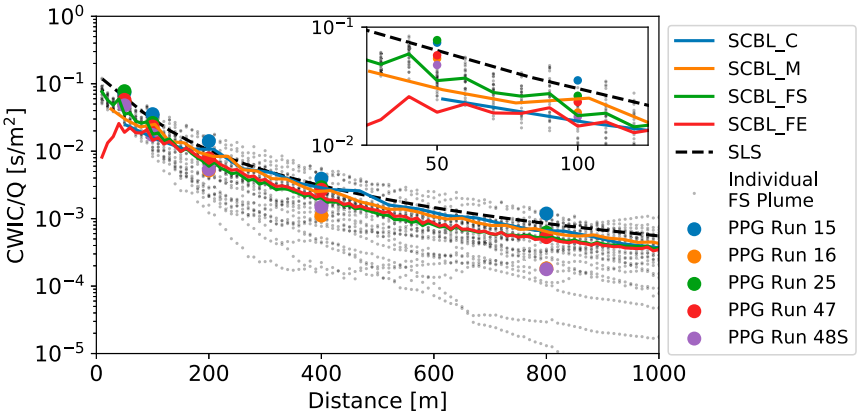

FIG. 3. SCBL observations and model predictions for the horizontal array. Ensemble average LES concentrations are shown as solid lines. SLS concentrations are calculated using u_* and L from SCBL_FS.

TABLE 6. LES performance in the SCBL. "Reject" denotes that the null hypothesis has been rejected.

	SCBL_C		SCBL_M		SCBL_FS	
	FB (%)	*t* test	FB (%)	*t* test	FB (%)	*t* test
50 m	100	Reject	84	Reject	71	Reject
100 m	64	Not reject	24	Not reject	56	Not reject
200 m	−3	Not reject	17	Not reject	54	Not reject
400 m	−7	Not reject	19	Not reject	41	Not reject
800 m	24	Not reject	27	Not reject	49	Not reject

As grid resolution is vertically refined, the height of the "surface" grid cell changes, thereby modifying the emission height. This height change produces differences in surface concentration that are significant close to the source but are insignificant farther downwind. Downwind of 200 m, mean surface concentrations from SCBL_FS (which emits at 1.5 m) and SCBL_FE (which emits at 10 m) overlap. However, between 10 and 200 m downwind of the source, mean surface concentrations diverge. Thus, it appears that the near-source concentration increase across all simulations primarily emerges from a change in source height. Therefore, when studying near-field trace concentrations through Eulerian dispersion simulations, it is critical to have a vertical grid fine enough that matches the desired emission height. This constraint can be avoided through the use of Lagrangian simulations (e.g., Weil et al. 2012).

Under strongly convective conditions, LES agrees moderately well with SLS theory, although agreement varies by resolution (Table 6). The moderate resolution simulation has the best agreement overall; while 50-m fractional bias is 83%, fractional bias is below 30% [the "good" performance threshold from Chang and Hanna (2004)] for all other downwind distances. Similarly, fractional bias in the coarse simulation fails to meet the 30% threshold at 50 and 100 m downwind, but it meets that threshold at all other downwind distances. Interestingly, the fine resolution, surface-release simulation fairs the worst when compared to SLS theory. This simulation fails to meet the 30% threshold at any downwind distance; fractional bias is approximately 50% at most downwind locations. While this is larger than the 30% threshold, these fractional bias values are approximately consistent downwind of 50 m. This offset points to a bias that likely arises from a mismatch in friction velocity and Obukhov length. We underscore that (just like LES) SLS theory is a model, and as such, it is an imperfect representation of real-world plume dynamics. To complement fractional bias calculations conduced with SLS theory, we additionally examine the results of *t* tests.

LES agrees well with observations in strongly convective conditions. We conduct a Welch's *t* test at each downwind location is used to assess whether the average LES and average measured concentrations differ significantly (Table 6). All three simulations perform equally well, irrespective of resolution. The distribution of simulated concentrations is statistically indistinct from measured concentrations everywhere aside from at 50 m.

It is crucial that these comparisons are rooted in a statistical framework, because LES ensemble members in the SCBL display a significant amount of scatter. At 200 m and beyond, the minimum and maximum concentrations differ by more than an order of magnitude. This scatter occurs even though all plumes are subject to the same geostrophic winds and surface heating. In the SCBL, individual plume behavior is strongly governed by the local presence of updrafts and downdrafts. To ensure that 30 members were sufficient to characterize surface concentrations, we conduct a bootstrap analysis to estimate confidence in mean concentrations as a function of ensemble size (Wilks, 2019). Bootstrap samples are generated with replacement for $n = 1, 3, \ldots, 30$ members using plumes from SCBL_FS. This process is repeated a large number of times for each sample size ($n_B = 1000$) and the coefficient of variance σ_B/μ (standard deviation divided by mean) of each bootstrap dataset is calculated. Uncertainty is reported as a function of bootstrap sample size and downwind distance (Fig. 4), and it is less than 10% for the 30-member ensemble at all downwind distances. As such, we find that 30 members are sufficient to

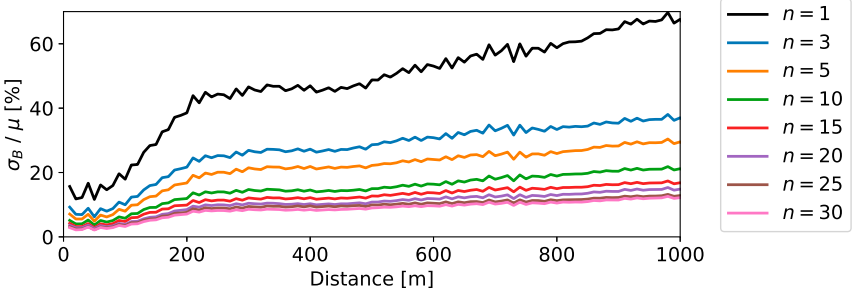

FIG. 4. Bootstrap analysis in the SCBL.

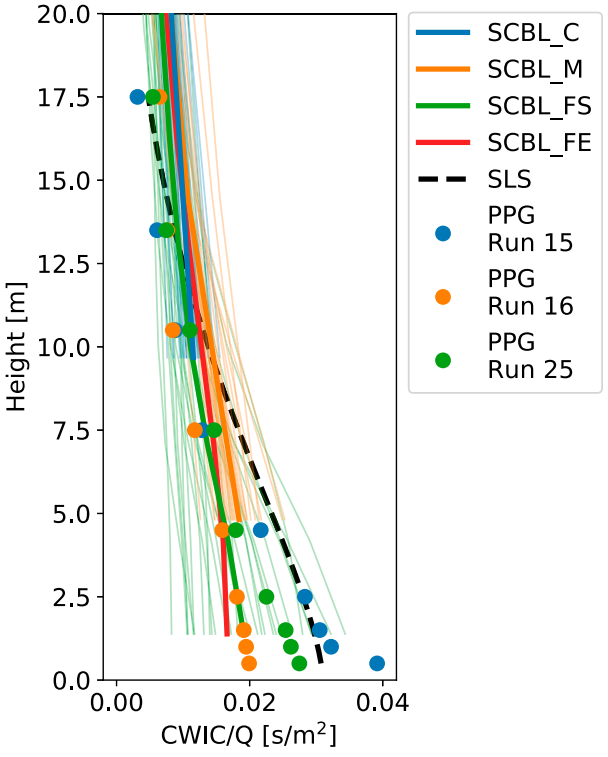

FIG. 5. SCBL observations and model predictions for the vertical array at 100-m downwind distance. Ensemble average LES concentrations are shown as solid lines, and individual plumes are shown as thin lines. SLS concentrations are calculated using u_* and L from SCBL_FS.

smooth the highly stochastic values of concentration from individual plumes.

b. *Vertical concentration profiles*

As with the horizontal array, we find that WRF-LES performs well against vertical profiles of concentration (Fig. 5). The average concentrations across all resolutions match at heights above 10 m, the height of the lowest concentration measurement from the coarse simulation. Nearly all measured concentrations fall within the 30-member spread of SCBL_FS. Additionally, the mean profile of SCBL_FS overlaps measured concentrations from Run 16. In the SCBL, LES quantitatively performs well against surface concentrations through fractional bias comparisons and t tests, and it qualitatively performs well at 100 m. Thus, in conjunction with the analysis of the horizontal array, we conclude that WRF-LES accurately models realistic plume dynamics in the SCBL.

4. Evaluation of LES in the weakly convective boundary layer

a. *Horizontal surface concentrations*

The simulated mean surface concentrations in the WCBL show similar convergence behavior to those in the SCBL (Fig. 6). Simulated concentrations from different resolutions converge in the far-field, and they diverge closer to the source. The emission heights in WCBL_M and WCBL_FS only differ by approximately 1 m (respectively 2.6 and 1 m), and as such, the near-field concentrations only slightly disagree near the source.

Unlike the SCBL, the LES simulations of the WCBL perform poorly relative to SLS theory and observations in the horizontal dimension (Table 7). Across all resolutions, most comparisons show |FB| $>$ 30%, which is outside the "good" performance threshold from Chang and Hanna (2004). While the magnitude of fractional bias at 100 or 200 m may be less than 30%, this downwind distance is simply the crossover point where LES transitions from overprediction to underprediction. In addition, every t-test comparison aside from the 200-m moderate resolution case fails. This single success case is dismissed as coincidental based on the results of the other t tests. Based on the different concentration decay rates, LES plumes in the WCBL are mixing more weakly than observed plumes.

b. *Vertical concentration profiles*

Surprisingly, the WCBL appears to perform somewhat accurately for the vertical array of measurements (Fig. 7). However, the surface concentration comparison reveals that

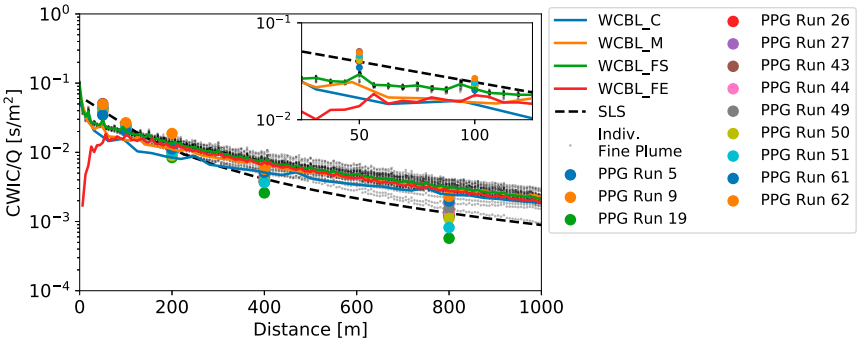

FIG. 6. WCBL observations and model predictions for the horizontal array. Ensemble average LES concentrations are shown as solid lines. SLS concentrations are calculated using u_* and L from WCBL_FS.

TABLE 7. LES performance in the WCBL.

	WCBL_C		WCBL_M		WCBL_FS	
	FB (%)	*t* test	FB (%)	*t* test	FB (%)	*t* test
50 m	90	Reject	65	Reject	41	Reject
100 m	65	Reject	62	Reject	32	Reject
200 m	33	Reject	5	Not reject	−13	Reject
400 m	−20	Reject	−40	Reject	−55	Reject
800 m	−60	Reject	−78	Reject	−80	Reject

this performance is coincidental, and it does not reflect that LES accurately models trace dynamics in weakly convective conditions. Unlike in the SCBL, concentration profiles from different resolution simulations fail to overlap above 10 m. This suggests that fields in the WCBL are grid-resolution dependent, and as such, they are unlikely to correspond to real-world dynamics. Correspondingly, LES infrequently overlaps with observations. The best agreement between LES and observations is found at 17.5 m, where both WCBL_FS and observations suggest negligible amounts of tracer. Notably, as resolution is increased, LES profiles tend toward observed and SLS profiles. Ardeshiri et al. (2020) found that resolution increases for LES of dispersion in neutral conditions produced marginal improvements for refinements at coarser resolutions but substantial improvements for refinements of finer grids. Thus, it is possible that additional grid resolution in the WCBL will lead to improved accuracy. But, while trends in WCBL profiles suggest that increased grid resolution may help, trends in WCBL surface concentrations do not suggest that this will be the case. Ardeshiri et al. (2020) also found that neutral LES underestimated crosswind variance σ_v^2 and vertical variance σ_w^2. This correlates with the undermixing observed in the WCBL LES.

5. Discussion

We find that WRF-LES of trace gas dispersion from a surface source performs well in strongly convective conditions but underperforms in weakly convective conditions. In this section, we discuss factors that may drive this behavior.

Near-surface turbulence within the atmospheric surface layer is complicated to model; it is characterized by anisotropy, a small outer length scale, a strong dependence on atmospheric stability, and a ''reverse turbulent cascade'' where small spatial scales transfer energy to larger scales (Sullivan et al. 2003; Klipp 2014). These characteristics make it challenging for LES to accurately model flow in this region, and the inability of our LES to capture all of these features likely drives the mismatched concentrations in the WCBL. Modelers are actively researching methods to improve LES accuracy near solid surfaces. Within the atmospheric surface layer, these techniques include improving subgrid-scale models (Porté-Agel et al. 2000; Bou-Zeid et al. 2005; Chung and Matheou 2014; Mokhtarpoor and Heinz 2017), improving wall models (Maronga et al. 2019), and refining grid size and aspect ratio (Brasseur and Wei 2010; Daniels et al. 2016).

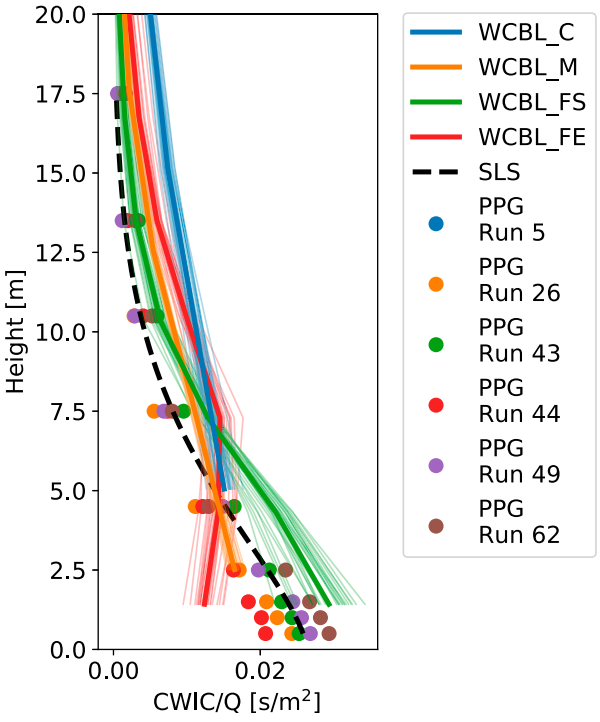

FIG. 7. WCBL observations and model predictions for the vertical array at 100-m downwind distance. Ensemble average LES concentrations are shown as solid lines, and individual plumes are shown as thin lines. SLS concentrations are calculated using u_* and L from WCBL_FS.

One common approach to diagnose the performance of LES in the atmospheric surface layer involves comparing simulated fields to MOST. This theory is derived for flat terrain under homogeneous forcing, as is the case in this LES study, and it has been shown to agree well with observations in these conditions (Businger et al. 1971; Dyer 1974). MOST is only accurate for moderately unstable to moderately stable surface layers, $-2 < z/L < 1$ (Foken 2006). In strict terms, MOST is valid for ensemble-averaged fields, but in many LES codes, including WRF-LES, it is applied to instantaneous fields at individual cells (Maronga et al. 2019). MOST describes wind and temperature profiles in the atmospheric surface layer based on L, a nondimensional wind shear ϕ_m, and a nondimensional temperature gradient ϕ_h (Stull 1988). This nondimensional function takes one of many similar empirical forms (Maronga and Reuder 2017), and it is calculated from either observations or LES as

$$\phi_m\left(\frac{z}{L}\right) = \frac{\partial \overline{u}_h}{\partial z}\frac{\kappa z}{u_*} \quad \text{and} \tag{5}$$

$$\phi_h\left(\frac{z}{L}\right) = \frac{\partial \overline{\theta}}{\partial z}\frac{\kappa z}{\theta_*}, \tag{6}$$

where $\overline{u}_h$ is the mean horizontal wind speed, $\overline{\theta}$ is the average potential temperature, θ_* is the kinematic heat flux divided by friction velocity, and κ is the von Kármán constant, taken to be 0.4. The expression for ϕ_m can be used to calculate wind speed as

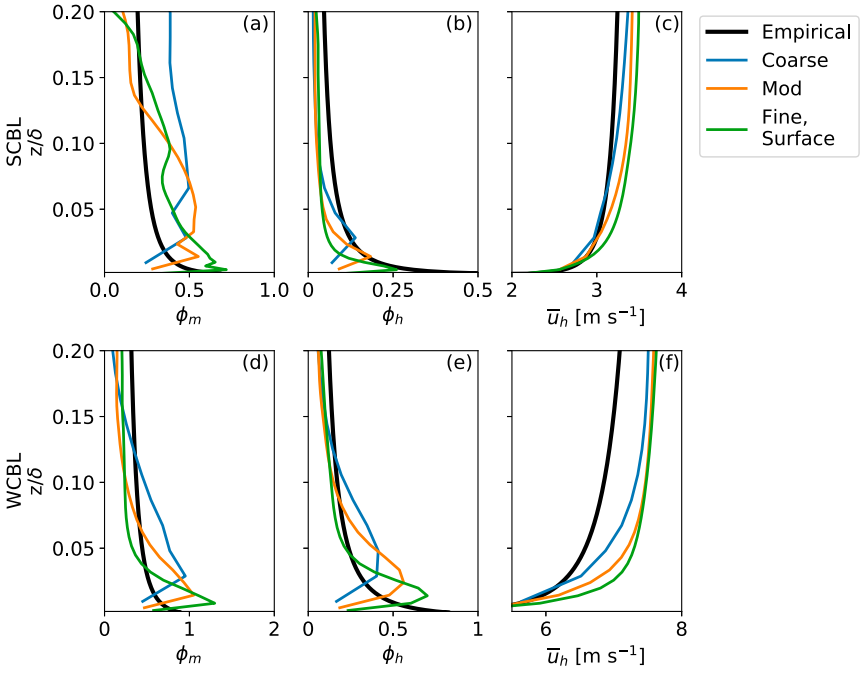

FIG. 8. (a),(d) Nondimensional wind shear ϕ_m; (b),(e) nondimensional temperature gradient ϕ_h; and (c),(f) wind speed profiles computed from LES (colored lines) and empirical fits (black line), scaled by boundary layer depth δ, for profiles corresponding to the (top) SCBL and (bottom) WCBL.

$$\overline{u}_h(z, z_0, L) = \frac{u_*}{\kappa}\left[\ln\frac{z}{z_0} + \Psi_m(z/L) - \Psi_m(z_0/L)\right], \quad (7)$$

where

$$\Psi_m(\zeta) = \int_0^{z/L} [\phi_m(\zeta) - 1]\zeta^{-1}\, d\zeta. \quad (8)$$

We examine the behavior of the WCBL near the ground with reference to MOST. We calculate ϕ_m and ϕ_h directly from WCBL LES wind fields and compare with empirical profiles based on the WCBL_FS values of u_* and L using the recommended parameterization from Foken (2006) (Figs. 8d,e). A large peak ("overshoot") is observed in the LES-based ϕ_m profiles, correlating with excessive shear observed in other studies (e.g., Maronga et al. 2019; Peña and Hahmann 2020) and the height of this overshoot decreases as resolution increases, as in Brasseur and Wei (2010). We additionally observe overshoot in ϕ_h profiles, as in Khanna and Brasseur (1997) for LES driven with Smagorinsky-type SGS models. The overshoot in ϕ_m causes LES winds to be stronger than those predicted by MOST (Fig. 8f). Stronger winds typically enhance tracer mixing. However, far-field LES concentrations overpredict relative to SLS theory and observations, suggesting weaker mixing. Thus, while ϕ_m and ϕ_h overshoot may correlate with inadequate winds in the WCBL surface layer, they do not appear to drive the LES far-field overprediction in the WCBL. The overshoot could instead be expected to produce far-field underprediction.

In the asymptotically unstable state, u_* tends toward 0 and MOST is superseded by local-free-convection (LFC) theory. Statistics of w and θ have been observed to agree well with LFC theory (Wyngaard 2010). In LFC theory, two key scaling parameters for the dry atmosphere are (Maronga and Reuder 2017) as follows:

$$w_{\mathrm{LF}} = \left(\frac{g}{\theta}\overline{w'\theta'}_0 z\right)^{1/3} \quad \text{and} \quad (9)$$

$$\theta_{\mathrm{LF}} = \frac{\overline{w'\theta'_0}}{w_{\mathrm{LF}}}. \quad (10)$$

The dimensionless mean gradient of temperature should follow

$$\frac{\partial\theta}{\partial z}\frac{z}{\theta_{\mathrm{LF}}} = A_\theta, \quad (11)$$

where A_θ is a universal constant that has been observed to be −0.35 (Sorbjan 1986).

We examine the behavior of the SCBL with reference to MOST and LFC theory. In strict terms, MOST should not be applied to the SCBL, as heights above approximately 3 m lie outside the z/L range of applicability for this similarity theory. As such, we provide these profiles simply as a reference. We highlight that overshoot in ϕ_m does not occur in the SCBL, even at the coarsest resolution. However, overshoot in ϕ_h is still observed. We calculate A_θ using 10 instantaneous temperature fields spaced 1 min apart (Fig. 9). Mirroring the

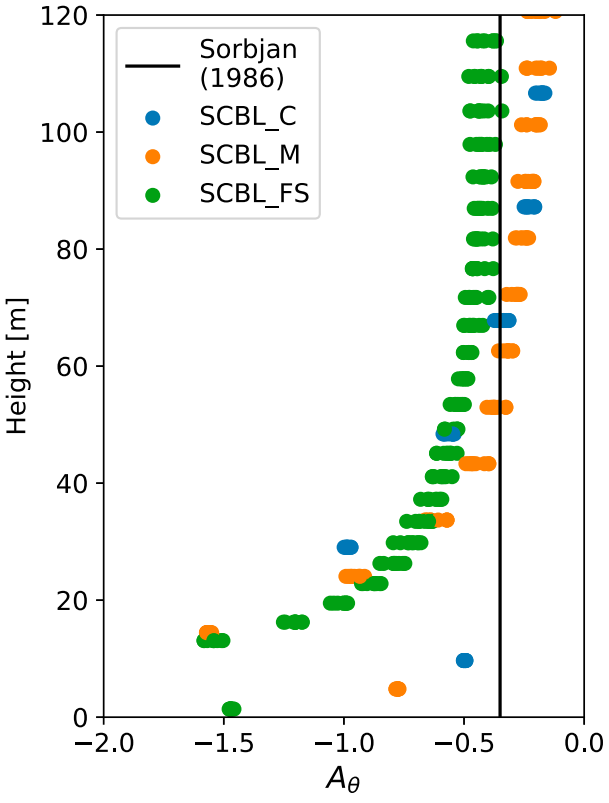

FIG. 9. Profile of the universal LFC constant A_θ.

LFC LES study of Maronga and Reuder (2017), LES values of A_θ approximately equal the observed value of 0.35, especially in the upper part of the surface layer. Maronga and Reuder (2017) exclude the lowest seven vertical levels in their high resolution simulation due to boundary condition concerns (Larsson et al. 2016). We similarly observe poor agreement for the lowest seven vertical levels of SCBL_FS. We additionally highlight a grid dependence, as SCBL_C agrees well with observations by the third vertical level. Overall, LES of the SCBL agrees with observed values of A_θ, correlating with the performance with regard to plume modeling. Interestingly, good LES plume dispersion performance is observed even though simulated values of A_θ tend to disagree with observed values at the lowest vertical levels. This surprising behavior may be explained by examining SGS energy.

Aside from employing MOST and LFC theory, studies in the atmospheric surface layer characterize characterize near-surface flow by quantifying the relative importance of SGS modeling. In LES models, resolved winds near surfaces are often underresolved, as the outer length scale of the flow can be smaller than the LES filter width. SGS models parameterize the impact of underresolved winds everywhere in the flow, but they often struggle to accurately parameterize near-surface turbulence due to its unique attributes (Sullivan et al. 2003). Many grid resolution guidelines have been proposed to ensure accurately modeled flow (e.g., Wurps et al. 2020). Pope (2004) suggest that the resolved turbulent kinetic energy (TKE) should compose at least 80% of the total TKE. We calculate total TKE from instantaneous fields as

$$\overline{e}_{\text{tot}} = \frac{1}{2}\overline{u_i' u_i'} + \frac{1}{2} m_{ii} + e_{\text{SGS,prog}}. \tag{12}$$

The first term corresponds to resolved TKE, whereas the last two terms correspond to SGS TKE. Velocity fluctuations u_i' at a height z are calculated with reference to winds that have been averaged horizontally across the domain. The subgrid TKE from the prognostic TKE equation is $e_{\text{SGS,prog}}$ and m_{ii} is the diagonal NBA model stress (Kosović 1997; Mirocha et al. 2010).

LES performance of plume dispersion correlates with adherence to the Pope guideline (Fig. 10). In the SCBL, resolved TKE composes over 80% of total TKE at all heights. In contrast, resolved TKE composes less than 80% of total TKE at the lowest grid cells for all three resolutions. Approximately 70% of TKE is resolved at the lowest vertical level. Thus, flow at emission height is underresolved in the WCBL. This may lead to the undermixing observed in plumes in the WCBL. This also suggests that a finer grid in the WCBL may lead to better agreement with observations.

6. Conclusions

In this study, we assess the accuracy of WRF-LES for simulating trace gas dispersion from a surface source in strongly convective and weakly convective boundary layers. We compare 30 plumes within each simulation with horizontal and vertical measurements from the Project Prairie Grass campaign (50–800 m downwind of a source, with measurements at heights of 0.5–17.5 m). We also compare WRF-LES simulations to surface layer similarity (SLS) theory. We evaluate the performance of WRF-LES dispersion using a statistical framework, relying on the fractional bias metric and Welch's t tests to compare distributions.

In strongly convective conditions, WRF-LES, the PPG measurements, and the SLS theory tend to agree well. Regardless of grid resolution, WRF-LES concentrations pass t tests relative to observations at 100 m and beyond. When compared with SLS theory, coarse- and moderate-resolution simulations tend to have a fractional bias magnitude of less than 30%. The largest discrepancies occur closest to the source at 50 m. The fine-resolution simulation produces fractional biases that are approximately 50% at 100 m and farther downwind. This persistent offset points to bias that arises from sensitivity in SLS theory to micrometeorological conditions as well as limitations that come with model-to-model comparisons.

In weakly convective conditions, WRF-LES agrees poorly with SLS theory and observations. Regardless of grid resolution, WRF-LES substantially overpredicts concentrations at 50 m and underpredicts concentrations beyond 200 m. WRF-LES agrees moderately well at 100 and 200 m, although this region simply corresponds to the crossover between overprediction and underprediction. This overall poor agreement suggests that plumes from a surface source are undermixed in LES of these atmospheric conditions.

We shed light on factors that drive agreement and disagreement in the atmospheric surface layer by turning to

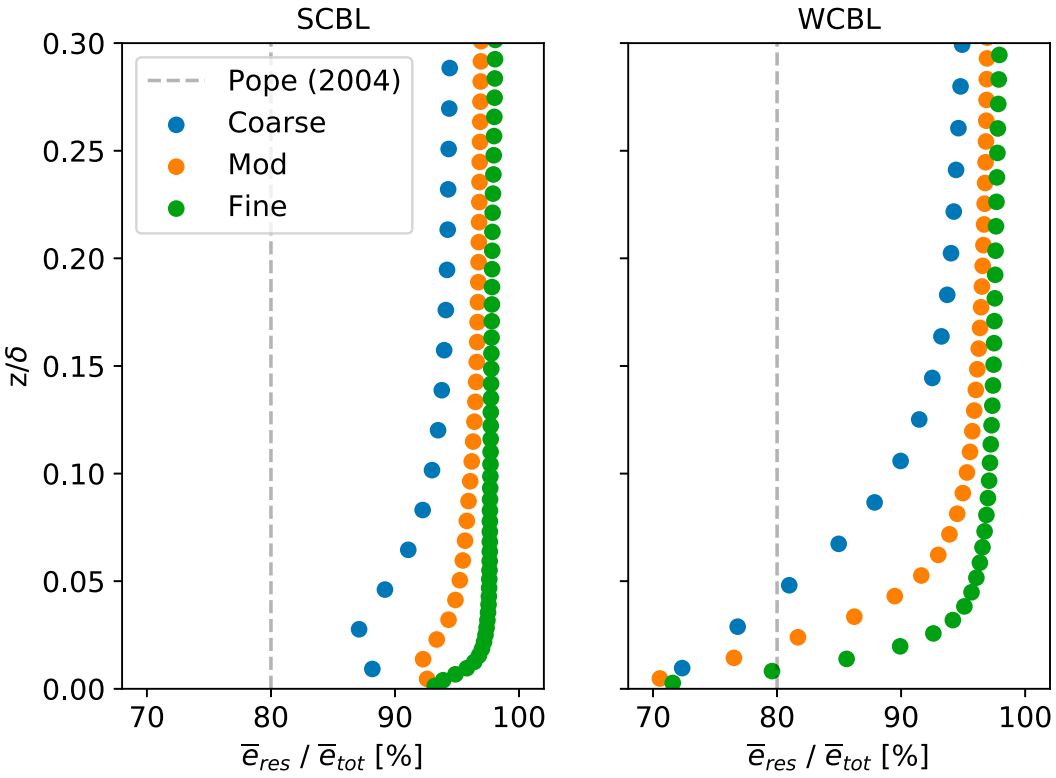

FIG. 10. The ratio of resolved TKE to total TKE, scaled by boundary layer depth δ. The filled circles correspond to vertical grid levels.

Monin–Obukhov similarity theory (MOST), local-free-convection (LFC) theory, and SGS modeling. We find that winds in the weakly convective LES poorly agree with MOST because of the "overshoot" problem (Brasseur and Wei, 2010). As a result, LES predicts stronger winds than would be anticipated. However, stronger winds should lead to overmixing, whereas undermixing is observed in these simulations. Thus, while overshoot is present and may correlate with poor accuracy of dispersion modeling, it does not appear to drive the poor agreement with observed concentrations. In contrast, LES of the strongly convective boundary layer agrees with LFC theory, further suggesting that LES is accurately modeling real-world dynamics. Last, we quantify the relative contribution of explicitly resolved turbulent kinetic energy (TKE) to parameterized TKE. We find that strongly convective simulations meet the criteria suggested by Pope (2004), as resolved TKE constitutes over 80% of total TKE at all heights. However, resolved TKE only constitutes 70% of the total TKE in weakly convective simulations at the surface, likely impacting the surface-height emission sources. There is evidence that additional grid resolution will improve LES agreement with observed concentrations (Ardeshiri et al. 2020). However, there is also evidence to the contrary; grid refinement from coarse to fine resolution did not substantially change surface concentrations in the weakly convective boundary layer in this study.

The results of this study caution that WRF-LES (and atmospheric LES codes in general) should be evaluated in a statistical framework with reference to available empirical datasets when possible. By simulating 30 plumes under identical large-scale forcing, we consider the stochastic nature of turbulent diffusion. At times we observe order-of-magnitude differences in 10-min-averaged concentrations between different plumes. This study examined the simple case of flat terrain and homogeneous forcing, but the conclusions are broadly applicable to studies examining dispersion in more challenging scenarios, such as complex terrain or urban environments.

LES has many unique features that makes it a uniquely useful tool for modeling emissions of trace gases. LES can be (and has been used) to improve measurement strategies for field campaigns. It can simulate dispersion in complex environments, which is valuable as regulators seek to characterize real-world emissions in industrial environments with complex terrain and time-varying emissions. This study also showcases the invaluable theoretical contributions that stem from work on wall-modeled LES and turbulence modeling. Through further theoretical improvements and comparisons with controlled releases, trust in LES dispersion can be fostered, and it can begin to take a more central role in the emission quantification challenge.

Acknowledgments. We effusively thank the anonymous reviewer whose feedback substantially improved the analysis and writing presented in this paper. In addition, we thank Ian Faloona, Branko Kosović, and Jeffery C. Weil for their insight while we prepared this paper. Authors Rybchuk, Alden, and

Rieker were supported by a grant from the U.S. Department of Energy (DOE) Office of Fossil Energy, National Energy Technology Laboratory (DE-FE0029168). Author Lundquist was supported by NSF Grant AGS-1565498. The simulations here were conducted with supercomputing resources from the University of Colorado Boulder Research Computing Group, which is supported by the National Science Foundation (Awards ACI-1532235 and ACI-1532236), the University of Colorado Boulder, and Colorado State University. This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the DOE under Contract DE-AC36-08GO28308. Funding was provided by the DOE Office of Energy Efficiency and Renewable Energy Wind Energy Technologies Office. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. government. The U.S. government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. government purposes.

Data availability statement. The namelists and 3D 10-min time-averaged WRF-LES meteorological and plume fields used in this study can be found online (https://doi.org/10.5281/zenodo.3909881). The digitized PPG measurements were provided through the courtesy of the initiative on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes (http://www.harmo.org/jsirwin; originally made available by the late John S. Irwin).

REFERENCES

Ardeshiri, H., M. Cassiani, S. Y. Park, A. Stohl, I. Pisso, and A. S. Dinger, 2020: On the convergence and capability of the large-eddy simulation of concentration fluctuations in passive plumes for a neutral boundary layer at infinite Reynolds number. *Bound.-Layer Meteor.*, **176**, 291–327, https://doi.org/10.1007/s10546-020-00537-6.

Barad, M. L., 1958: Project Prairie Grass, a field program in diffusion. Volume I. Air Force Cambridge Research Center Geophysics Research Directorate Geophysical Research Papers 59, AFCRC-TR-58-235(I), 300 pp., http://www.harmo.org/jsirwin/PGrassVolumeI.pdf.

Beare, R. J., and Coauthors, 2006: An intercomparison of large-eddy simulations of the stable boundary layer. *Bound.-Layer Meteor.*, **118**, 247–272, https://doi.org/10.1007/s10546-004-2820-6.

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.

Brasseur, J. G., and T. Wei, 2010: Designing large-eddy simulation of the turbulent boundary layer to capture law-of-the-wall scaling. *Phys. Fluids*, **22**, 021303, https://doi.org/10.1063/1.3319073.

Businger, J. A., J. C. Wyngaard, Y. Izumi, and E. F. Bradley, 1971: Flux-profile relationships in the atmospheric surface layer. *J. Atmos. Sci.*, **28**, 181–189, https://doi.org/10.1175/1520-0469(1971)028<0181:FPRITA>2.0.CO;2.

Caulton, D. R., and Coauthors, 2018: Quantifying uncertainties from mobile-laboratory-derived emissions of well pads using inverse Gaussian methods. *Atmos. Chem. Phys.*, **18**, 15 145–15 168, https://doi.org/10.5194/acp-18-15145-2018.

——, and Coauthors, 2019: Importance of superemitter natural gas well pads in the Marcellus shale. *Environ. Sci. Technol.*, **53**, 4747–4754, https://doi.org/10.1021/acs.est.8b06965.

Chang, J. C., and S. R. Hanna, 2004: Air quality model performance evaluation. *Meteor. Atmos. Phys.*, **87**, 167–196, https://doi.org/10.1007/s00703-003-0070-7.

Chung, D., and G. Matheou, 2014: Large-eddy simulation of stratified turbulence. Part I: A vortex-based subgrid-scale model. *J. Atmos. Sci.*, **71**, 1863–1879, https://doi.org/10.1175/JAS-D-13-0126.1.

Cimorelli, A. J., and Coauthors, 2005: AERMOD: A dispersion model for industrial source applications. Part I: General model formulation and boundary layer characterization. *J. Appl. Meteor.*, **44**, 682–693, https://doi.org/10.1175/JAM2227.1.

Coburn, S., and Coauthors, 2018: Regional trace-gas source attribution using a field-deployed dual frequency comb spectrometer. *Optica*, **5**, 320–327, https://doi.org/10.1364/OPTICA.5.000320.

Conley, S., G. Franco I. Faloona, D. R. Blake, J. Peischl, and T. B. Ryerson, 2016: Methane emissions from the 2015 Aliso Canyon blowout in Los Angeles, CA. *Science*, **351**, 1317–1320, https://doi.org/10.1126/science.aaf2348.

——, and Coauthors, 2017: Application of Gauss's theorem to quantify localized surface emissions from airborne measurements of wind and trace gases. *Atmos. Meas. Tech.*, **10**, 3345–3358, https://doi.org/10.5194/amt-10-3345-2017.

Daniels, M. H., K. A. Lundquist, J. D. Mirocha, D. J. Wiersema, and F. K. Chow, 2016: A new vertical grid nesting capability in the Weather Research and Forecasting (WRF) Model. *Mon. Wea. Rev.*, **144**, 3725–3747, https://doi.org/10.1175/MWR-D-16-0049.1.

De Visscher, A., 2013: *Air Dispersion Modeling: Foundations and Applications*. John Wiley and Sons, 664 pp.

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

Dyer, A. J., 1974: A review of flux-profile relationships. *Bound.-Layer Meteor.*, **7**, 363–372, https://doi.org/10.1007/BF00240838.

Eberhard, W. L., W. R. Moninger, and G. A. Briggs, 1988: Plume dispersion in the convective boundary layer. Part I: CONDORS field experiment and example measurements. *J. Appl. Meteor.*, **27**, 599–616, https://doi.org/10.1175/1520-0450(1988)027<0599:PDITCB>2.0.CO;2.

EIA, 2020: United States dry natural gas production. Accessed 22 June 2020, https://www.eia.gov/dnav/ng/hist/n9070us2A.htm.

Foken, T., 2006: 50 Years of the Monin–Obukhov Similarity Theory. *Bound.-Layer Meteor.*, **119**, 431–447, https://doi.org/10.1007/s10546-006-9048-6.

Fox, T. A., T. E. Barchyn, D. Risk, A. P. Ravikumar, and C. H. Hugenholtz, 2019: A review of close-range and screening technologies for measuring fugitive methane emissions in upstream oil and gas. *Environ. Res. Lett.*, **14**, 053002, https://doi.org/10.1088/1748-9326/ab0cc3.

Frankenberg, C., and Coauthors, 2016: Airborne methane remote measurements reveal heavy-tail flux distribution in Four Corners region. *Proc. Natl. Acad. Sci. USA*, **113**, 9734–9739, https://doi.org/10.1073/pnas.1605617113.

Golder, D., 1972: Relations among stability parameters in the surface layer. *Bound.-Layer Meteor.*, **3**, 47–58, https://doi.org/10.1007/BF00769106.

Harper, L. A., O. T. Denmead, and T. K. Flesch, 2011: Micrometeorological techniques for measurement of enteric greenhouse gas emissions. *Anim. Feed Sci. Technol.*, **166–167**, 227–239, https://doi.org/10.1016/j.anifeedsci.2011.04.013.

Horst, T. W., J. C. Doran, and P. W. Nickola, 1979: Evaluation of empirical atmospheric diffusion data. Battelle Pacific Northwest Laboratories Tech Rep. PNL-2599/U.S. Nuclear Regulatory Commission Tech. Rep. NUREG/CR-0798, 137 pp., https://doi.org/10.2172/5716471.

Jiménez, P. A., J. Dudhia, J. F. González-Rouco, J. Navarro, J. P. Montávez, and E. García-Bustamante, 2012: A revised scheme for the WRF surface layer formulation. *Mon. Wea. Rev.*, **140**, 898–918, https://doi.org/10.1175/MWR-D-11-00056.1.

Jongaramrungruang, S., C. Frankenberg, G. Matheou, A. K. Thorpe, D. R. Thompson, L. Kuai, and R. M. Duren, 2019: Towards accurate methane point-source quantification from high-resolution 2-D plume imagery. *Atmos. Meas. Tech.*, **12**, 6667–6681, https://doi.org/10.5194/amt-12-6667-2019.

Kang, M., and Coauthors, 2016: Identification and characterization of high methane-emitting abandoned oil and gas wells. *Proc. Natl. Acad. Sci. USA*, **113**, 13 636–13 641, https://doi.org/10.1073/pnas.1605913113.

Karion, A., and Coauthors, 2013: Methane emissions estimate from airborne measurements over a western United States natural gas field. *Geophys. Res. Lett.*, **40**, 4393–4397, https://doi.org/10.1002/grl.50811.

Khanna, S., and J. G. Brasseur, 1997: Analysis of Monin–Obukhov similarity from large-eddy simulation. *J. Fluid Mech.*, **345**, 251–286, https://doi.org/10.1017/S0022112097006277.

Klipp, C., 2014: Turbulence anisotropy in the near-surface atmosphere and the evaluation of multiple outer length scales. *Bound.-Layer Meteor.*, **151**, 57–77, https://doi.org/10.1007/s10546-013-9884-0.

Kosović, B., 1997: Subgrid-scale modelling for the large-eddy simulation of high-Reynolds-number boundary layers. *J. Fluid Mech.*, **336**, 151–182, https://doi.org/10.1017/S0022112096004697.

Lamb, R. G., 1978: A numerical simulation of dispersion from an elevated point source in the convective planetary boundary layer. *Atmos. Environ.*, **12**, 1297–1304, https://doi.org/10.1016/0004-6981(78)90068-9.

Larsson, J., S. Kawai, J. Bodart, and I. Bermejo-Moreno, 2016: Large eddy simulation with modeled wall-stress: Recent progress and future directions. *Mech. Eng. Rev.*, **3**, 15-00418, https://doi.org/10.1299/mer.15-00418.

Lundquist, K. A., F. K. Chow, and J. K. Lundquist, 2012: An immersed boundary method enabling large-eddy simulations of flow over complex terrain in the WRF Model. *Mon. Wea. Rev.*, **140**, 3936–3955, https://doi.org/10.1175/MWR-D-11-00311.1.

Maronga, B., and J. Reuder, 2017: On the formulation and universality of Monin–Obukhov similarity functions for mean gradients and standard deviations in the unstable surface layer: Results from surface-layer-resolving large-eddy simulations. *J. Atmos. Sci.*, **74**, 989–1010, https://doi.org/10.1175/JAS-D-16-0186.1.

——, C. Knigge, and S. Raasch, 2019: An improved surface boundary condition for large-eddy simulations based on Monin–Obukhov similarity theory: Evaluation and consequences for grid convergence in neutral and stable conditions. *Bound.-Layer Meteor.*, **174**, 297–325, https://doi.org/10.1007/s10546-019-00485-w.

Mason, P. J., 1994: Large-eddy simulation: A critical review of the technique. *Quart. J. Roy. Meteor. Soc.*, **120**, 1–26, https://doi.org/10.1002/qj.49712051503.

Mirocha, J. D., J. K. Lundquist, and B. Kosović, 2010: Implementation of a nonlinear subfilter turbulence stress model for large-eddy simulation in the advanced Research WRF Model. *Mon. Wea. Rev.*, **138**, 4212–4228, https://doi.org/10.1175/2010MWR3286.1.

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.

Mokhtarpoor, R., and S. Heinz, 2017: Dynamic large eddy simulation: Stability via realizability. *Phys. Fluids*, **29**, 105104, https://doi.org/10.1063/1.4986890.

Nieuwstadt, F. T. M., and J. P. J. M. M. de Valk, 1987: A large eddy simulation of buoyant and non-buoyant plume dispersion in the atmospheric boundary layer. *Atmos. Environ.*, **21**, 2573–2587, https://doi.org/10.1016/0004-6981(87)90189-2.

Nottrott, A., J. Kleissl, and R. Keeling, 2014: Modeling passive scalar dispersion in the atmospheric boundary layer with WRF large-eddy simulation. *Atmos. Environ.*, **82**, 172–182, https://doi.org/10.1016/j.atmosenv.2013.10.026.

Nunalee, C. G., B. Kosović, and P. E. Bieringer, 2014: Eulerian dispersion modeling with WRF-LES of plume impingement in neutrally and stably stratified turbulent boundary layers. *Atmos. Environ.*, **99**, 571–581, https://doi.org/10.1016/j.atmosenv.2014.09.070.

Pasquill, F., 1972: Some aspects of boundary layer description. *Quart. J. Roy. Meteor. Soc.*, **98**, 469–494, https://doi.org/10.1002/qj.49709841702.

Peña, A., and A. N. Hahmann, 2020: Evaluating planetary boundary-layer schemes and large-eddy simulations with measurements from a 250-m meteorological mast. *J. Phys.: Conf. Ser.*, **1618**, 062001, https://doi.org/10.1088/1742-6596/1618/6/062001.

Pope, S. B., 2004: Ten questions concerning the large-eddy simulation of turbulent flows. *New J. Phys.*, **6**, 35, https://doi.org/10.1088/1367-2630/6/1/035.

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

Powers, J. G., and Coauthors, 2017: The Weather Research and Forecasting Model: Overview, system efforts, and future directions. *Bull. Amer. Meteor. Soc.*, **98**, 1717–1737, https://doi.org/10.1175/BAMS-D-15-00308.1.

Rao, K. S., 2005: Uncertainty analysis in atmospheric dispersion modeling. *Pure Appl. Geophys.*, **162**, 1893–1917, https://doi.org/10.1007/s00024-005-2697-4.

Saide, P. E., D. Steinhoff, B. Kosovic, J. Weil, N. Downey, D. Blewitt, S. Hanna, and L. Delle Monache, 2018: Evaluating methods to estimate methane emissions from oil and gas production facilities using LES simulations. *Environ. Sci. Technol.*, **52**, 11 206–11 214, https://doi.org/10.1021/acs.est.8b01767.

Sawford, B. L., 2001: Project Prairie Grass—A classic atmospheric dispersion experiment revisited. *14th Australasian Fluid Mechanics Conf.*, Adelaide, Australia, Adelaide University, https://people.eng.unimelb.edu.au/imarusic/proceedings/14/FM010206.PDF.

Skamarock, W. C., and Coauthors, 2019: A description of the Advanced Research WRF Model version 4. NCAR Tech. Note NCAR/TN-556+STR, 145 pp., https://doi.org/10.5065/1dfh-6p97.

Sorbjan, Z., 1986: On similarity in the atmospheric boundary layer. *Bound.-Layer Meteor.*, **34**, 377–397, do,:https://doi.org/10.1007/BF00120989.

Steinfeld, G., S. Raasch, and T. Markkanen, 2008: Footprints in homogeneously and heterogeneously driven boundary layers derived from a Lagrangian stochastic particle model embedded into large-eddy simulation. *Bound.-Layer Meteor.*, **129**, 225–248, https://doi.org/10.1007/s10546-008-9317-7.

Stoll, R., J. A. Gibbs, S. T. Salesky, W. Anderson, and M. Calaf, 2020: Large-eddy simulation of the atmospheric boundary layer. *Bound.-Layer Meteor.*, **177**, 541–581, https://doi.org/10.1007/s10546-020-00556-3.

Stull, R. B., 1988: *An Introduction to Boundary Layer Meteorology*. Kluwer Academic, 666 pp.

Sullivan, P. P., T. W. Horst, D. H. Lenschow, C.-H. Moeng, and J. C. Weil, 2003: Structure of subfilter-scale fluxes in the atmospheric surface layer with application to large-eddy simulation modelling. *J. Fluid Mech.*, **482**, 101–139, https://doi.org/10.1017/S0022112003004099.

Taylor, D. M., F. K. Chow, M. Delkash, and P. T. Imhoff, 2016: Numerical simulations to assess the tracer dilution method for measurement of landfill methane emissions. *Waste Manage.*, **56**, 298–309, https://doi.org/10.1016/j.wasman.2016.06.040.

Thorpe, A. K., and Coauthors, 2020: Methane emissions from underground gas storage in California. *Environ. Res. Lett.*, **15**, 045005, https://doi.org/10.1088/1748-9326/ab751d.

U.S. EPA, 2014: Other Test Method (OTM) 33 and 33A Geospatial Measurement of Air Pollution-Remote Emissions Quantification Direct Assessment (GMAP-REQ-DA). EPA Tech. Rep., 91 pp., https://www3.epa.gov/ttnemc01/prelim/otm33a.pdf.

van Ulden, A. P., 1978: Simple estimates for vertical diffusion from sources near the ground. *Atmos. Environ.*, **12**, 2125–2129, https://doi.org/10.1016/0004-6981(78)90167-1.

Varon, D. J., D. J. Jacob, J. McKeever, D. Jervis, B. O. A. Durak, Y. Xia, and Y. Huang, 2018: Quantifying methane point sources from fine-scale satellite observations of atmospheric methane plumes. *Atmos. Meas. Tech.*, **11**, 5673–5686, https://doi.org/10.5194/amt-11-5673-2018.

Venkatram, A., 1996: An examination of the Pasquill–Gifford–Turner dispersion scheme. *Atmos. Environ.*, **30**, 1283–1290, https://doi.org/10.1016/1352-2310(95)00367-3.

Weil, J. C., W. H. Snyder, R. E. Lawson, and M. S. Shipman, 2002: Experiments on buoyant plume dispersion in a laboratory convection tank. *Bound.-Layer Meteor.*, **102**, 367–414, https://doi.org/10.1023/A:1013874816509.

——, P. P. Sullivan, and C.-H. Moeng, 2004: The use of large-eddy simulations in Lagrangian particle dispersion models. *J. Atmos. Sci.*, **61**, 2877–2887, https://doi.org/10.1175/JAS-3302.1.

——, ——, E. G. Patton, and C.-H. Moeng, 2012: Statistical variability of dispersion in the convective boundary layer: Ensembles of simulations and observations. *Bound.-Layer Meteor.*, **145**, 185–210, https://doi.org/10.1007/s10546-012-9704-y.

Wilks, D. S., 2019: *Statistical Methods in the Atmospheric Sciences*. 4th ed. Elsevier, 840 pp.

Willis, G. E., and J. W. Deardorff, 1976: A laboratory model of diffusion into the convective planetary boundary layer. *Quart. J. Roy. Meteor. Soc.*, **102**, 427–445, https://doi.org/10.1002/qj.49710243212.

Wurps, H., G. Steinfeld, and S. Heinz, 2020: Grid-resolution requirements for large-eddy simulations of the atmospheric boundary layer. *Bound.-Layer Meteor.*, **175**, 179–201, https://doi.org/10.1007/s10546-020-00504-1.

Wyngaard, J. C., Ed., 2010: The atmospheric surface layer. *Turbulence in the Atmosphere*, Cambridge University Press, 215–240, https://doi.org/10.1017/CBO9780511840524.011.

Xue, F., H. Kikumoto X. Li, and R. Ooka, 2018: Bayesian source term estimation of atmospheric releases in urban areas using LES approach. *J. Hazard. Mater.*, **349**, 68–78, https://doi.org/10.1016/j.jhazmat.2018.01.050.