1  **Evaluating WRF-GC v2.0 predictions of boundary layer and vertical ozone**
2  **profiles during the 2021 TRACER-AQ campaign in Houston, Texas**

3  **Xueying Liu[1], Yuxuan Wang[1], Shailaja Wasti[1], Wei Li[1], Ehsan Soleimanian[1], James**
4  **Flynn[1], Travis Griggs[1], Sergio Alvarez[1], John T. Sullivan[2], Maurice Roots[3], Laurence**
5  **Twigg[4], Guillaume Gronoff[5], Timothy Berkoff[5], Paul Walter[6], Mark Estes[6], Johnathan W.**
6  **Hair[5], Taylor Shingler[5], Amy Jo Scarino[4], Marta Fenn[4], Laura Judd[5]**

7  [1]Department of Earth and Atmospheric Sciences, University of Houston, Houston, TX, USA
8  [2]NASA Goddard Space Flight Center, Greenbelt, MD, USA
9  [3]Department of Physics, University of Maryland Baltimore County, MD, USA
10  [4]Science Systems and Applications, Inc., Hampton, VA, USA
11  [5]NASA Langley Research Center, Hampton, VA, USA
12  [6]School of Natural Sciences, St. Edward's University, Austin, TX, USA

13  **Correspondence:** Yuxuan Wang (ywang246@central.uh.edu)

14  **Abstract.** The Tracking Aerosol Convection Experiment Air Quality (TRACER-AQ) campaign
15  probed Houston air quality with a comprehensive suite of ground-based and airborne remote
16  sensing measurements during the intensive operating period in September 2021. Two post-frontal
17  high-ozone episodes (September 6-11 and 23-26) were recorded during the said period. In this
18  study, we evaluated the simulation of the planetary boundary layer (PBL) height and the vertical
19  ozone profile by a high-resolution (1.33 km) 3-D photochemical model, Weather Research and
20  Forecasting (WRF)-driven GEOS-Chem (WRF-GC). We contrasted the model performance
21  between ozone-episode days and non-episode days. The model captures the diurnal variations of
22  the PBL during ozone episodes (R = 0.72-0.77; normal mean bias (NMB) = 3%-22%) and non-
23  episode days (R = 0.88; NMB = -21%), compared with the ceilometer at La Porte. Land-water
24  differences in PBL heights are captured better during non-episode days than episode days,
25  compared with the airborne High Spectral Resolution Lidar-2 (HSRL-2). During ozone episodes,
26  the simulated land-water differences are 50-60 m (morning), 320-520 m (noon), and 440-560 m
27  (afternoon) in comparison with the observed values of 190 m, 130 m, and 260 m, respectively.
28  During non-episode days, the simulated land-water differences are 140-220 m (morning) and 360-
29  760 m (noon) in comparison with the observed values of 210 m and 420 m, respectively. For
30  vertical ozone distributions, the model was evaluated against vertical profile measurements from
31  the Tropospheric Ozone lidar (TROPOZ), the HSRL-2, and ozonesondes, as well as at the surface
32  from a model 49i ozone analyzer and a site from the Continuous Ambient Monitoring Stations
33  (CAMS) at La Porte. The model underestimates free tropospheric ozone (2-3 km aloft) by 9%-22%
34  but overestimates near-ground ozone (< 50 m aloft) by 6%-39% during the two ozone episodes.
35  Boundary layer ozone (0.5-1 km aloft) is underestimated by 1%-11% during September 8-11 but
36  overestimated by 0%-7% during September 23-26. Based on these evaluations, we identified two
37  model limitations: the single-layer PBL representation and free tropospheric ozone
38  underestimation. These limitations have implications for the predictivity of ozone's vertical
39  mixing and distribution in other models.
40
41

**1. Introduction**

The Houston metropolitan area has experienced nonattainment of the US National Ambient Air
Quality Standards (NAAQS) for ozone over decades (TCEQ, 2022). Ozone exceedances in
Houston usually occur in two peaks, a spring peak in April–May and a late summer peak in
August–October (Zhou et al., 2014). Such seasonal behavior is driven by diverse meteorological
conditions that influence ozone development. The passages of synoptic-scale cold fronts (~ 1000
km horizontally and ~ 5 km vertically; a timescale of days) are known to bring high background
ozone air from the continent into the Houston area (Lefer et al., 2010; McMillan et al., 2010;
Haman et al., 2014). Mesoscale sea breeze recirculation (~ 20 km horizontally and ~ 1 km
vertically; a timescale of hours) is found to be associated with ozone exceedances (Li et al.,
2020; Banta et al., 2005, 2011; Caicedo et al., 2019). Meanwhile, microscale-to-mesoscale
vertical mixing (< 1 km vertically; a timescale of hours) of the lower troposphere is shown to be
a significant factor in near-surface ozone air quality (Morris et al., 2010; Haman et al., 2014;
Sullivan et al., 2017; Xu et al., 2018; Caputi et al., 2019). Favored by these meteorological
conditions of different scales, local emissions of ozone precursors from the urban center and the
nearby Houston Ship Channel stay locally in the area and lead to high-ozone events. This study
will focus on the impact of mixing between lower free tropospheric layers on vertical ozone
distribution and the impact of chemistry is outside the scope of this analysis.

The planetary boundary layer (PBL) is the lower part (e.g. < 2 km) of the troposphere that is
directly influenced by the presence of the Earth's surface and responds to surface forcings with a
timescale of an hour or less. A stable capping layer at the top of the PBL, where temperature
increases with height, is known as the capping inversion (CI) layer (e.g. ~ 2 km). With the cap in
place, air exchange is inhibited between the overlying free troposphere (FT) (e.g. >2 km) and the
underlying PBL (e.g. < 2 km). During the daytime, there is strong turbulence production
throughout the PBL, generating a buoyant layer called the convective boundary layer (CBL). The
CBL is characterized by intense mixing in a statically unstable situation where warm air rises
from the ground, growing from a few hundred meters in the early morning (e.g. ~ 0.5 km)
towards the top of the PBL in the afternoon (e.g. ~ 2 km). As the sun sets, convectively driven
turbulence decays in the formerly well-mixed CBL. The remnant of the recently decayed CBL
will remain aloft in the less-turbulent residual layer (RL) at around 1–2 km. As the night
progresses, the bottom portion of the RL transforms into a stable boundary layer (SBL) (e.g. <
0.5 km) due to its contact with the ground, characterized by statically stable air with weak and
sporadic turbulence. The PBL is commonly considered as the CBL under certain conditions
during the daytime and the SBL during the nighttime (Tangborn et al., 2021).

The heights of the PBL (incl. CBL and SBL) and other lower tropospheric layers (e.g. RL, CI)
are defined mainly by temperature inversions. It is primarily a thermodynamic-based definition.
Atmospheric models adopt the thermodynamic concept and rely on parameterization schemes to
define the structure of the PBL and compute the height of the PBL. Two major types of data
have been commonly used to validate the modeled PBL height. The first type is the PBL height
derived from the profiles of thermodynamic properties measured by ozonesondes and
radiosondes (Zhang et al., 2019; Zhang et al., 2020; Morris et al., 2010; Rappenglück et al.,
2008). These observations share a similar thermodynamic definition with the modeled PBL
height and are widely used to validate model prediction of the PBL height under various
conditions (day, night, land, water). The second type of data is remotely-sensed mixed layer

1  height as defined by aerosol backscatter gradients, which is becoming more widely available
2  with ceilometer data and aircraft lidars and can be adopted for model evaluation (Caicedo et al.,
3  2017, 2020; Knepp et al., 2017; Li et al., 2021; Wang et al., 2020).

4  Mixed layer height, defined as the volume of atmosphere in which aerosols are well mixed and
5  dispersed, can be derived from the unattenuated backscatter signal of aerosols alone (e.g. the
6  High Spectral Resolution Lidar-2 (HSRL-2)) or the attenuated total backscatter signal produced
7  by aerosols and molecules combined (e.g. CHM 15k-x ceilometers). Both signals have been used
8  to derive mixed layer height for model comparisons (Scarino et al., 2014; Li et al., 2022). Mixed
9  layer height does not equal PBL height by definition; it approximates the CBL height during the
10 daytime and can represent the height of the RL or the SBL depending on retrieval algorithms
11 applied to lidar signals at night (Wang et al., 2020; Vivone et al., 2021). Mixed layer height is
12 often a good proxy for the heights of different lower tropospheric layers determined
13 thermodynamically in models during the daytime (Scarino et al., 2014) and throughout the day
14 (Kuik et al., 2016; Haman et al., 2014) and serves as an input parameter of PBL heights for
15 meteorological and photochemical models (Tangborn et al., 2021; Knote et al., 2015; Geiß et al.,
16 2017).

17 Vertical mixing between different layers of the lower troposphere, such as boundary layer
18 mixing with the FT flow at its upper interface (through entrainment processes), mixing between
19 the RL and the SBL (through surface exchange processes) and the RL mixing through the growth
20 of the CBL, etc., strongly influences surface ozone concentrations. Entrainment can occur during
21 the daytime when strong convective thermals penetrate the laminar FT above and then sink back
22 into the CBL, bringing the FT air towards the surface and thus affecting surface ozone
23 concentrations (Parrish et al., 2010; Jaffe et al., 2011). Located between the FT and the CBL, the
24 strength of the CI layer limits the upward penetration of thermals and is thus used to indicate the
25 influence of the FT air on surface ozone (Kaser et al., 2017; Morris et al., 2010; Rappenglück et
26 al., 2008). Meanwhile, surface exchange processes occur when a low-level jet exists between the
27 RL and the underlying SBL and drives the shear production of turbulence between these layers.
28 Since the RL is a known ozone reservoir with limited $NO_x$ titration and ozone deposition, ozone-
29 rich air in the RL can be mixed down into the SBL effectively, where it is subject to dry
30 deposition to the surface, affecting surface ozone concentrations (Tucker et al., 2010; Sullivan et
31 al., 2017; Caputi et al., 2019; Bernier et al., 2019; Zhao et al., 2022; Xu et al., 2018).

32 The Tracking Aerosol Convection Experiment Air Quality (TRACER-AQ, https://www-
33 air.larc.nasa.gov/missions/tracer-aq/) campaign, led by NASA with contributions from the Texas
34 Commission on Environmental Quality (TCEQ), probed Houston air quality with a
35 comprehensive suite of remote sensing and in situ measurements of ozone, ozone precursors, and
36 meteorology from ground-based, airborne, balloon-borne and shipborne platforms (Jensen et al.,
37 2022). The operational period occurred from July–September 2021, with intensive measurements
38 during September 2021. Combining field campaign observations with a high-resolution 3-D
39 photochemical model, the goals of this study are to (1) evaluate the PBL prediction in the model,
40 (2) examine the vertical distribution of ozone, and (3) identify specific model limitations that
41 prevent accurate prediction of the PBL height and the vertical ozone distribution.

42

1 **2. Model and Data**
2 **2.1 Observations**
3 To evaluate the PBL and the vertical ozone distribution, this study adopted continuous, high-
4 resolution profiles (i.e., 1–10 minutes) from ground-based measurements at the La Porte site and
5 airborne measurements covering urban Houston and the Galveston Bay in September of 2021
6 (Fig. 1). Compared with discrete or low-resolution PBL measurements (*e.g.* hourly) used in
7 previous studies in Houston (Haman et al., 2014; Cuchiara et al., 2014; Rappenglück et al.,
8 2008), the high-resolution measurements in TRACER-AQ field campaign are capable to probe
9 into the fine PBL structure and its development as well as the associated vertical ozone profiles.

10 The La Porte site was equipped with (1) semi-continuous vertical ozone profiles from the NASA
11 Goddard Space Flight Center (GSFC) Tropospheric Ozone (TROPOZ) Differential Absorption
12 Lidar (DIAL) (Sullivan et al., 2014), (2) continuous aerosol mixed layer height derived from
13 atmospheric backscatter profiling with a CHM 15k-x ceilometer, (3) multiple ozonesonde
14 launches, and (4) continuous surface ozone and meteorology measurements.

15 The TROPOZ, as part of the ground-based Tropospheric Ozone Lidar Network (TOLNet,
16 https://www-air.larc.nasa.gov/missions/TOLNet/), has been used to provide continuous, high-
17 resolution profile measurements of vertical ozone profile during various campaigns for satellite
18 and model evaluation (Sullivan et al., 2014, 2015, 2019, 2022; Bernier et al., 2022; Kotsakis et
19 al., 2022; Dacic et al., 2020; Johnson et al., 2016; Dreessen et al., 2016). The TROPOZ data can
20 be used to identify pollutant transport to understand the vertical mixing of ozone. Similar to the
21 TROPOZ at the La Porte site, the University of Houston site measured semi-continuous vertical
22 ozone profiles with the Langley Mobile Ozone Lidar (LMOL) (Gronoff et al., 2019, 2021).

23 The CHM 15k-x ceilometer measured continuous atmospheric attenuated backscatter profiles at
24 a wavelength of 1064 nm. The signals were corrected due to the incomplete superposition of the
25 laser and the receiver field of view by the overlapping correction function from the manufacturer
26 (Rizza et al., 2017). The normalized range corrected signals (RCS) is shown in this paper. The
27 sharp gradients in the collected backscatter were then used to detect up to three aerosol layers by
28 the standard retrieval algorithm provided by the ceilometer manufacturer (Lufft, 2016). The
29 lowest determined aerosol layer is characterized as mixed layer height. It depends on the users to
30 determine whether the derived mixed layer height can be used as a proxy for thermodynamically-
31 defined layers such as the CBL, the SBL and the RL (Caicedo et al., 2017, 2020; Knepp et al.,
32 2017; Li et al., 2021; Wang et al., 2020).

33 Ozonesondes were often launched multiple times in a day at several locations and measured
34 vertical profiles of ozone and meteorological variables including temperature, humidity, and
35 winds. This study uses ozone and potential temperature profiles from eight ozonesondes at La
36 Porte launched on 10:00-15:00 CDT during ozone episodes.

37 Surface measurements at La Porte included ozone, air temperature, relative humidity, and wind
38 speed and direction. This study uses surface ozone from a Thermo Scientific model 49i ozone
39 analyzer operated by the GSFC and a TCEQ Continuous Ambient Monitoring Stations (CAMS)
40 site named La Porte Sylvan Beach, as well as surface meteorology from a Lufft WS-501B
41 operated by the GSFC.

1    In September 2021, the NASA Gulfstream-V aircraft flew on ten flight days. This analysis uses
2    the High Spectral Resolution Lidar-2 (HSRL-2) datasets collected over the Houston area up to
3    three times per day, roughly at 8:00-10:00, 11:00-13:00, and 14:00-16:00 local time, covering an
4    area of approximately 50 km x 135 km. With its high resolution and vertically resolved
5    measurements, the HSRL-2 demonstrated reliable performances on many previous airborne
6    campaigns (Hair et al., 2018; Hair et al., 2008; Burton et al., 2015).The HSRL-2 provides below
7    aircraft retrievals of the spatial and vertical distributions of ozone, aerosols, aerosol optical
8    properties, and mixed layer heights. This paper only reports on (1) mixed layer height derived
9    from gradients in the aerosol backscatter profiles measured at 532 nm and (2) ozone mixing ratio
10   along one flight track that has the nearest distance to the La Porte site (Fig. 1).

11   Mixed layer heights from the ceilometer at La Porte and the HSRL-2 are derived differently. The
12   ceilometer at the La Porte site measures attenuated total backscatter profiles of the atmosphere
13   (incl. aerosols and molecules), while the HSRL-2 can measure the unattenuated aerosol
14   backscatter profile. Both ceilometer and the HSRL-2 signals can be used to derive mixed layer
15   height. This study uses mixed layer heights from the ceilometer and the HSRL-2 to evaluate the
16   WRF-GC prediction of PBL heights.
17
18   The supplement includes (1) surface measurements from the TCEQ CAMS and the boats
19   throughout July to October used to identify ozone exceedance days, (2) details on the
20   assimilation and evaluation of the modeled meteorology with these measurements, and (3)
21   vertical ozone distribution at University of Houston by LMOL.
22
23   Apart from the observations above, we used geopotential heights and winds at 850 hPa from the
24   European Centre for Medium Range Weather Forecast (ECMWF) reanalysis version5 (ERA5)
25   dataset (description in Sect. 2.3.2) to derive the synoptic conditions in Fig. 2.
26
27   **2.2 Identification of ozone episodes**
28   Ozone exceedance days were identified according to surface measurements from the TCEQ
29   CAMS (onshore) and the boats (offshore). The criteria used in this study is that any onshore site
30   from the CAMS network in Houston and Galveston or offshore boat ozone observations
31   registered daily maximum 8-hour average (MDA8) ozone in exceedance of 70 ppbv, the current
32   U.S. Environmental Protection Agency (EPA) NAAQS air quality standard for ozone. Three
33   high ozone episodes in September of 2021 were identified based on the above criteria:
34   September 6-11, September 17-19, and September 23-26, consisting of 13 ozone exceedance
35   days. We excluded analysis from the September 17-19 episode because it happened right after
36   tropical cyclone Nicholas, which made landfall 125 km south-southwest of Houston and
37   hindered measurements at the ground-sites and aircraft due to clouds and power outages. The
38   model meteorology was not designed to capture the cyclone either. Other September days were
39   used as a control to represent non-episode days.
40
41   **2.3 Model**
42   **2.3.1 Model description**
43   WRF-GC v2.0 is a regional air quality model (Feng et al., 2021; Lin et al., 2020) that couples the
44   Weather Research and Forecasting (WRF) meteorological model (v3.9.1.1) with the GEOS-
45   Chem atmospheric chemistry model (v12.7.2). The WRF and GEOS-Chem versions are

1  benchmarks of WRF-GC v2.0 with the proven performance of meteorology, PBL heights, and
2  aerosol simulation in Feng et al. (2021) and Lin et al. (2020). We evaluated WRF-GC's
3  prediction of ozone during the TRACER-AQ study. We set up three domains with different
4  horizontal resolutions that cover the contiguous United States, Southeast Texas, and the
5  Houston-Galveston region, referred to as d01, d02, and d03, respectively, as shown in Figure 1.
6  The corresponding horizontal resolutions for d01–d03 are 12 km, 4 km, and 1.33 km,
7  respectively. All domains have identical vertical resolutions with 50 hybrid sigma-eta vertical
8  levels spanning from the surface to 10 hPa. Vertical resolution ranges from ~70 m (near the
9  ground) to ~700 m (aloft); the first 2 km above the ground has 10 model layers, and the first 4
10  km has 14 model layers.
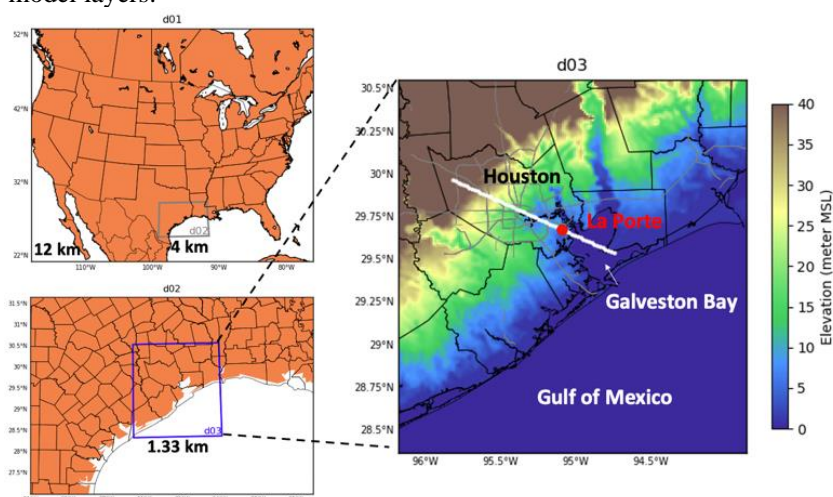


11
12  Figure 1. WRF-GC nested domains and their horizontal resolutions. The La Porte site is labeled
13  as a red dot. The white line represents a flight track that is chosen because of its nearest distance
14  to the La Porte site.
15
16  WRF-GC uses the most updated full $O_x$-$NO_x$-VOC-halogen-aerosol chemistry from GEOS-
17  Chem. The anthropogenic emissions used are the 2019 TCEQ emission inventory for Houston
18  and southeastern Texas, the 2013 National Emission Inventory for the rest of the US, and the
19  2014 Community Emissions Data System (CEDS) for regions outside of the US. Biomass
20  burning emissions are from the 2019 Global Fire Emissions Database (GFED). Biogenic
21  emissions are from the Model of Emissions of Gases and Aerosols from Nature (MEGAN)
22  (Guenther et al., 2012). Soil $NO_x$ (Hudman et al., 2012) and lightning $NO_x$ (Murray et al., 2012)
23  emissions are also included.
24
25  **2.3.2 Model configurations**
26  Boundary and initial conditions for WRF employed three alternative meteorological analyses.
27  They were (1) the National Centers for Environmental Prediction (NCEP)-Final Analysis (FNL)
28  (https://rda.ucar.edu/datasets/ds083.3/), (2) the fifth generation of European Centre for Medium-
29  Range Weather Forecasts (ECMWF) atmospheric reanalysis (ERA5) data
30  (https://rda.ucar.edu/datasets/ds633.0/), and (3) the High-Resolution Rapid Refresh (HRRR)
31  from NOAA Amazon Web Service (https://registry.opendata.aws/noaa-hrrr-pds). The temporal

1  resolution for FNL, ERA5, and HRRR is 6-hourly, hourly, and hourly, respectively. The
2  horizontal resolution for FNL, ERA5, and HRRR is 0.25°, 0.25°, and 3 km, respectively.
3
4  WRF has different schemes or options to represent physics and dynamics processes. Three PBL
5  schemes were used to investigate the effect of different parameterizations of heat, moisture, and
6  momentum exchange between the surface and PBL on the simulated PBL structure and height.
7  They are the local closure Mellor-Yamada-Nakanishi-Niino (MYNN) scheme (Nakanishi and
8  Niino, 2009), the non-local closure Yonsei University (YSU) scheme (Hong et al., 2006), and the
9  hybrid local-nonlocal Asymmetric Convective Model version 2 (ACM2) scheme (Pleim, 2007).
10 Details of the PBL schemes are in Sect. 2.3.3. Two microphysics schemes were used: the
11 Morrison double moment (2M) scheme (Morrison et al., 2009) and the single-moment 6-class
12 (WSM6) scheme (Hong and Lim, 2006). Other schemes adopted were the Monin-Obukhov
13 Similarity surface layer, the Noah land surface scheme (Chen and Dudhia, 2001), the Rapid
14 Radiative Transfer Model (RRTM) longwave and shortwave radiation schemes (Iacono et al.,
15 2008), and the New Tiedtke cumulus scheme (Zhang et al., 2011; Tiedtke, 1989).
16
17 To select the best model configuration to represent meteorology during the 2021 TRACER-AQ
18 campaign, we designed eight model experiments with different physics options, boundary
19 meteorology, data assimilation, and reinitializing option, as listed in Table S2. First, [Base] is the
20 baseline configuration: MYNN for PBL, 2M for microphysics, NCEP FNL for boundary
21 conditions, no nudging for assimilation, and no reinitialization. Second, [YSU] and [ACM2]
22 experiments used the YSU and ACM2 PBL schemes, respectively, while keeping other options
23 the same as [Base]. Differences between [Base], [YSU], and [ACM2] show the effects of
24 different PBL parameterizations. Third, the [WSM6] experiment differs from [Base] by
25 replacing the 2M microphysics scheme with WSM6. Differences between [Base] and [WSM6]
26 show the effects of different microphysics schemes. Next, [ERA5] and [HRRR] were designed to
27 show the effects of different meteorological initial and boundary conditions on the WRF
28 performance by using ERA5 and HRRR instead of NCEP FNL, respectively. We examined the
29 effects of data assimilation options in [Nudged]. [Nudged] adopted observation nudging and
30 surface analysis nudging to assimilate both onshore and offshore measurements from multiple
31 platforms, including the TCEQ CAMS, boats, and the NCEP surface and upper air measurements
32 into WRF meteorology (see Text S2 for details). Differences between [Base] and [Nudged] show
33 the effects of assimilation. Last, [Reinit] used daily reinitialization where the simulation was
34 broken into many 30-hour segments with the first 6 hours of each segment (18:00-23:00
35 Coordinated Universal Time (UTC) of a previous day) as spin-up and the subsequent 24 hours
36 (0:00-23:00 UTC of the following day) used for analysis (Yahya et al., 2015; Otte et al., 2008).
37 Differences between [Base] and [Reinit] show the effects of a free-running option versus model
38 reinitialization.
39
40 Among the above simulations, we chose four simulations (Table 1), including the three
41 simulations with different PBL schemes and the best simulation [HRRR] determined by
42 campaign-wide statistics (see Text S3 for details) in the analysis below. The surface layer, land
43 surface, longwave and shortwave radiation, and Tiedtke cumulus schemes remain unchanged in
44 all simulations.
45
46

1
2
3    Table 1. List of simulations used in this study.

| Simulations | Meteorology for Boundary Condition | PBL scheme |
|---|---|---|
| [Base] | NCEP FNL | MYNN |
| [YSU] | NCEP FNL | YSU |
| [ACM2] | NCEP FNL | ACM2 |
| [HRRR] | HRRR | MYNN |

4
5    **2.3.3 Determination of PBL height in different schemes**
6    The heights of the PBL are determined differently among different PBL schemes in the WRF
7    model. The intra-scheme differences can originate from (1) the vertical profile of thermodynamic
8    quantities simulated with different assumptions of the vertical exchange of heat, moisture, and
9    momentum and (2) the diagnosis of the PBL height from these thermodynamic quantities. The
10   PBL heights determined by different schemes can differ by 20-30% (Hu et al., 2010; Xie et al.,
11   2013).

12
13   First, the common parameterizations of vertical exchange include local and non-local closure
14   schemes. Local closure schemes estimate the turbulent fluxes at each point in model grids from
15   the mean atmospheric variables and their gradients at that point. In contrast, non-local closure
16   schemes include the nonlocal upward transport by buoyant plumes, representing large-scale
17   motions. Among the three PBL schemes used in this study, the MNYY scheme is local, the YSU
18   is nonlocal, and the ACM2 is hybrid local-nonlocal.

19
20   Second, the bulk Richardson number (BRN) and the turbulent kinetic energy (TKE) methods are
21   the two common methods to determine PBL height. The BRN method diagnoses PBL height
22   thermodynamically by potential temperature with wind speeds and is adopted in the YSU and the
23   ACM2 schemes. The PBL heights under this condition are defined as the height of the model
24   layer where the bulk Richardson number reaches a critical value. The two schemes have two
25   major differences. The YSU scheme calculates the bulk Richardson number starting from the
26   surface while the ACM2 scheme calculates it above the neutral buoyancy level (Hu et al., 2010;
27   Hong et al., 2006; Pleim, 2007). The critical value is 0.25 for stable conditions and 0 for unstable
28   conditions in the YSU scheme and it is 0.25 for both stable and unstable conditions in the ACM2
29   scheme (Xie et al., 2013). Meanwhile, the TKE method diagnoses PBL height by horizontal and
30   vertical winds and is adopted in the MYJ scheme (not used in this study). The PBL height under
31   this condition is diagnosed when the TKE decreases to a minimum of 0.1 $m^2 s^{-2}$. A hybrid
32   definition that combines the BRN and the TKE methods is implemented in the MYNN scheme.
33   The hybrid method weights the TKE method more during stable conditions when the BRN-based
34   PBL height is below ~0.5 km, while it weights the TKE-based definition negligible when the
35   BRN-based PBL height is above ~1 km.

36
37   Previous studies have demonstrated that the mentioned schemes outperform each other under
38   different conditions across regions, evaluated with various metrics (Hu et al., 2010; Xie et al.,
39   2012; Xie et al., 2013). No conclusion is reached as to which scheme is universally the best. No
40   systematic higher or lower PBL height is expected from one scheme relative to one another.

**3. Lower tropospheric layering for ozone-episode and non-ozone-episode days**

The geopotential heights at 850 hPa in Figure 2 show different synoptic conditions are seen between ozone-episode and non-episode days in September 2021. The non-episode days experienced clean southerlies from the Gulf of Mexico (Fig. 2a), while the ozone episodes of September 6-11 and 23-26 both happened after a cold frontal passage with a low pressure sitting in the northeast US and a high pressure located in eastern Texas (Fig. 2b, 2c). This synoptic structure puts the Houston region under northerly wind conditions, which bring colder and more polluted continental air to the region, leading to relatively lower temperature (Fig. 3b) and relative humidity (Fig. 3d) than non-episode days (Fig. 3a; Fig. 3c).
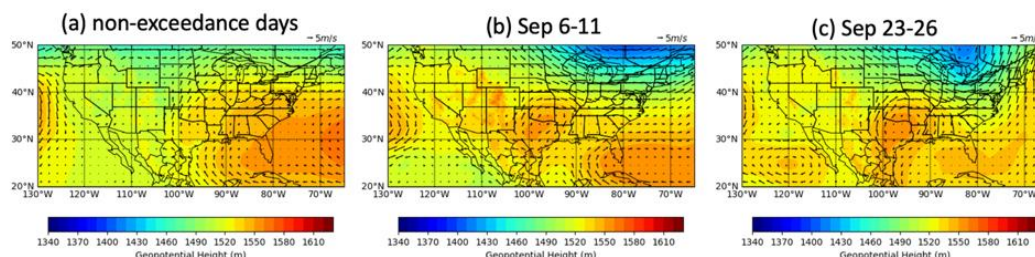


Figure 2. Synoptic conditions denoted by geopotential height at 850 hPa and the associated winds for (a) the non-episode days and the two ozone episodes of (b) September 6-11 and (c) September 23-26, 2021.

Apart from differences in meteorological variables, synoptic high-pressure centers during ozone episodes tend to create a clear, calm condition with light horizontal winds at night when the RL and the multiple layer structure of the lower troposphere (incl. an SBL, an RL, and a CI layer) are prone to form, while the RL structure tends to be disrupted due to shear effects under meteorological conditions during non-episode days (Stull et al., 1988; Yi et al., 2001). We find mixed layer heights derived from the ceilometer at La Porte during non-episode days (Fig. 3i) and ozone-episode (Fig. 3j) are similar during the daytime, while the nocturnal mixed layer heights (*e.g.* 0:00–10:00 CDT) are greater on ozone-episode days than on non-episode days. Such differences can also be seen from the direct measurements of the ceilometer, i.e. atmospheric backscatter profiles, in Figures 3k and 3l. During ozone episodes, the high-pressure center creates favorable meteorological condition for multiple nocturnal layers to form. Among these, the RL contains much of the aerosol remnant left by the daytime CBL and is therefore detected by the ceilometer during ozone episodes (Fig. 3l). In contrast, no such multiple layers form under meteorological conditions on non-episode days. Much of the aerosol remnant above the SBL is dissipated with the disruption of RL by wind shear such that the SBL contains more aerosol than above. Therefore, the ceilometer detects the SBL on non-episode days (Fig. 3k). In this study, the mixed layer heights derived from the ceilometer represent the RL during ozone episodes but the SBL during non-episode days.

Mixed layer height is often a good proxy for the heights of different lower tropospheric layers determined thermodynamically in models (Scarino et al., 2014; Kuik et al., 2016; Haman et al., 2014). We refer to the standard mixed layer retrievals, that is the CBL during the daytime, the SBL at night during non-episode days, and the RL at night during ozone episodes, respectively

1    as observed CBL, SBL or RL hereafter in a manner consistent with the modeled equivalents. The
2    next section evaluates the observed and the modeled PBL heights.
3
4    **4. PBL height evaluation**
5    In this section, we evaluate the modeled PBL height with two types of independent field
6    measurements. The ground-based ceilometer at the La Porte site is used to evaluate the diurnal
7    variation, given its continuous measurements throughout the day. Meanwhile, the HSRL-2
8    instrument acquired data over much of the urban Houston region and adjacent waters and is thus
9    used to evaluate spatial and temporal (daytime) variations of the PBL.
10
11   **4.1 Evaluation with ceilometer**
12   Diurnal variations of the PBL heights averaged during non-episode periods and ozone episodes
13   are separately evaluated in Figure 3. The observations represent the daytime CBL on both types
14   of days. At night, the observations represent the SBL on non-episode days but the RL on ozone-
15   episode days. The modeled equivalents are needed to yield equal comparisons between the
16   models and observations. The model diagnoses the CBL height as a standard output for the PBL
17   height during the daytime. However, the model only diagnoses the SBL as the standard output
18   for nighttime PBL rather than other nocturnal layers such as RL (Fig. 3j). Therefore, the modeled
19   RL needs to be extracted for a valid comparison with the observed RL during ozone episodes.
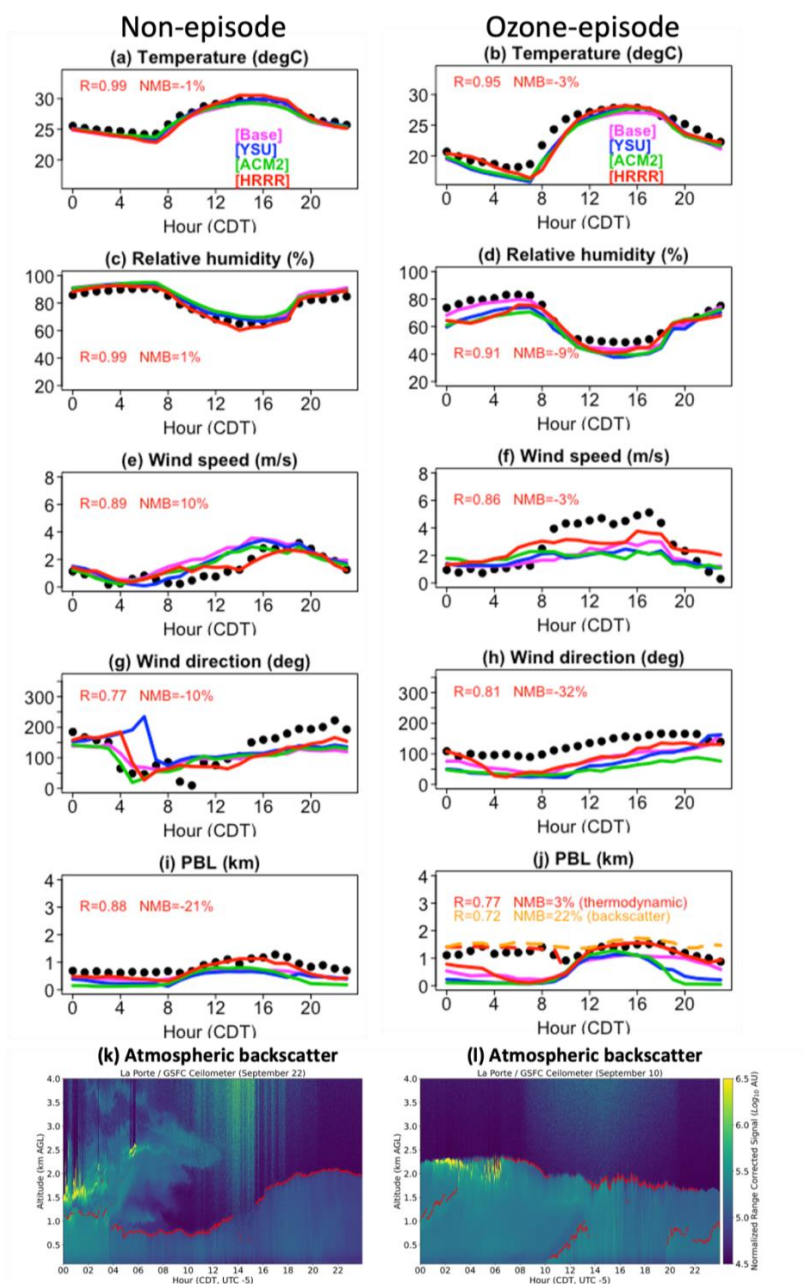
1
2  Figure 3. Diurnal variation in observed versus modeled surface meteorology and PBL height
3  averaged over different days during non-episode periods (left) and ozone episodes (right) in
4  September 2021. For the first five rows, black dots are NASA GSFC observations at the La Porte
5  site while lines are equivalent data simulated by the WRF-GC model. Different colors denote
6  different model configurations. In (j), dashed lines represent residual layers identified by aerosol
7  backscatter and potential temperature from the [HRRR] configuration. The last row shows the

1  ceilometer-measured atmospheric backscatter profiles overlaid with mixed layer heights of two
2  typical days; non-episode day September 22 (k) and ozone-episode day September 10 (l).
3
4  Before extracting RL characteristics in the model, we first selected one simulation with the best
5  daytime performance among the four simulations and examine its nighttime performance. On
6  non-episode days (Fig. 3i), [Base], [YSU], [ACM2], and [HRRR] respectively show the diurnal
7  mean and standard deviation of the PBL height of 0.52±0.14 km, 0.43±0.17 km, 0.39±0.27 km,
8  and 0.66±0.28 km in comparison with the observed value of 0.83±0.22 km. On ozone-episode
9  days (Fig.3j), the model simulations show the CBL height variation of 0.96±0.18 km ([Base]),
10  0.60±0.37 km ([YSU]), 0.50±0.5 km ([ACM2]), and 1.25±0.29 km ([HRRR]) in comparison
11  with the observed value of 1.26±0.24 km during the afternoon and evening hours (15:00-23:00
12  CDT). During the same period, the model simulations show the PBL decay rates of 53 m h$^{-1}$
13  ([Base]), 102 m h$^{-1}$ ([YSU]), 135 m h$^{-1}$ ([ACM2]), 59 m h$^{-1}$ ([HRRR]) in comparison with the
14  observed 60 m h$^{-1}$. All model simulations generally underestimate the PBL: 180–450 m
15  throughout the day on non-episode days and 10–760 m during the daytime on ozone-episode
16  days. The model underestimations are relative to the observed mixed layer height. The actual
17  PBL biases in the model can be larger or smaller than those underestimations depending on the
18  relationship between the backscatter-defined mixed layer and the thermodynamically defined
19  CBL. Among the four simulations, the [HRRR] best captures the observed mean height and
20  decay rate during the daytime and is thus examined further for the nighttime hours.
21
22  Second, the simulated aerosol backscatter (Fig. 4b, 5b) and potential lapse rates (Fig. 4c, 5c) of
23  the [HRRR] simulation are used to extract the modeled RL heights. The modeled aerosol
24  backscatter shows the volume of the atmosphere in which aerosol species are mixed and
25  dispersed. Substantially stronger backscatter signals are found within the first ~2 km than the
26  free troposphere at 3-4 km aloft with background backscatter of 0.01-0.02 km$^{-1}$ sr$^{-1}$. Therefore,
27  we take the sharpest vertical gradient in the backscatter signal (i.e. the largest first derivative of
28  backscatter) to estimate the modeled mixed layer height. The extracted layers have daytime
29  variations of 1.58±0.13 km and nighttime variations of 1.50±0.06 km during ozone episodes. The
30  modeled aerosol backscatter in Figures 4b and 5b is not equivalent to the ceilometer-measured
31  atmospheric backscatter, which includes both aerosol and molecular backscatter signals in
32  Figures 4a and 5a. The modeled aerosol backscatter is presented here instead of the total
33  atmospheric backscatter because the latter is not diagnosed by the model. The modeled aerosol
34  backscatter is the closest product from the model to denote the modeled mixed layer heights.
35
36  Potential lapse rate or potential temperature gradient $\left(\frac{d\theta}{dz}\right)$, defined thermodynamically as the
37  changes of potential temperature (θ) with height (z), is commonly used to distinguish
38  atmospheric layers according to their instability. Figures 4c and 5c show that the modeled
39  nocturnal PBL consists of a stable SBL, a neutrally stratified RL, and a CI layer during most
40  ozone-episode days. The modeled RL top is identified from where the RL (little or low
41  temperature increases at 0-3 °C km$^{-1}$) shifts to the CI layer (drastic temperature increases at 8-
42  14 °C km$^{-1}$). Therefore, it can be identified by the sharpest gradient in the potential lapse rate,
43  which is 6.6 °C km$^{-1}$ on average. The modeled RL top identified here has a variation of
44  1.39±0.03 km during ozone episodes, slightly lower than the 1.50±0.06 km identified by
45  backscatter.
46

1  The observed RL validates the backscatter-identified and thermodynamically-identified layers
2  from the model in Figure 3j. Model results have a slightly better agreement with the ceilometer
3  defined MLH for the thermodynamically-identified layer, with a correlation coefficient (R) of
4  0.77 and normalized mean bias (NMB) of 3%, than for the backscatter-identified layer, with
5  R=0.72 and NMB=22%, during ozone episodes.
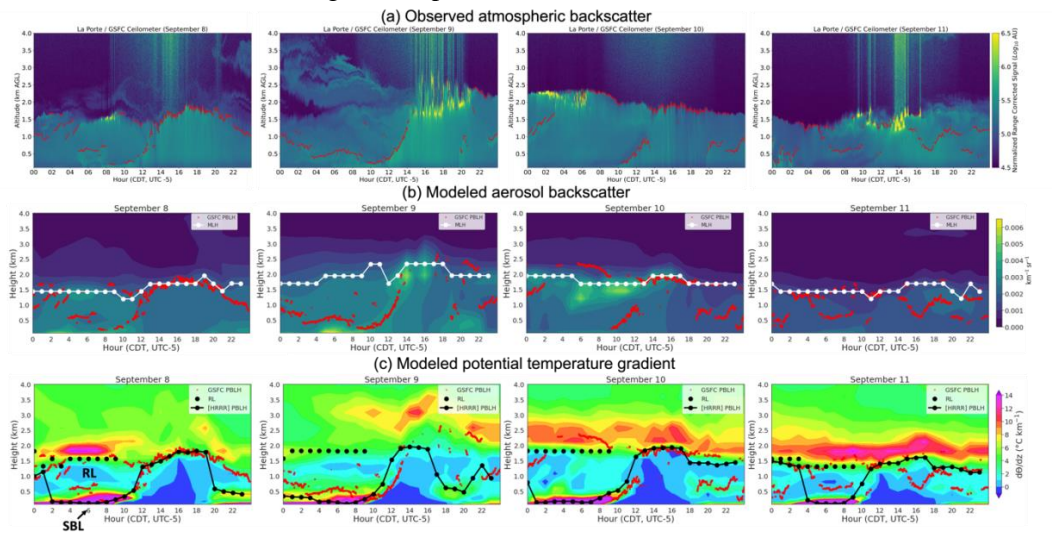


6
7  Figure 4. Observed and modeled heights of lower tropospheric layers at the La Porte site during
8  September 8-11. The contours show (a) ceilometer-observed attenuated atmospheric backscatter
9  signal produced by aerosols and molecules combined at 1064 nm, (b) modeled unattenuated
10  backscatter of aerosols alone at 1000 nm, and (c) modeled potential temperature gradient. Red
11  dots are ceilometer-observed mixed layer. White and black lines are backscatter-defined and
12  thermodynamically-defined mixed layers from the [HRRR] model simulation.
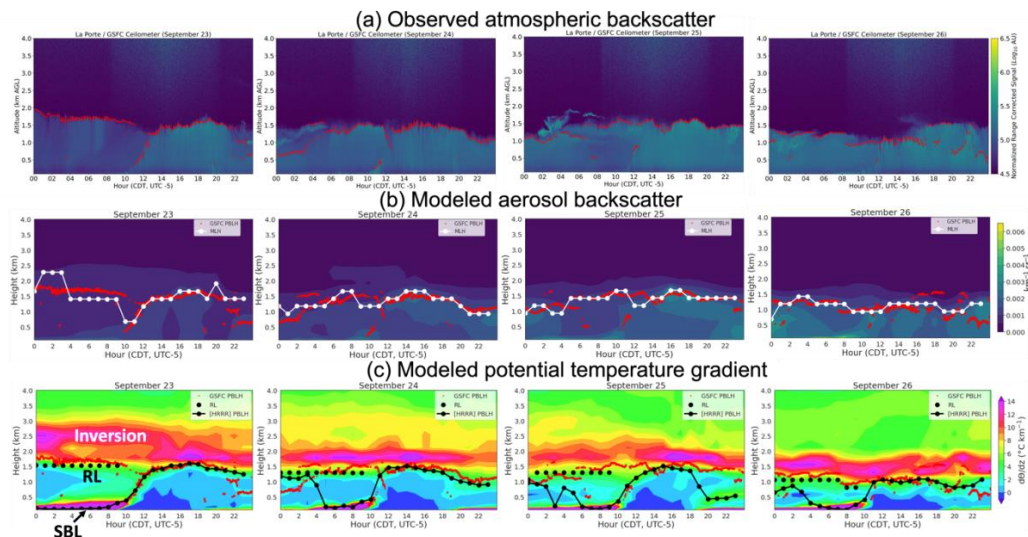


13
14  Figure 5. Same as Figure 4 but for September 23-26.

13

1
2  **4.2 Evaluation with HSRL-2**
3  This section evaluates spatial and temporal (daytime) variations of the modeled PBL heights
4  with the HSRL-2 mixed layer heights. The HSRL-2 collected measurements over the Houston
5  region and the adjacent Galveston Bay typically three times per day for ten days in September
6  2021. As stated in sections above, the mixed layer can represent the PBL under different
7  conditions and we refer to the observed mixed layer heights as observed PBL hereafter in a
8  manner consistent with the modeled equivalents.
9
10  The four model simulations underestimate the observed PBL heights under most conditions.
11  During ozone episodes, the urban Houston region observes the PBL heights of 0.63±0.25 km in
12  the morning (8:00-10:00 CDT), 1.27±0.38 km at noon (11:00-13:00 CDT), and 1.69±0.23 km in
13  the afternoon (14:00-16:00 CDT). The observed heights are consistently lower over water with
14  values of 0.44±0.34 km, 1.14±0.45 km, and 1.43±0.24 km for the three time periods,
15  respectively. Compared to observations, the four model simulations underestimate the heights
16  both over land (by 0.24-0.39 km in the morning, 0.02-0.25 km at noon, and 0.06-0.37 km in the
17  afternoon) and over water (by 0.02-0.25 km in the morning, 0.23-0.59 km at noon and 0.30-0.60
18  km in the afternoon). During non-episode days, the observed PBL height over land is 0.78±0.14
19  km in the morning and 1.07±0.24 km at noon. The morning heights are underestimated by 0.10-
20  0.34 km in the four model simulations, while the noon heights can be either underestimated by
21  0.25-0.37 km or overestimated by 0.05-0.23 km. In contrast to land, the observed height over
22  water is consistently lower with values of 0.57±0.28 km in the morning and 0.65±0.34 km at
23  noon; these are underestimated in the model by 0.03-0.28 km and 0.11-0.21 km, respectively.
24  Among the four simulations, the [HRRR] is the best to reproduce observed values with the
25  minimum model-observation differences under most conditions (e.g. different times and
26  regions), as shown in Table 2.
27
28  The observed land-water differences in PBL heights are larger in the afternoon than in the
29  morning during both ozone-episode and non-episode days. The four model simulations capture
30  the land-water differences better on non-episode days than ozone-episode days. During ozone
31  episodes, the observed mean land-water difference (land minus water) is 0.19 km while the
32  model predicts smaller differences of –0.05~0.06 km in the morning; this is because the model
33  shows consistent larger underestimations over land than water. During noon and afternoon hours,
34  the observed mean land-water differences of 0.13 km and 0.26 km are predicted to be higher in
35  the model with values of 0.32-0.52 km and 0.44-0.56 km, respectively; this is because the model
36  shows consistent smaller underestimations over land than water during this period. During non-
37  episode days, the observed land-water gradients of 0.21 km (morning) and 0.42 km (noon) are
38  simulated to be 0.14-0.22 km and 0.36-0.76 km, respectively. The [ACM2] and the [HRRR]
39  slightly outperform the other two simulations for land-water differences (Table 2).
40
41  One ozone-episode day, September 9, is selected to show the spatial characteristics of different
42  simulations (Fig. 6). The four simulations match the observed mixed layer heights with high
43  spatial correlation at noon (R=0.62-0.77) and in the afternoon (R=0.71-0.82). Among all
44  simulations, the [HRRR] and the [Base] show the best spatial correlations at noon and in the
45  afternoon, respectively. The [HRRR] shows sharp transitions for the different flight legs over
46  urban Houston at noon on September 9 because of drastic changes in the modeled PBL heights at

1  an hourly interval. The morning mixed layer heights can be difficult to retrieve with the
2  influences from multiple layers (e.g. SBL and RL), and they can differ substantially from the
3  thermodynamically-defined PBL. Therefore, we do not expect the model to capture the spatial
4  patterns of mixed layer heights in the morning.
5
6
7  Table 2. Differences of the HSRL-2 mixed layer height and the WRF-GC thermodynamic PBL
8  height during ozone-episode days (September 8-10 and 23-26) and non-episode days (September
9  1 and 3). Land and water are defined by the gray boxes in Figure 6. The bias difference and the
10 root mean square (RMS) difference are calculated by model simulations minus HSRL-2, with the
11 unit of kilometers.

|  | Simulations | Differences | Morning (8:00-10:00 CDT) | | Noon (11:00-13:00 CDT) | | Afternoon (14:00-16:00 CDT) | |
|---|---|---|---|---|---|---|---|---|
|  |  |  | Land | Water | Land | Water | Land | Water |
| Ozone-episode | [Base] | Bias | -0.242 | -0.086 | -0.240 | -0.497 | -0.372 | -0.578 |
|  |  | RMS | 0.381 | 0.318 | 0.436 | 0.668 | 0.464 | 0.696 |
|  | [YSU] | Bias | -0.392 | -0.250 | -0.194 | -0.589 | -0.301 | -0.610 |
|  |  | RMS | 0.488 | 0.409 | 0.406 | 0.785 | 0.409 | 0.807 |
|  | [ACM2] | Bias | -0.294 | -0.167 | -0.076 | -0.471 | -0.278 | -0.457 |
|  |  | RMS | 0.430 | 0.378 | 0.376 | 0.683 | 0.476 | 0.665 |
|  | [HRRR] | Bias | -0.262 | -0.026 | -0.040 | -0.232 | -0.068 | -0.303 |
|  |  | RMS | 0.384 | 0.312 | 0.289 | 0.462 | 0.223 | 0.455 |
|  | Simulations | Differences | Morning | | Noon | | Afternoon | |
|  |  |  | Land | Water | Land | Water | Land | Water |
| Non-episode | [Base] | Bias | -0.211 | -0.218 | -0.243 | -0.276 |  |  |
|  |  | RMS | 0.309 | 0.353 | 0.433 | 0.440 |  |  |
|  | [YSU] | Bias | -0.348 | -0.282 | -0.363 | -0.304 |  |  |
|  |  | RMS | 0.434 | 0.397 | 0.506 | 0.471 |  |  |
|  | [ACM2] | Bias | -0.236 | -0.236 | 0.050 | -0.261 |  |  |
|  |  | RMS | 0.371 | 0.377 | 0.460 | 0.440 |  |  |
|  | [HRRR] | Bias | -0.100 | -0.029 | 0.237 | -0.107 |  |  |
|  |  | RMS | 0.243 | 0.301 | 0.377 | 0.364 |  |  |

12
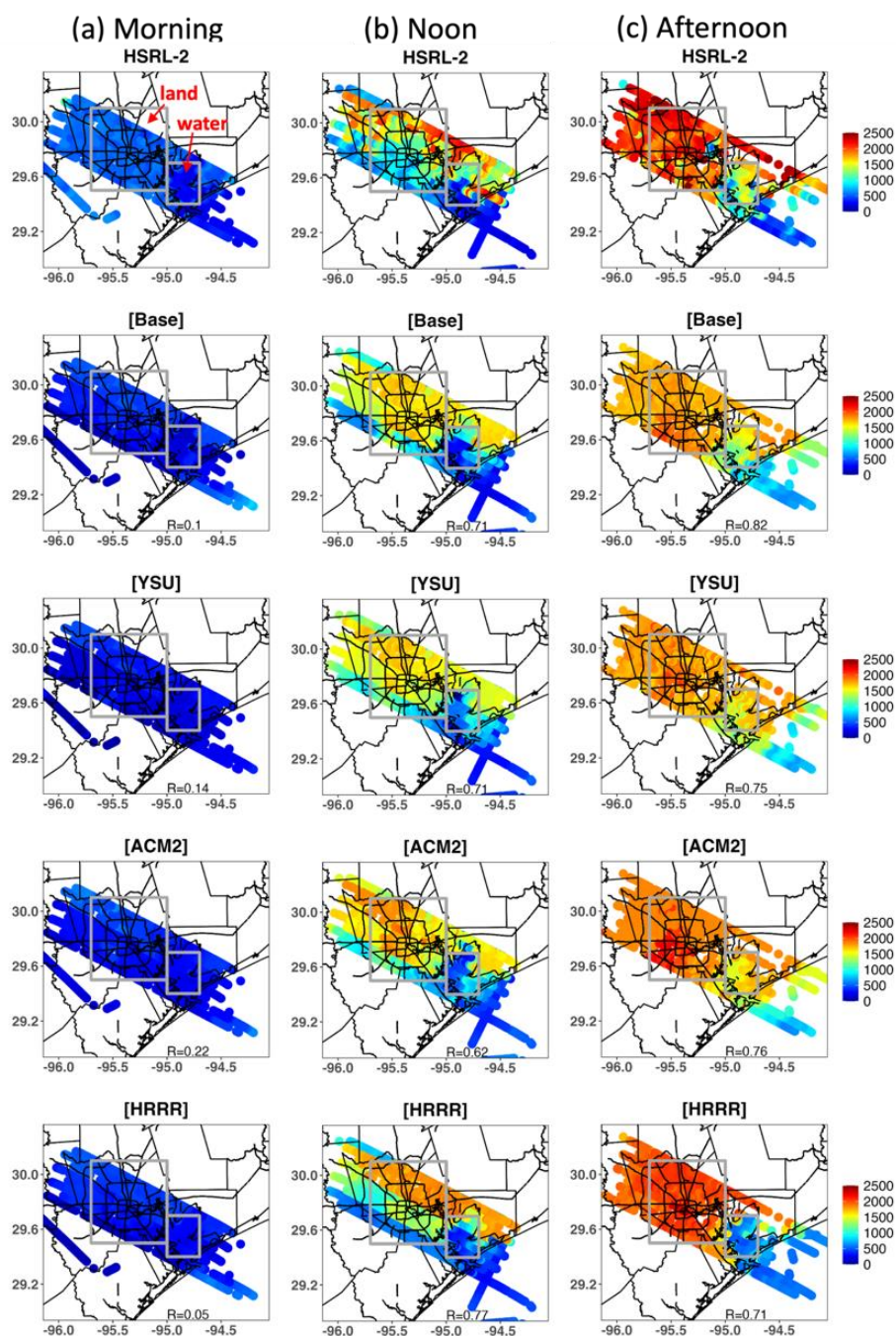
Figure 6. Spatial variabilities of the PBL heights (in meters) from the HSRL-2 and different
WRF-GC simulations (a) in the morning (8:00-10:00 CDT), (b) at noon (11:00-13:00 CDT), and
(c) in the afternoon (14:00-16:00 CDT) of September 9, 2021.

**5. Ozone vertical mixing and distribution**

Boundary layer mixing can bring air aloft towards the surface and vice versa, leading to uneven vertical distribution of ozone which accordingly affects surface ozone concentrations. This section uses independent field measurements at La Porte (incl. TROPOZ, HSRL-2, ozonesondes, a model 49i ozone analyzer, and a CAMS site named La Porte Sylvan Beach) to validate the modeled vertical ozone profiles at three layers, including the lower free troposphere (2-3 km aloft), the boundary layer (0.5-1 km aloft), and the ground level (<50 m). Since the [HRRR] simulation best represents the PBL variations in Section 4, it is used to investigate vertical ozone profiles in this section.

**5.1 Free tropospheric ozone entrainment**

The strength of the CI layer regulates the gas exchange between the FT and the PBL. Strong convection can penetrate a weak CI layer and entrain FT air into the PBL (i.e. entrainment), while a strong CI layer acts as a lid to restrict gas exchange between the PBL and the FT. The potential temperature differences between the top and bottom of the CI layer are often used to indicate the strength of the CI layer and the extent of entrainment processes (Kaser et al., 2017; Morris et al., 2010; Rappenglück et al., 2008). We first identified the modeled CI layers at 1.5–3 km aloft during ozone episodes (Fig. 4c and Fig. 5c), and then calculate the temperature differences in the model between the top and bottom of the CI layers in each day. The corresponding daily inversion strength is 2.3 °C, 2.8 °C, 6.8 °C and 6.4 °C during September 8-11 and 13.6 °C, 7.5 °C, 7.8 °C, and 8.4 °C during September 23-26, respectively. Among these days, September 8 and 9 experienced the weakest inversions. To examine if the modeled inversion strength is representative of the observations, we evaluate the modeled potential temperature profiles with ozonesonde measurements in Fig. 7a. Results show that the model simulates the vertical profiles of potential temperature well across different days with high correlation (R=0.99) and low biases (MB= -0.64 °C ~ -0.17 °C).

Combining the inversion strengths (Fig. 4c; Fig. 5c) and the vertical ozone distributions from the TROPOZ lidar (Fig. 8) helps to identify potential entrainment of the FT air into the underlying PBL on September 8 and 9 at the La Porte site. On September 8, strong convection associated with a rapid CBL growth penetrates the thin and weak inversion at 2 km aloft at around noon (Fig. 4c) and allows the ozone-rich air above to entrain into the CBL, adding to afternoon ozone buildup (Fig. 8a). Similarly, there is no CI layer present overnight from 20 CDT on September 8 to 10 CDT on September 9 (Fig. 4c) and thus long-lasting ozone entrainment into the RL (Fig. 8a). Conversely, a strong and thick inversion at 1.5-3 km decouples the FT and the underlying PBL during September 23-24 (Fig. 5c) and the ozone layer remains aloft at 2-3 km (Fig. 8d). The inversion strength presented here is one way to approach the potential entrainment, follow-up studies can probe into the detailed dynamics. It is also noteworthy that the presented vertical distribution of ozone is also largely shaped by local ozone production in the boundary layer. Since this study is focused on the vertical ozone distribution impacted by mixing between lower free tropospheric layers, the vertical ozone distribution impacted by chemistry and differentiating between the contributions from dynamics and chemistry are outside the scope of this analysis.
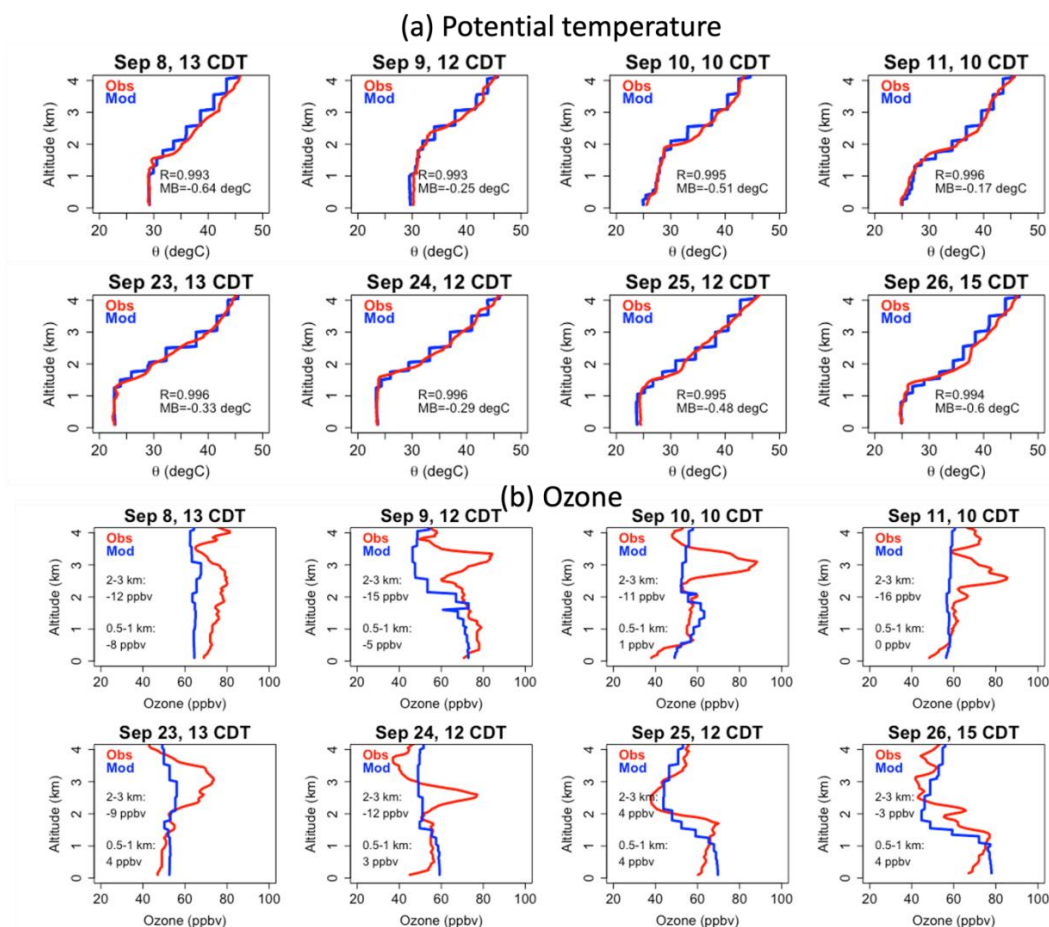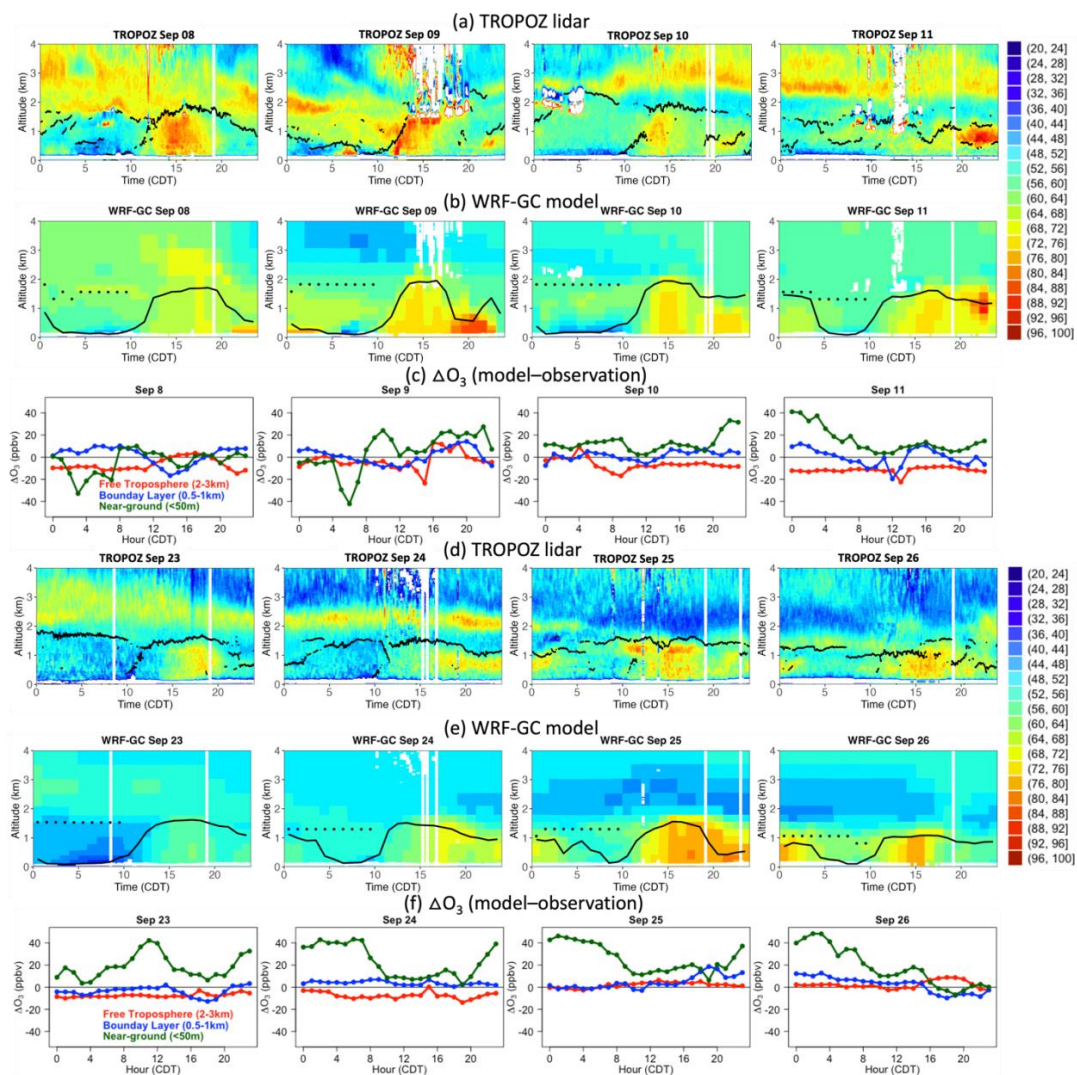
Figure 7. Vertical profiles of (a) potential temperature and (b) ozone from ozonesonde measurements and the WRF-GC [HRRR] simulation at La Porte during September 8-11 and September 23-26.

Figure 8. Time series of the vertical ozone profile from the TROPOZ ozone lidar (a, d) and the
WRF-GC [HRRR] simulation (b, e) at La Porte. Observed and modeled boundary layer heights
are inserted, respectively. Dots represent the modeled residual layer identified in this study. Line
plots (c, f) show ozone differences (model minus observation) at the free troposphere (2-3 km)
and the boundary layer (0.5-1 km) from the TROPOZ as well as the near-ground (< 50m) from
the model 49i ozone analyzer.

1
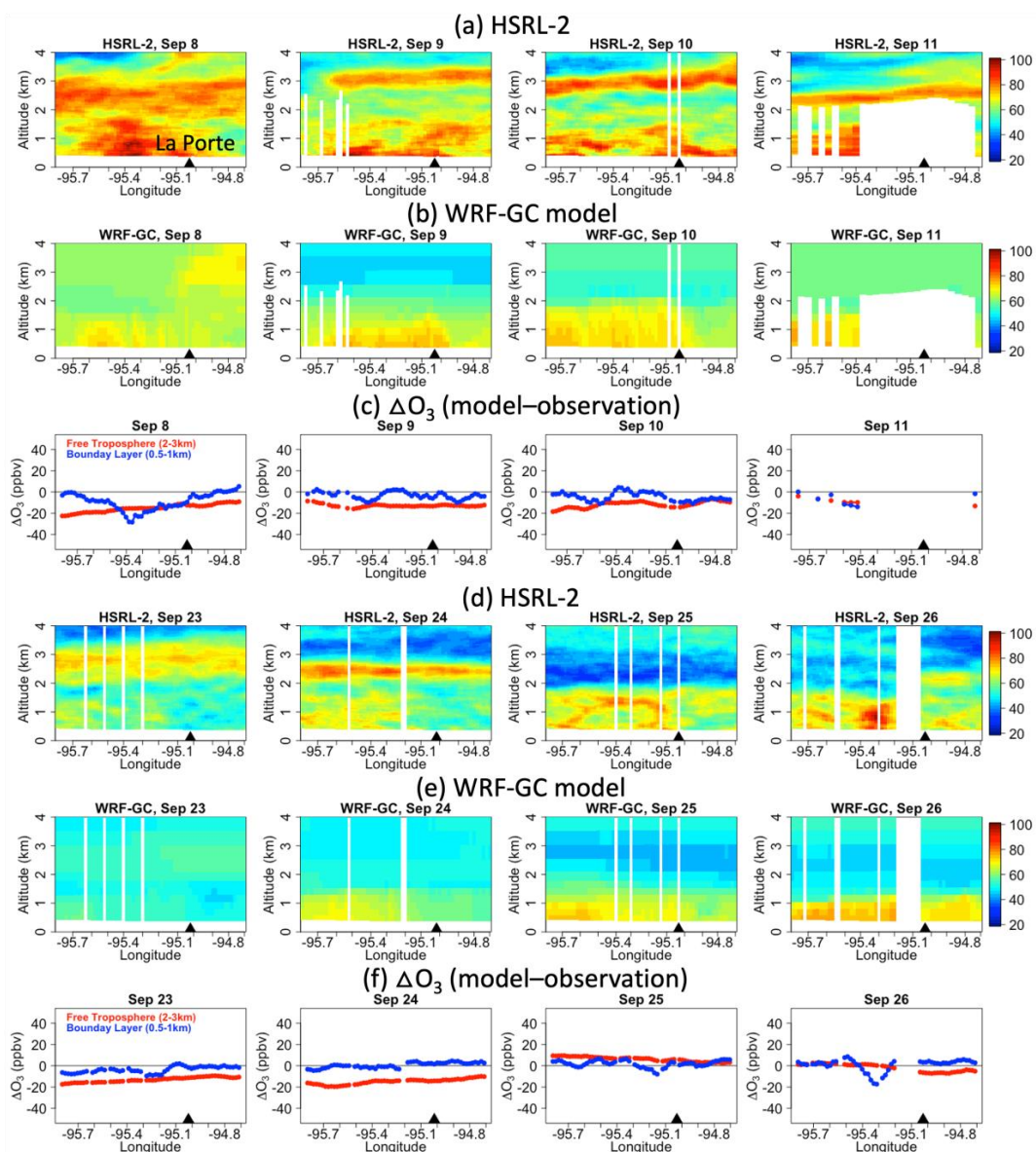


Figure 9. Vertical ozone profiles from (a, d) the HSRL-2 and (b, e) the WRF-GC [HRRR] simulation. The profiles are taken from a flight track (Fig.1) over urban Houston and Galveston Bay at around 11:00-13:00 CDT each day. Line plots (c, f) show ozone differences (model minus observation) at the free troposphere (2-3km) and the boundary layer (0.5-1km).

**5.2 Evaluation of ozone vertical distribution**

Multiple field measurements at La Porte are used to evaluate the modeled vertical ozone distribution at the free troposphere (2-3 km), the boundary layer (0.5-1 km), and the near-ground level (< 50m). According to data availabilities at different levels, the free troposphere and boundary layer are evaluated by the TROPOZ, the HSRL-2, and ozonesondes (Table 3), while the ground level is evaluated by a model 49i ozone analyzer and a CAMS site named La Porte Sylvan Beach (Table 4). To cross compare among multiple measurements, we present the model-observation differences at a common site (La Porte) during a common time slot (11:00-13:00 CDT). Larger ozone differences are found at the near-ground level than for the boundary layer and lower free troposphere (Table 3; Table 4).

As shown in Table 3, the model underestimates the layer of enhanced ozone at 2-3 km aloft in the free troposphere by 9%-21% (TROPOZ), 15%-22% (HSRL-2), and 14%-22% (ozonesondes) at La Porte at 11:00-13:00 CDT during September 8-11 and September 23-24. Unlike most of the campaign's ozone exceedance days, September 25 and 26 do not have an enhanced ozone layer at 2-3 km aloft but have a lower ozone layer relative to the background tropospheric values instead; this low ozone layer is overestimated by 9%-12% on September 25 but underestimated by 3-12% on September 26. Meanwhile, the model underestimates the boundary layer ozone at 0.5-1 km aloft by 6%-10% (TROPOZ), 6%-12% (HSRL-2), and 1%-11% (ozonesondes) during the first ozone episode of September 8-11, but overestimate it by 0%-6% (TROPOZ), 3%-5% (HSRL-2), and 5%-7% (ozonesondes) during the second episode of September 23-26. Above model-observation differences are based on the common site (La Porte) and common time (11:00-13:00 CDT) among different measurements, the temporal (Figures 8c and 8f) and spatial (Figures 9c and 9f) variations of these differences are shown in Figures 8 (TROPOZ) and 9 (HSRL-2).

While free tropospheric and boundary layer ozone are important components of the vertical ozone distribution due to their thickness, the thin layer of near-ground ozone affects human and vegetation health the most and thus receives more attention. In Table 4, the model overestimates near-ground ozone by 6%-24% (model 49i ozone analyzer) and 8%-39% (CAMS La Porte Sylvan Beach) at La Porte at 11:00-13:00 CDT during the two ozone episodes. Figures 8c and 8f show the temporal variations of model-observation differences from the model 49i ozone analyzer. Most near-ground ozone differences occur at night, consistent with the known problem of overestimating nighttime ozone common to many photochemical models (Schnell et al., 2015; Travis et al., 2016; Jaffe et al.,2018). The WRF-GC model adopts a chemical module from GEOS-Chem. Thus, the two share the difficulties replicating nighttime ozone due to reasons such as the insufficient representation of the stratification of multiple nocturnal atmospheric layers, uncertainties in gas exchanges between the residual layer and the underlying surface layer, and difficulties in simulating the timing of changes in PBL dynamics (Travis and Jacob, 2019).

1 Table 3. Absolute (abs.) and percentage (pct.) ozone differences between field measurements and
2 the model at free troposphere and boundary layer at La Porte.

| | | TROPOZ (11-13 CDT) | | HSRL-2 (11-13 CDT) | | Ozonesonde (10-15 CDT) | |
|---|---|---|---|---|---|---|---|
| | | abs. (ppbv) | pct. | abs. (ppbv) | pct. | abs. (ppbv) | pct. |
| Free Troposphere (2-3km aloft) | 8-Sep | -7 | -9% | -12 | -15% | -12 | -16% |
| | 9-Sep | -8 | -13% | -13 | -20% | -15 | -22% |
| | 10-Sep | -8 | -13% | -14 | -21% | -11 | -18% |
| | 11-Sep | -16 | -21% | | | -16 | -21% |
| | 23-Sep | -8 | -13% | -11 | -17% | -9 | -14% |
| | 24-Sep | -9 | -15% | -14 | -22% | -12 | -20% |
| | 25-Sep | 5 | 12% | 6 | 15% | 4 | 9% |
| | 26-Sep | -2 | -3% | -6 | -12% | -3 | -6% |
| Boundary Layer (0.5-1km aloft) | 8-Sep | -5 | -7% | -8 | -11% | -8 | -11% |
| | 9-Sep | -8 | -10% | -5 | -6% | -5 | -7% |
| | 10-Sep | -4 | -6% | -9 | -12% | 1 | 2% |
| | 11-Sep | -5 | -7% | | | -0.4 | -1% |
| | 23-Sep | 0 | 0% | -2 | -3% | 4 | 7% |
| | 24-Sep | 2 | 4% | 2 | 4% | 3 | 5% |
| | 25-Sep | 1 | 2% | 2 | 3% | 4 | 7% |
| | 26-Sep | 4 | 6% | 3 | 5% | 4 | 5% |

3

4 Table 4. Absolute (abs.) and percentage (pct.) ozone differences between field measurements and
5 the model at the near-ground level at La Porte.

| | | Model 49i (11-13 CDT) | | CAMS La Porte Sylvan Beach (11-13 CDT) | |
|---|---|---|---|---|---|
| | | abs. (ppbv) | pct. | abs. (ppbv) | pct. |
| Near-ground (< 50m) | 8-Sep | 4 | 7% | | |
| | 9-Sep | 8 | 12% | | |
| | 10-Sep | 4 | 6% | 16 | 31% |
| | 11-Sep | 6 | 9% | 15 | 29% |
| | 23-Sep | 4 | 9% | 4 | 8% |
| | 24-Sep | 10 | 20% | 15 | 34% |
| | 25-Sep | 13 | 24% | 18 | 33% |
| | 26-Sep | 12 | 21% | 21 | 39% |

6

**6. Conclusion**

We used ground-based and aircraft observations collected during the TRACER-AQ campaign in September 2021 to evaluate WRF-GC simulation of the PBL height and ozone in Houston, including two ozone episodes characterized by MDA8 ozone exceeding 70 ppbv. The combined suite of ground-based and airborne meteorological and chemical observations are critical in thoroughly evaluating the spatial and temporal variations of the PBL and vertical ozone distributions during multi-day ozone episodes, as presented in this work.

The modeled PBL heights are evaluated with mixed layer heights retrieved by the ground-based ceilometer and the airborne HSRL-2. Compared with the ceilometer, the four model simulations of [Base], [YSU], [ACM2], and [HRRR] generally underestimate the PBL heights to different extents: 180–450 m throughout the day on non-episode days and 10–760 m during the daytime on ozone-episode days. As the best simulation, the [HRRR] captures the diurnal variations during non-episode days (R=0.88; NMB=-21%). Standard models do not diagnose RL heights, unlike ceilometers. Therefore, we separately identified the modeled RL following the practices using aerosol backscatter signals and potential temperature gradients during the ozone episodes. As a result, the diurnal variation of the thermodynamically-identified layer (R=0.77; NMB=3%) compares slightly better than that of the backscatter-identified layer (R=0.72; NMB=22%) during ozone episodes. Meanwhile, when compared with the HSRL-2, the four simulations underestimates PBL heights by 20-390 m over the urban Houston region and by 20-600 m over the adjacent Galveston Bay during ozone episodes. On non-episode days, the PBL heights over urban region are either underestimated by 100-370 m or overestimated by 50-230 m and those over the Bay are underestimated by 30-210 m by the four simulations. On both ozone-episode and non-episode days, the observed land-water differences in PBL heights are larger in the afternoon than in the morning; the model captures such daytime trends. The four model simulations capture the land-water differences better on non-episode days than ozone-episode days.

We evaluated the vertical ozone distribution with multiple field measurements, including TROPOZ, HSRL-2, ozonesonde, a model 49i ozone analyzer, and a CAMS site named La Porte Sylvan Beach. Evaluations were done at three lower tropospheric layers: the free troposphere (2-3 km aloft), the boundary layer (0.5-1 km aloft), and the ground level (< 50 m aloft). As a result, the model underestimates the high ozone layer in the free troposphere by 9%-21% (TROPOZ), 15%-22% (HSRL-2), and 14%-22% (ozonesondes) on most ozone-episode days. The boundary layer ozone is underestimated by 6%-10% (TROPOZ), 6%-12% (HSRL-2), and 1%-11% (ozonesondes) during September 8-11, but overestimated by 0%-6% (TROPOZ), 3%-5% (HSRL-2), and 5%-7% (ozonesondes) during September 23-26. Meanwhile, the model overestimates near-ground ozone by 6%-24% (model 49i ozone analyzer) and 8%-39% (CAMS La Porte Sylvan Beach) during the two ozone episodes.

Based on these evaluations, we summarized model limitations that prevent more accurate simulation of PBL heights and the vertical ozone distribution during TRACER-AQ. The first limitation is the single-layer PBL representation. The WRF model only diagnoses the SBL at night, despite the model simulating different physical and thermodynamic properties of multiple nocturnal layers above the SBL. For example, the RL is not identified by the model as a standard diagnosis; this prevents the direct comparison of the model outputs with the observed RL at

1  night. The second limitation is the underestimation of the layer of enhanced ozone 2-3 km aloft
2  in the free troposphere that was often present on ozone-episode days during the campaign. Given
3  its height of 2-3 km and a lifetime of around a week, the layer of enhanced ozone was likely
4  transported into Houston by synoptic flows of cold fronts from the north. The
5  underrepresentation of the synoptic layer of enhanced ozone affects model representations across
6  regions horizontally and atmospheric layers vertically, making it particularly important to model
7  vertical ozone distributions and the effects of entrainment accurately.
8
9  Our findings of the model limitations have implications for the predictivity of ozone's vertical
10  mixing and distribution across different modeling systems. For example, WRF is widely used in
11  various meteorology-chemistry coupling systems with different treatments of boundary layer
12  mixing. In WRF-Chem, boundary layer mixing in the chemistry part uses a mixing coefficient
13  originating in WRF such that the boundary layer mixing calculations in the meteorology and
14  chemistry parts share the same set of coefficients. In WRF-GC, the chemistry part from GEOS-
15  Chem only takes the PBL height from WRF as the maximum height for boundary layer mixing
16  and conducts independent calculations of boundary layer mixing from WRF. Unlike online
17  coupled WRF-Chem and WRF-GC, WRF is offline coupled to CAMx in the WRF-CAMx
18  system, and the boundary layer mixing in the chemistry part of CAMx is subject to WRF output
19  frequency instead of the native transport time step in WRF. Thus, it is essential to understand
20  how the model limitation of a single-layer PBL representation affects boundary layer mixing in
21  chemical simulations among different meteorology-chemistry coupling systems. Follow-up
22  studies to this work will address these aspects with detailed model intercomparisons.
23

24  **Code availability.** WRF-GC is a free and open source model (http://wrf.geos-chem.org; last
25  access: 29 May 2023) (Lin et al., 2019; Feng et al., 2021). The two parent models, WRF and
26  GEOS-Chem, are also open source and can be obtained from their developers at
27  https://github.com/wrf-model/WRF (last access: 29 May 2023) and http://www.geos-chem.org
28  (last access: 29 May 2023), respectively. The version of WRF-GC (v2.0) described in this paper
29  couples WRF v3.9.1.1 and GEOS-Chem v12.7.2 and is archived in Zenodo at
30  https://doi.org/10.5281/zenodo.4395258 (last access: 29 May 2023).

31  **Data availability.** All observation datasets, model configuration files, model boundary
32  conditions, model input files, and scripts used in this paper are archived in Zenodo at
33  https://doi.org/10.5281/zenodo.7983449 (last access: 29 May 2023).

34
35  **Author contributions.** XL and YW conceived the research idea. XL wrote the initial draft of the
36  paper and performed the analyses and model simulations. JF, TG, and SA provided the shipborne
37  data. JS, MR, and LT provided the TROPOZ and ceilometer data. GG and TB provided the
38  LMOL data. PW and JS provided the ozonesonde data. JH, TS, AJS, and MF provided the
39  HSRL-2 data. All authors contributed to the interpretation of the results and the preparation of
40  the paper.
41
42  **Competing interests.** The contact author has declared that none of the authors has any
43  competing interests.

11
12  **References**
13

14  Banta, R. M., Senff, C. J., Nielsen-Gammon, J., Darby, L. S., Ryerson, T. B., Alvarez, R. J.,
15  Sandberg, S. R., Williams, E. J., and Trainer, M.: A bad air day in Houston, B. Am. Meteorol.
16  Soc., 86, 657–669, doi:10.1175/bams-86-5-657, 2005.
17
18  Banta, R. M., Senff, C. J., Alvarez, R. J., Langford, A. O., Parrish, D. D., Trainer, M. K., Darby,
19  L. S., Hardesty, R. M., Lambeth, B., Neuman, J. A., Angevine, W. M., Nielsen-Gammon, J.,
20  Sandberg, S. P., and White, A. B.: Dependence of daily peak O3 concentrations near Houston,
21  Texas on environmental factors: wind speed, temperature, and boundary-layer depth, Atmos.
22  Environ., 45, 162–173, doi:10.1016/j.atmosenv.2010.09.030, 2011.

23  Bernier, C., Wang, Y., Estes, M., Lei, R., Jia, B., Wang, S., and Sun, J.: Clustering Surface
24  Ozone Diurnal Cycles to Understand the Impact of Circulation Patterns in Houston, TX, J.
25  Geophys. Res.-Atmos, 124, 13457–13474., https://doi.org/10.1029/2019jd031725, 2019.

26  Bernier, C., Wang, Y., Gronoff, G., Berkoff, T., Knowland, K.E., Sullivan, J.T., Delgado, R.,
27  Caicedo, V. and Carroll, B.: Cluster-based characterization of multi-dimensional tropospheric
28  ozone variability in coastal regions: an analysis of lidar measurements and model results, Atmos.
29  Chem. Phys., 15313–15331, https://doi.org/10.5194/acp-22-15313-2022, 2022.

30  Burton, S. P., Hair, J. W., Kahnert, M., Ferrare, R. A., Hostetler, C. A., Cook, A. L., Harper, D.
31  B., Berkoff, T. A., Seaman, S. T., Collins, J. E., Fenn, M. A., and Rogers, R. R.: Observations of
32  the spectral dependence of linear particle depolarization ratio of aerosols using NASA Langley
33  airborne High Spectral Resolution Lidar, Atmos. Chem. Phys., 15, 13453–13473,
34  https://doi.org/10.5194/acp-15-13453-2015, 2015.

35  Caicedo, V., Rappenglück, B., Lefer, B., Morris, G., Toledo, D., and Delgado, R.: Comparison
36  of aerosol lidar retrieval methods for boundary layer height detection using ceilometer aerosol
37  backscatter data, Atmos. Meas. Tech., 10, 1609–1622, https://doi.org/10.5194/amt-10-1609-
38  2017, 2017.

1   Caicedo, V., Rappenglück, B., Cuchiara, G., Flynn, J., Ferrare, R., Scarino, A., Berkoff, T.,
2   Senff, C., Langford, A., and Lefer, B.: Bay breeze and sea breeze circulation impacts on the
3   planetary boundary layer and air quality from an observed and modeled DISCOVER-AQ Texas
4   case study, J. Geophys. Res.-Atmos., 124, 7359–7378, https://doi.org/10.1029/2019JD030523,
5   2019.
6
7   Caicedo, V., Delgado, R., Sakai, R., Knepp, T., Williams, D., Cavender, K., Lefer, B., and
8   Szykman, J.: An Automated Common Algorithm for Planetary Boundary Layer Retrievals Using
9   Aerosol Lidars in Support of the U.S. EPA Photochemical Assessment Monitoring Stations
10  Program, J. Atmos. Oceanic Technol., 37, 1847–1864, https://doi.org/10.1175/JTECH-D-20-
11  0050.1, 2020.
12
13  Caputi, D. J., Faloona, I., Trousdell, J., Smoot, J., Falk, N., and Conley, S.: Residual layer ozone,
14  mixing, and the nocturnal jet in California's San Joaquin Valley, Atmos. Chem. Phys., 19, 4721–
15  4740, https://doi.org/10.5194/acp-19-4721-2019, 2019.
16
17  Chen, F. and Dudhia, J.: Coupling an advanced land surface–hydrology model with the Penn
18  State–NCAR MM5 modeling system. Part II: Preliminary model validation, Monthly Weather
19  Review, 129(4), 587-604, 2001.
20
21  Cuchiara, G. C., Li, X., Carvalho, J., and Rappenglück, B.: Intercomparison of planetary
22  boundary layer parameterization and its impacts on surface ozone concentration in the
23  WRF/Chem model for a case study in Houston/Texas, Atmos. Environ., 96, 175– 185,
24  doi:10.1016/j.atmosenv.2014.07.013, 2014.
25
26  Dacic, N., Sullivan, J. T., Knowland, K. E., Wolfe, G. M., Oman, L. D., Berkoff, T. A., and
27  Gronoff, G. P.: Evaluation of NASA's high-resolution global composition simulations:
28  Understanding a pollution event in the Chesapeake Bay during the summer 2017 OWLETS
29  campaign, Atmos. Environ., 222, 117133, https://doi.org/10.1016/j.atmosenv.2019.117133,
30  2020.
31
32  Dreessen, J., Sullivan, J., and Delgado, R.: Observations and impacts of transported Canadian
33  wildfire smoke on ozone and aerosol air quality in the Maryland region on June 9–12, 2015, J.
34  Air Waste Manage., 66, 842–862, https://doi.org/10.1080/10962247.2016.1161674, 2016.
35
36  Feng, X., Lin, H., Fu, T.-M., Sulprizio, M. P., Zhuang, J., Jacob, D. J., Tian, H., Ma, Y., Zhang,
37  L., Wang, X., Chen, Q., and Han, Z.: WRF-GC (v2.0): online two-way coupling of WRF
38  (v3.9.1.1) and GEOS-Chem (v12.7.2) for modeling regional atmospheric chemistry–meteorology
39  interactions, Geosci. Model Dev., 14, 3741–3768, https://doi.org/10.5194/gmd-14-3741-2021,
40  2021.
41
42  Geiß, A., Wiegner, M., Bonn, B., Schäfer, K., Forkel, R., von Schneidemesser, E., Münkel, C.,
43  Chan, K. L., and Nothard, R.: Mixing layer height as an indicator for urban air quality?, Atmos.
44  Meas. Tech., 10, 2969–2988, https://doi.org/10.5194/amt10-2969-2017, 2017.
45

1   Gronoff, G., Robinson, J., Berkoff, T., Swap, R., Farris, B., Schroeder, J., Halliday, H. S.,
2   Knepp, T., Spinei, E., Car- rion, W., Adcock, E. E., Johns, Z., Allen, D., and Pip- pin, M.: A
3   Method for Quantifying near Range Point Source Induced O3 Titration Events Using Co-
4   Located Li- dar and Pandora Measurements, Atmos. Environ., 204, 43–52,
5   https://doi.org/10.1016/j.atmosenv.2019.01.052, 2019.
6
7   Gronoff, G., Berkoff, T., Knowland, K. E., Lei, L., Shook, M., Fab- bri, B., Carrion, W., and
8   Langford, A. O.: Case study of strato- spheric Intrusion above Hampton, Virginia: lidar-
9   observation and modeling analysis, Atmos. Environ., 259, 1352–2310,
10  https://doi.org/10.1016/j.atmosenv.2021.118498, 2021.
11
12  Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., and
13  Wang, X.: The Model of Emissions of Gases and Aerosols from Nature version 2.1
14  (MEGAN2.1): an extended and updated framework for modeling biogenic emissions, Geosci.
15  Model Dev., 5, 1471–1492, doi:10.5194/gmd-5- 1471-2012, 2012.
16
17  Haman, C. L., Couzo, E., Flynn, J. H., Vizuete, W., Heffron, B., and Lefer, B. L.: Relationship
18  between boundary layer heights and growth rates with ground-level ozone in Houston, Texas, J.
19  Geophys. Res.-Atmos., 119, 6230–6245, https://doi.org/10.1002/2013jd020473, 2014.
20
21  Hair, J. W., Hostetler, C. A., Cook, A. L., Harper, D. B., Ferrare, R. A., Mack, T. L., Welch, W.,
22  Izquierdo, L. R., and Hovis, F. E.: Airborne High Spectral Resolution lidar for profiling aerosol
23  optical properties, Appl. Opt., 47, 6734–6752, 2008.
24
25  Hair, J., Hostetler, C., Cook, A., Harper, D., Notari, A., Fenn, M., Newchurch, M., Wang, L.,
26  Kuang, S., Knepp, T., Burton, S., Ferrare, R., Butler, C., Collins, J., and Nehrir, A.: New
27  capability for ozone dial profiling measurements in the troposphere and lower stratosphere from
28  aircraft, EPJ Web Conf., 176, 01006, https://doi.org/10.1051/epjconf/201817601006, 2018.
29
30  Hong, S.-Y., Noh, Y., and Dudhia, J.: A new vertical diffusion package with an explicit
31  treatment of entrainment processes, Mon. Weather Rev., 134, 2318–2341,
32  https://doi.org/10.1175/MWR3199.1, 2006.
33
34  Hong, S. Y. and Lim, J. O. J.: The WRF Single-Moment 6-Class Microphysics Scheme
35  (WSM6), J. Korean Meteor. Soc., 42, 129– 151, 2006.
36
37  Hu, X. M., Nielsen-Gammon, J. W., and Zhang, F.: Evaluation of three planetary boundary layer
38  schemes in the WRF model, J. Appl. Meteorol. Clim., 49, 1831–1844,
39  doi:10.1175/2010JAMC2432.1, 2010.
40
41  Hudman, R. C., Moore, N. E., Mebust, A. K., Martin, R. V., Russell, A. R., Valin, L. C., and
42  Cohen, R. C.: Steps towards a mechanistic model of global soil nitric oxide emissions:
43  implementation and space based-constraints, Atmos. Chem. Phys., 12, 7779–7795,
44  https://doi.org/10.5194/acp-12-7779-2012, 2012.
45

1  Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W.
2  D.: Radiative forcing by longlived greenhouse gases: Calculations with the AER radiative
3  transfer models, J. Geophys. Res.-Atmos., 113, D13103, https://doi.org/10.1029/2008JD009944,
4  2008.
5
6  Jaffe, D.: Relationship between Surface and Free Tropospheric Ozone in the Western U.S,
7  Environ. Sci. Technol, 45, 432–438, doi:10.1021/es1028102, 2011.
8
9  Jaffe, D. A., Cooper, O. R., Fiore, A. M., Henderson, B. H., Tonnesen, G. S., Russell, A. G.,
10 Henze, D. K., Langford, A. O., Lin, M., and Moore, T.: Scientific assessment of background
11 ozone over the U.S.: Implications for air quality management, Elem. Sci. Anth., 6, p. 56,
12 https://doi.org/10.1525/elementa.309, 2018.
13
14 Johnson, M., Kuang, S., Wang, L., and Newchurch, M.: Evaluating summer-time ozone
15 enhancement events in the southeast United States, Atmosphere, 7, 108,
16 https://doi.org/10.3390/atmos7080108, 2016.
17
18 Kaser, L., Patton, E. G., Pfister, G. G., Weinheimer, A. J., Montzka, D. D., Flocke, F.,
19 Thompson, A. M., Stauffer, R. M., and Halliday, H. S.: The effect of entrainment through
20 atmospheric boundary layer growth on observed and modeled surface ozone in the Colorado
21 Front Range, J. Geophys. Res.-Atmos., 122, 6075– 6093, https://doi.org/10.1002/2016JD026245,
22 2017.
23
24 Knepp, T. N., Szykman, J. J., Long, R., Duvall, R. M., Krug, J., Beaver, M., Cavender, K.,
25 Kronmiller, K., Wheeler, M., Delgado, R., Hoff, R., Berkoff, T., Olson, E., Clark, R., Wolfe, D.,
26 Van Gilst, D., and Neil, D.: Assessment of mixed-layer height estimation from single-
27 wavelength ceilometer profiles, Atmos. Meas. Tech., 10, 3963–
28 3983, https://doi.org/10.5194/amt-10-3963-2017, 2017.
29
30 Knote, C., Tuccella, P., Curci, G., Emmons, L., Orlando, J. J., Madronich, S., Baró, R., Jiménez-
31 Guerrero, P., Luecken, D., Hogrefe, C., Forkel, R., Werhahne, J., Hirtl, M., Pérez, J., José, R.,
32 Giordano, L., Brunner, D., Yahya, K., and Zhang, Y.: Influence of the choice of gas-phase
33 mechanism on predictions of key gaseous pollutants during the AQMEII phase-2
34 intercomparison, Atmos. Environ., 115, 553– 568,
35 https://doi.org/10.1016/j.atmosenv.2014.11.066, 2015.
36
37 Jensen, M. P., Flynn, J. H., Judd, L. M., Kollias, P., Kuang, C., Mcfarquhar, G., Nadkarni, R.,
38 Powers, H., and Sullivan, J.: A Succession of Cloud, Precipitation, Aerosol, and Air Quality
39 Field Experiments in the Coastal Urban Environment, B. Am. Meteorol. Soc., 103, 103–105,
40 2022.
41
42 Kotsakis, A., Sullivan, J.T., Hanisco, T.F., Swap, R.J., Caicedo, V., Berkoff, T.A., Gronoff, G.,
43 Loughner, C.P., Ren, X., Luke, W.T. and Kelley, P.: Sensitivity of total column NO2 at a marine
44 site within the Chesapeake Bay during OWLETS-2, Atmos. Environ., 277, 119063, 2022.
45

Kuik, F., Lauer, A., Churkina, G., Denier van der Gon, H. A. C., Fenner, D., Mar, K. A., and Butler, T. M.: Air quality modelling in the Berlin–Brandenburg region using WRF-Chem v3.7.1: sensitivity to resolution of model grid and input data, Geosci. Model Dev., 9, 4339–4363, https://doi.org/10.5194/gmd-9-4339-2016, 2016.

Lefer, B., Rappenglück, B., Flynn, J., and Haman, C.: Photochemical and meteorological relationships during the Texas-II Radical and Aerosol Measurement Project (TRAMP), Atmos. Environ., 44, 4005–4013, doi:10.1016/j.atmosenv.2010.03.011, 2010.

Li, W., Wang, Y., Bernier, C., and Estes, M.: Identification of Sea Breeze Recirculation and Its Effects on Ozone in Houston, TX, during Discover-Aq 2013, J. Geophys. Res.-Atmos, 125, e2020JD033165, https://doi.org/10.1029/2020jd033165, 2020.

Li, D., Wu, Y., Gross, B. and Moshary, F.: Capabilities of an Automatic Lidar Ceilometer to Retrieve Aerosol Characteristics within the Planetary Boundary Layer, Remote Sensing, 13(18), 3626, 2021.

Lin, H., Feng, X., Fu, T.-M., Tian, H., Ma, Y., Zhang, L., Jacob, D. J., Yantosca, R. M., Sulprizio, M. P., Lundgren, E. W., Zhuang, J., Zhang, Q., Lu, X., Zhang, L., Shen, L., Guo, J., Eastham, S. D., and Keller, C. A.: WRF-GC (v1.0): online coupling of WRF (v3.9.1.1) and GEOS-Chem (v12.2.1) for regional atmospheric chemistry modeling – Part 1: Description of the one-way model, Geosci. Model Dev., 13, 3241–3265, https://doi.org/10.5194/gmd-13-3241-2020, 2020.

Lufft. Manual ceilometer CHM 15k Nimbus. Revision R07, 2016.

McMillan, W. W., Pierce, R. B., Sparling, L. C., Osterman, G., McCann, K., Fischer, M. L., Rappenglück, B., Newsom, R., ̈ Turner, D., Kittaka, C., Evans, K., Biraud, S., Lefer, B., Andrews, A., and Oltmans, S.: An observational and modeling strategy to investigate the impact of remote sources on local air quality: A Houston, Texas case study from the Second Texas Air Quality Study (TexAQS II), J. Geophys. Res., 115, D01301, doi:10.1029/2009JD011973, 2010.

Morrison, H., Thompson, G., and Tatarskii, V.: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: comparison of one- and two-moment schemes, Mon. Weather Rev., 137, 991–1007, 2009.

Morris, G. A., Ford, B., Rappenglück, B., Thompson, A. M., Mefferd, A., Ngan, F., and Lefer, B.: An evaluation of the interaction of morning residual layer and afternoon mixed layer ozone in Houston using ozonesonde data, Atmos. Environ., 44, 4024– 4034, 2010.

Murray, L. T., Jacob, D. J., Logan, J. A., Hudman, R. C., and Koshak, W. J.: Optimized regional and interannual variability of lightning in a global chemical transport model constrained by LIS/OTD satellite data, J. Geophys. Res., 117, D20307, https://doi.org/10.1029/2012jd017934, 2012.

1  Nakanishi, M. and Niino, H.: Development of an improved turbulence closure model for the
2  atmospheric boundary layer, J. Meteorol. Soc. Jpn., 87, 895–912,
3  https://doi.org/10.2151/jmsj.87.895, 2009.
4
5  Otte, T. L.: The impact of nudging in the meteorological model for retrospective air quality
6  simulations. Part I: Evaluation against national observation networks, J. Appl. Meteor. Climatol.,
7  47, 1853–1867, 2008.
8
9  Parrish, D. D., Aikin, K. C., Oltmans, S. J., Johnson, B. J., Ives, M., and Sweeny, C.: Impact of
10 transported background ozone inflow on summertime air quality in a California ozone
11 exceedance area, Atmos. Chem. Phys., 10, 10093–10109, doi:10.5194/acp10-10093-2010, 2010.
12
13 Pleim, J. E.: A combined local and nonlocal closure model for the atmospheric boundary layer,
14 Part I: model description and testing, J. Appl. Meteor. Clim., 46, 1383–1395, 2007a.
15
16 Pleim, J. E.: A combined local and nonlocal closure model for the atmospheric boundary layer,
17 Part II: application and evaluation in a mesoscale meteorological model, J. Appl. Meteor. Clim.,
18 46, 1396–1409, 2007b.
19
20 Rappenglück, B., Perna, R., Zhong, S., and Morris, G. A.: An analysis of the vertical structure of
21 the atmosphere and the upper-level meteorology and their impact on surface ozone levels in
22 Houston, Texas, J. Geophys. Res.-Atmos., 113, D17315, doi:10.1029/2007JD009745, 2008.
23
24 Rizza, U., Barnaba, F., Miglietta, M. M., Mangia, C., Di Liberto, L., Dionisi, D., Costabile, F.,
25 Grasso, F., and Gobbi, G. P.: WRF-Chem model simulations of a dust outbreak over the central
26 Mediterranean and comparison with multi-sensor desert dust observations, Atmos. Chem. Phys.,
27 17, 93–115, https://doi.org/10.5194/acp-17-93-2017, 2017.
28
29 Scarino, A. J., Obland, M. D., Fast, J. D., Burton, S. P., Ferrare, R. A., Hostetler, C. A., Berg, L.
30 K., Lefer, B., Haman, C., Hair, J. W., Rogers, R. R., Butler, C., Cook, A. L., and Harper, D. B.:
31 Comparison of mixed layer heights from airborne high spectral resolution lidar, ground-based
32 measurements, and the WRFChem model during CalNex and CARES, Atmos. Chem. Phys., 14,
33 5547–5560, doi:10.5194/acp-14-5547-2014, 2014.
34
35 Schnell, J. L., Prather, M. J., Josse, B., Naik, V., Horowitz, L. W., Cameron-Smith, P.,
36 Bergmann, D., Zeng, G., Plummer, D. A., Sudo, K., Nagashima, T., Shindell, D. T., Faluvegi,
37 G., and Strode, S. A.: Use of North American and European air quality networks to evaluate
38 global chemistry–climate modeling of surface ozone, Atmos. Chem. Phys., 15, 10581–10596,
39 doi:10.5194/acp-15-10581-2015, 2015.
40
41 Stull, R.: An introduction to Atmospheric Boundary Layer Meteorology, Kluwer Academic
42 Publishers, Dordrecht, The Netherlands, 1988.
43
44 Sullivan, J. T., McGee, T. J., Sumnicht, G. K., Twigg, L. W., and Hoff, R. M.: A mobile
45 differential absorption lidar to measure sub-hourly fluctuation of tropospheric ozone profiles in

1    the Baltimore–Washington, D.C. region, Atmos. Meas. Tech., 7, 3529–
2    3548, https://doi.org/10.5194/amt-7-3529-2014, 2014.
3
4    Sullivan, J. T., McGee, T. J., Thompson, A. M., Pierce, R. B., Sumnicht, G. K., Twigg, L. W.,
5    Eloranta, E., and Hoff, R. M.: Characterizing the lifetime and occurrence of stratospheric-
6    tropospheric exchange events in the rocky mountain region using high-resolution ozone
7    measurements, J. Geophys. Res.-Atmos., 120, 12410–12424, 2015.
8
9    Sullivan, J. T., Rabenhorst, S. D., Dreessen, J., McGee, T. J., Delgado, R., Twigg, L., and
10   Sumnicht, G.: Lidar observations revealing transport of O3 in the presence of a nocturnal low-
11   level jet: Regional implications for "next-day" pollution, Atmos. Environ., 158, 160–171, 2017.
12
13   Sullivan, J. T., Berkoff, T., Gronoff, G., Knepp, T., Pippin, M., Allen, D., Twigg, L., Swap, R.,
14   Tzortziou, M., Thompson, A. M., and Stauffer, R. M.: The ozone water–land environmental
15   transition study: An innovative strategy for understanding Chesapeake Bay pollution events, B.
16   Am. Meteorol. Soc., 100, 291–306, 2019.
17
18   Sullivan, J.T., Apituley, A., Mettig, N., Kreher, K., Knowland, K.E., Allaart, M., Piters, A., Van
19   Roozendael, M., Veefkind, P., Ziemke, J.R. and Kramarova, N.: Tropospheric and stratospheric
20   ozone profiles during the 2019 TROpomi vaLIdation eXperiment (TROLIX-19), Atmos. Chem.
21   Phys., 22(17), 11137-11153, 2022.
22
23   Tangborn, A., Demoz, B., Carroll, B. J., Santanello, J., and Anderson, J. L.: Assimilation of lidar
24   planetary boundary layer height observations, Atmos. Meas. Tech., 14, 1099–1110,
25   https://doi.org/10.5194/amt-14-1099-2021, 2021.
26
27   Tiedtke, M.: A Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale
28   Models, Monthly Weather Review, 117, 1779–1800, 1989.
29
30   Tucker, S. C., Banta, R. M., Langford, A. O., Senff, C. J., Brewer, W. A., Williams, E. J.,
31   Lerner, B. M., Osthoff, H., and Hardesty, R. M.: Relationships of coastal nocturnal boundary
32   layer winds and turbulence to Houston ozone concentrations during TexAQS 2006, J. Geophys.
33   Res., 115, D10304, doi:10.1029/2009JD013169, 2010.
34
35   TCEQ: Texas Commission on Environmental Quality (2022). Houston-Galveston-Brazoria:
36   Current attainment status, available at: https://www.tceq.texas.gov/airquality/sip/hgb/hgb-status
37   (last access: 12 Jan 2023)
38
39   Travis, K. R. and Jacob, D. J.: Systematic bias in evaluating chemical transport models with
40   maximum daily 8 h average (MDA8) surface ozone for air quality applications: a case study with
41   GEOS-Chem v9.02, Geosci. Model Dev., 12, 3641–3648, https://doi.org/10.5194/gmd-12-3641-
42   2019, 2019.
43
44   Travis, K. R., Jacob, D. J., Fisher, J. A., Kim, P. S., Marais, E. A., Zhu, L., Yu, K., Miller, C. C.,
45   Yantosca, R. M., Sulprizio, M. P., Thompson, A. M., Wennberg, P. O., Crounse, J. D., St. Clair,
46   J. M., Cohen, R. C., Laughner, J. L., Dibb, J. E., Hall, S. R., Ullmann, K., Wolfe, G. M., Pollack,

I. B., Peischl, J., Neuman, J. A., and Zhou, X.: Why do models overestimate surface ozone in the Southeast United States?, Atmos. Chem. Phys., 16, 13561–13577, https://doi.org/10.5194/acp-16-13561-2016, 2016.

Vivone, G., D'Amico, G., Summa, D., Lolli, S., Amodeo, A., Bortoli, D., and Pappalardo, G.: Atmospheric boundary layer height estimation from aerosol lidar: A new approach based on morphological image processing techniques, Atmos. Chem. Phys., 21, 4249–4265, https://doi.org/10.5194/ACP-21-4249-2021, 2021.

Wang, H., Li, Z., Lv, Y., Zhang, Y., Xu, H., Guo, J., and Goloub, P.: Determination and climatology of the diurnal cycle of the atmospheric mixing layer height over Beijing 2013–2018: lidar measurements and implications for air pollution, Atmos. Chem. Phys., 20, 8839–8854, https://doi.org/10.5194/acp-20- 8839-2020, 2020.

Xie, B., Hunt, J.C.R., Carruthers, D.J., Fung, J.C.H. and Barlow, J.F.: Structure of the planetary boundary layer over Southeast England: Modeling and measurements, J. Geophys. Res.-Atmos., 118(14), 7799-7818, 2013.

Xu, Z., Huang, X., Nie, W., Shen, Y., Zheng, L., Xie, Y., Wang, T., Ding, K., Liu, L., Zhou, D., Qi, X., and Ding, A.: Impact of Biomass Burning and Vertical Mixing of Residual-Layer Aged Plumes on Ozone in the Yangtze River Delta, China: A Tethered-Balloon Measurement and Modeling Study of a Multiday Ozone Episode, J. Geophys. Res.-Atmos., 123, 11786–711803, https://doi.org/10.1029/2018jd028994, 2018.

Yahya, K., Wang, K., Gudoshava, M., Glotfelty, T., and Zhang, Y.: Application of WRF/Chem over North America under the AQMEII Phase 2: Part I. Comprehensive evaluation of 2006 simulation, Atmos. Environ., 115, 733–755, https://doi.org/10.1016/j.atmosenv.2014.08.063, 2015.

Yi, C., Davis, K. J., Berger, B. W., and Bakwin P. S.: Long-term observations of the dynamics of the continental planetary boundary layer, J. Atmos. Sci., 58, 1288–1299, 2001.

Zhang, C., Wang, Y., and Hamilton, K.: Improved representation of boundary layer clouds over the southeast Pacific in ARW-WRF using a modified Tiedtke cumulus parameterization scheme, Mon. Weather Rev., 139, 3489–3513, https://doi.org/10.1175/MWR-D-10-05091.1, 2011.

Zhang, Y., Li, D., Lin, Z., Santanello Jr., J. A., and Gao, Z.: Development and evaluation of a long-term data record of planetary boundary layer profiles from aircraft meteorological reports, J. Geophys. Res.-Atmos., 124, 2008–2030, 2019.

Zhang, Y., Sun, K., Gao, Z., Pan, Z., Shook, M. A., and Li, D.: Diurnal climatology of planetary boundary layer height over the contiguous United States derived from AMDAR and reanalysis data, J. Geophys. Res.-Atmos., 125, e2020JD032803, https://doi.org/10.1029/2020JD032803, 2020.

1  Zhao, D., Xin, J., Wang, W., Jia, D., Wang, Z., Xiao, H., Liu, C., Zhou, J., Tong, L., Ma, Y. and
2  Wen, T.: Effects of the sea-land breeze on coastal ozone pollution in the Yangtze River Delta,
3  China, Science of The Total Environment, 807, 150306, 2022.
4
5  Zhou, W., Cohan, D. S., and Henderson, B. H.: Slower ozone production in Houston, Texas
6  following emission reductions: evidence from Texas Air Quality Studies in 2000 and 2006,
7  Atmos. Chem. Phys., 14, 2777–2788, doi:10.5194/acp-14-2777- 2014, 2014.
8