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

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

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

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

Abstract. The Tracking Aerosol Convection Experiment Air Quality (TRACER-AQ) campaign probed Houston air quality with a comprehensive suite of ground-based and airborne remote sensing measurements during the intensive operating period in September 2021. Two post-frontal high-ozone episodes (September 6-11 and 23-26) were recorded during the said period. In this study, we evaluated the simulation of the planetary boundary layer (PBL) height and the vertical ozone profile by a high-resolution (1.33 km) 3-D photochemical model, Weather Research and Forecasting (WRF)-driven GEOS-Chem (WRF-GC). We evaluated the PBL heights with a ceilometer at the coastal site La Porte and the airborne High Spectral Resolution Lidar-2 (HSRL-2) flying over urban Houston and adjacent waters. Compared with the ceilometer at La Porte, the model captures the diurnal variations of the PBL heights with a very strong temporal correlation ($R > 0.7$) and ±20% biases. Compared with the airborne HSRL-2, the model exhibits moderate to strong spatial correlation (R=0.26-0.68) with ±20% biases during noon and afternoon hours during ozone episodes. For land-water differences in PBL heights, the water has shallower PBL heights compared to land. The model predicts larger land-water differences than observations because the model consistently underestimates PBL heights over land compared to water. We evaluated vertical ozone distributions by comparing the model against vertical measurements from the Tropospheric Ozone lidar (TROPOZ), the HSRL-2, and ozonesondes, as well as surface measurements at La Porte from a model 49i ozone analyzer and one Continuous Ambient Monitoring Stations (CAMS). The model underestimates free tropospheric ozone (2-3 km aloft) by 9%-22% but overestimates near-ground ozone (< 50 m aloft) by 6%-39% during the two ozone episodes. Boundary layer ozone (0.5-1 km aloft) is underestimated by 1%-11% during September 8-11 but overestimated by 0%-7% during September 23-26. Based on these evaluations, we identified two model limitations: the single-layer PBL representation and free tropospheric ozone underestimation. These limitations have implications for the predictivity of ozone's vertical mixing and distribution in other models.

Deleted: s

Deleted: ~~We contrasted the model performance between ozone-episode days and non-episode days.~~

Deleted: T

Deleted: during ozone episodes (R = 0.72-0.77; normal mean bias (NMB) = 3%-22%) and non-episode days (R = 0.88; NMB = -21%)

Deleted: , compared with the ceilometer at La Porte

Deleted: Land-water differences in PBL heights are captured better during non-episode days than episode days, compared with the airborne High Spectral Resolution Lidar-2 (HSRL-2). During ozone episodes, the simulated land-water differences are 50-60 m (morning), 320-520 m (noon), and 440-560 m (afternoon) in comparison with the observed values of 190 m, 130 m, and 260 m, respectively. During non-episode days, the simulated land-water differences are 140-220 m (morning) and 360-760 m (noon) in comparison with the observed values of 210 m and 420 m, respectively. For …

Deleted: ,

Deleted: was evaluated

Deleted: profile

Deleted: at the

Deleted: a site from the

Deleted: at La Porte

Formatted: Font: Not Bold

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, but various types of measurements can be used to calculate the height of the PBL. These measurements include (1) thermodynamical quantities, (2) atmospheric aerosol particle characteristics, (3) atmospheric gases, and (4) wind and turbulence (Kotthaus et al., 2023). The first type measures thermodynamic properties (e.g., temperature, water vapor mixing ratio, etc.) with microwave radiometer (MWR), infrared spectrometer (IRS), Raman lidar, radio acoustic sounding system (RASS), etc (Cimini et al., 2013; Wulfmeyer et al., 2010). The second type measures backscatter signals (e.g., attenuated backscatter, particle backscatter, etc.) with aerosol lidars, ceilometers, etc (Caicedo et al., 2017; Knepp et al., 2017; Li et al., 2022). The third type

Deleted: e.g.

Deleted: e.g.

Deleted: e.g.

Deleted: e.g.

Deleted: e.g.

Deleted: e.g.

Deleted: e.g.

Deleted: . 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.

Formatted: Font: (Default) Times New Roman, (Asian) +Body Asian (DengXian), Font color: Text 1, Pattern: Clear

Formatted: Font: (Default) Times New Roman, (Asian) +Body Asian (DengXian), Font color: Text 1, Pattern: Clear

Formatted: Font: (Default) Times New Roman, (Asian) +Body Asian (DengXian), Font color: Text 1, Pattern: Clear

Deleted: Two major types of data have been commonly used to validate the modeled PBL height

Formatted: Font: (Default) Times New Roman, (Asian) +Body Asian (DengXian), 12 pt, Font color: Text 1

Formatted: Font: (Default) Times New Roman, (Asian) +Body Asian (DengXian), 12 pt, Font color: Text 1

Formatted: Font: (Default) Times New Roman, (Asian) +Body Asian (DengXian), 12 pt, Font color: Text 1

Formatted: Font: (Default) Times New Roman, (Asian) +Body Asian (DengXian), 12 pt, Font color: Text 1

Formatted: Font: (Default) Times New Roman, (Asian) +Body Asian (DengXian), 12 pt, Font color: Text 1

Deleted: is the PBL height derived from the profiles of

Deleted: measured

Deleted: by

Formatted: Font: (Default) Times New Roman, (Asian) +Body Asian (DengXian), 12 pt, Font color: Text 1

Formatted: Font: (Default) Times New Roman, (Asian) +Body Asian (DengXian), 12 pt, Font color: Text 1

Formatted: Font: (Default) Times New Roman, (Asian) +Body Asian (DengXian), 12 pt, Font color: Text 1

Deleted: ozonesondes and radiosondes (Zhang et al., 2019; Zhang et al., 2020; Morris et al., 2010; Rappenglück et al., 2008)…

Deleted: 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).

Formatted: Font: (Default) Times New Roman, (Asian) +Body Asian (DengXian), 12 pt, Font color: Text 1

Deleted: of data is remotely-sensed mixed layer height as defined by aerosol

Deleted: gradients

measures the mass or number concentration of gases with differential absorption lidar (DIAL) (Hair et al., 2008). The fourth type measures dynamic and turbulent processes (e.g., horizontal wind speed, variances of the velocity components, turbulent kinetic energy (TKE), eddy dissipation rate, etc.) with Doppler wind lidar (DWL), radar wind profiler (RWP), sodar, etc (Bonin et al., 2016; Bodini et al., 2018; Bonin et al., 2018; Angevine et al., 2003).

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

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

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

Formatted: Font: (Default) Times New Roman, (Asian) +Body Asian (DengXian), 12 pt, Font color: Text 1

Formatted: Font: (Default) Times New Roman, (Asian) +Body Asian (DengXian), 12 pt, Font color: Text 1

Formatted: Font: (Default) Times New Roman, (Asian) +Body Asian (DengXian), 12 pt, Font color: Text 1

Formatted: Font: (Asian) +Body Asian (DengXian), Font color: Text 1

Deleted: , which is becoming more widely available with ceilometer data and aircraft lidars and can be adopted for model evaluation (Caicedo et al., 2017, 2020; Knepp et al., 2017; Li et al., 2021; Wang et al., 2020). ¶

Formatted: Normal (Web)

Deleted: ¶

2. Model and Data

2.1 Observations

In this study, to evaluate PBL heights and vertical ozone distribution, continuous and high-resolution measurements (i.e., 1–10 minutes) were obtained from two observational sources including (1) the ground-based instruments at the La Porte site and (2) the airborne instrument flying over urban Houston and the Galveston Bay in September of 2021 (Fig. 1). Compared with discrete or low-resolution measurements (*e.g.*, hourly) of PBL heights used in previous studies in Houston (Haman et al., 2014; Cuchiara et al., 2014; Rappenglück et al., 2008), the high-resolution measurements in TRACER-AQ field campaign are capable to probe into the fine PBL structure and its development as well as the associated vertical ozone profiles.

The first observational source includes multiple measurements at the La Porte site, that is (1) continuous vertical ozone profiles from the NASA Goddard Space Flight Center (GSFC) Tropospheric Ozone (TROPOZ) Differential Absorption Lidar (DIAL) (Sullivan et al., 2014), (2) continuous aerosol mixed layer height derived from atmospheric backscatter profiling with a CHM 15k-x ceilometer, (3) multiple ozonesonde launches, and (4) continuous surface ozone and meteorology measurements. The following paragraphs provide detailed introductions for the four types of measurements mentioned above.

First, the TROPOZ, as part of the ground-based Tropospheric Ozone Lidar Network (TOLNet, https://www-air.larc.nasa.gov/missions/TOLNet/), has been used to provide continuous, high-resolution profile measurements of vertical ozone profile during various campaigns for satellite and model evaluation (Sullivan et al., 2014, 2015, 2019, 2022; Bernier et al., 2022; Kotsakis et al., 2022; Dacic et al., 2020; Johnson et al., 2016; Dreessen et al., 2016). The TROPOZ data can be used to identify pollutant transport to understand the vertical mixing of ozone.

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

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

Last, surface measurements at La Porte included ozone, air temperature, relative humidity, and wind speed and direction. This study uses surface ozone from a Thermo Scientific model 49i ozone analyzer operated by the GSFC and a TCEQ Continuous Ambient Monitoring Stations

Deleted: T
Deleted: the
Deleted: the
Deleted: this study adopted
Deleted: ,
Deleted: profiles
Deleted: measurements
Deleted: measurements
Deleted: covering
Deleted: PBL
Deleted: (*e.g.* hourly)
Deleted: The La Porte site was equipped with
Deleted: semi-
Formatted: Font: (Default) Times New Roman, (Asian) +Body Asian (DengXian), Font color: Text 1, Pattern: Clear
Deleted: ¶
Deleted: T
Deleted: Similar to the TROPOZ at the La Porte site, the University of Houston site measured semi-continuous vertical ozone profiles with the Langley Mobile Ozone Lidar (LMOL) (Gronoff et al., 2019, 2021).
Deleted: T
Deleted: is
Deleted: O
Deleted: on
Deleted: S

(CAMS) site named La Porte Sylvan Beach, as well as surface meteorology from a Lufft WS-501B operated by the GSFC.

The second observational source is the airborne High Spectral Resolution Lidar-2 (HSRL-2) datasets collected over the Houston area up to three times per day, roughly at 8:00-10:00, 11:00-13:00, and 14:00-16:00 local time, covering an area of approximately 50 km x 135 km. With its high resolution and vertically resolved measurements, the HSRL-2 demonstrated reliable performances on many previous airborne campaigns (Hair et al., 2018; Hair et al., 2008; Burton et al., 2015). The HSRL-2 provides below aircraft retrievals of the spatial and vertical distributions of ozone, aerosols, and mixed layer heights on ten flight days. This paper only reports on (1) mixed layer height derived from gradients in the aerosol backscatter profiles measured at 532 nm and (2) ozone mixing ratio along one flight track that has the nearest distance to the La Porte site (Fig. 1). It is worth noting that mixed layer heights from the HSRL-2 and the ceilometer at La Porte are derived differently. The HSRL-2 measures unattenuated aerosol backscatter profile while the ceilometer at the La Porte site measures attenuated total backscatter profiles of the atmosphere (incl. aerosols and molecules).

2.2 Identification of ozone episodes

Ozone exceedance days used in this study were identified by the same criteria used in Li et al. (2023) and Soleimanian et al. (2023), where (1) any onshore site from the TCEQ CAMS network in Houston and Galveston or (2) offshore ozone observations by boat operating in Galveston Bay during the field campaign registered daily maximum 8-hour average (MDA8) ozone in exceedance of the current NAAQS air quality standard of 70 ppbv (see Text S1 for details; refer to Li et al. (2023) for full description of the boat observations). Three high ozone episodes in September of 2021 were identified based on the above criteria: September 6-11, September 17-19, and September 23-26, consisting of 13 ozone exceedance days. We excluded analysis from the September 17-19 episode because it happened right after tropical cyclone Nicholas, which made landfall 125 km south-southwest of Houston and hindered measurements at the ground sites and aircraft due to clouds and power outages. The model meteorology was not designed to capture the cyclone either. Other September days were used as a control to represent non-episode days.

2.3 Model

2.3.1 Model description

WRF-GC v2.0 is a regional air quality model (Feng et al., 2021; Lin et al., 2020) that couples the Weather Research and Forecasting (WRF) meteorological model (v3.9.1.1) with the GEOS-Chem atmospheric chemistry model (v12.7.2). The WRF and GEOS-Chem versions are benchmarks of WRF-GC v2.0 with the proven performance of meteorology, PBL heights, and aerosol simulation in Feng et al. (2021) and Lin et al. (2020). We evaluated WRF-GC's prediction of ozone during the TRACER-AQ study. We set up three domains with different horizontal resolutions that cover the contiguous United States, Southeast Texas, and the Houston-Galveston region, referred to as d01, d02, and d03, respectively, as shown in Figure 1. The corresponding horizontal resolutions for d01–d03 are 12 km, 4 km, and 1.33 km, respectively. All domains have identical vertical resolutions with 50 hybrid sigma-eta vertical levels spanning from the surface to 10 hPa. Vertical resolution ranges from ~70 m (near the

Deleted: ¶
In September 2021, the NASA Gulfstream-V aircraft flew on ten flight days. This analysis uses

Formatted: Normal (Web)

Deleted:).The

Deleted: , aerosol optical properties

Deleted: ¶

Deleted: M

Deleted: the ceilometer at La Porte and

Moved down [2]: The ceilometer at the La Porte site measures attenuated total backscatter profiles of the atmosphere (incl. aerosols and molecules), while

Deleted: t

Deleted: can

Deleted: the

Moved (insertion) [2]

Deleted: T

Deleted: , while

Deleted: Both ceilometer and the HSRL-2 signals can be used to derive mixed layer height. This study uses mixed layer heights from the ceilometer and the HSRL-2 to evaluate the WRF-GC prediction of PBL heights.

Deleted: ¶

Moved down [1]: Apart from the observations above, we used geopotential heights and winds at 850 hPa from the European Centre for Medium Range Weather Forecast (ECMWF) reanalysis version5 (ERA5) dataset (description in Sect. 2.3.2) to derive the synoptic conditions in Fig. 2. ¶

Deleted: The supplement includes (1) surface measurements from the TCEQ CAMS and the boats throughout July to October used to identify ozone exceedance days, (2) details on the assimilation and evaluation of the modeled meteorology with these measurements, and (3) vertical ozone distribution at University of Houston by LMOL. ¶

Deleted: ¶

Moved (insertion) [3]

Deleted: ozone observations

Deleted: of 70 ppbv,

Deleted: U.S. Environmental Protection Agency (EPA)

Deleted: for ozone

Deleted: ground-sites

Deleted: ¶

Moved up [3]: any onshore site from the CAMS network in Houston and Galveston or offshore boat ozone observations registered daily maximum 8-hour average (MDA8) ozone in

Deleted: according to surface measurements from the T ... [1]

Formatted: Left

ground) to ~700 m (aloft); the first 2 km above the ground has 10 model layers, and the first 4 km has 14 model layers.

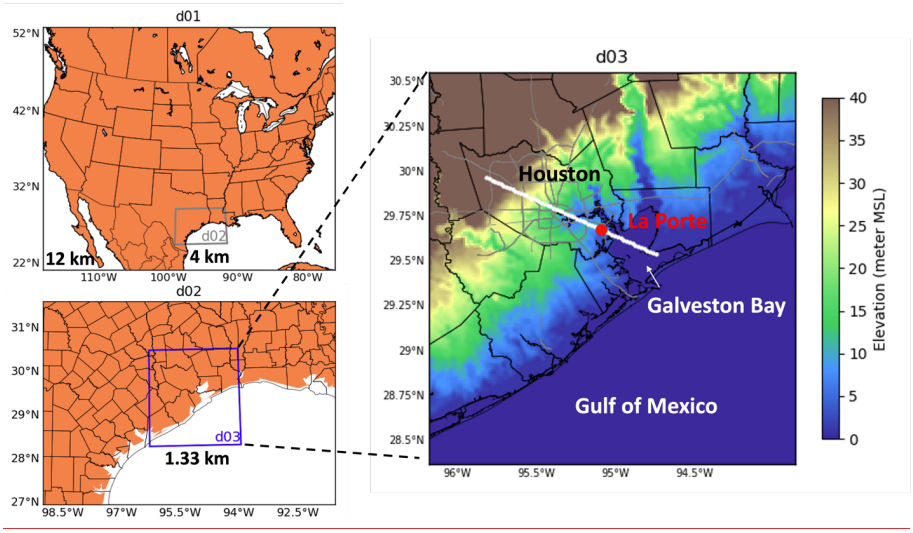


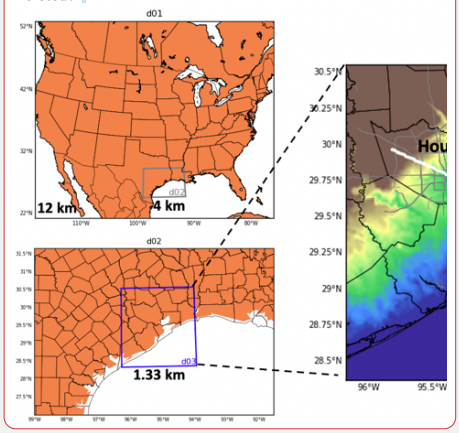
Figure 1. WRF-GC nested domains and their horizontal resolutions. The La Porte site is labeled as a red dot. The white line represents a flight track that is chosen because of its nearest distance to the La Porte site.

WRF-GC uses the most updated full O_x-NO_x-VOC-halogen-aerosol chemistry from GEOS-Chem. The anthropogenic emissions used are the 2019 TCEQ emission inventory for Houston and southeastern Texas, the 2013 National Emission Inventory for the rest of the US, and the 2014 Community Emissions Data System (CEDS) for regions outside of the US. Biomass burning emissions are from the 2019 Global Fire Emissions Database (GFED). Biogenic emissions are from the Model of Emissions of Gases and Aerosols from Nature (MEGAN) (Guenther et al., 2012). Soil NO_x (Hudman et al., 2012) and lightning NO_x (Murray et al., 2012) emissions are also included.

2.3.2 Model configurations

Boundary and initial conditions for WRF employed three alternative meteorological analyses. They were (1) the National Centers for Environmental Prediction (NCEP)-Final Analysis (FNL) (https://rda.ucar.edu/datasets/ds083.3/), (2) the fifth generation of European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric reanalysis (ERA5) data (https://rda.ucar.edu/datasets/ds633.0/), and (3) the High-Resolution Rapid Refresh (HRRR) from NOAA Amazon Web Service (https://registry.opendata.aws/noaa-hrrr-pds). The temporal resolution for FNL, ERA5, and HRRR is 6-hourly, hourly, and hourly, respectively. The horizontal resolution for FNL, ERA5, and HRRR is 0.25°, 0.25°, and 3 km, respectively. We used geopotential heights and winds at 850 hPa from the ERA5 dataset to derive synoptic conditions in Fig. 2.

Deleted: ¶



Moved (insertion) [1]

Deleted: Apart from the observations above, we used geopotential heights and winds at 850 hPa from the European Centre for Medium Range Weather Forecast (ECMWF) reanalysis version5 (ERA5) dataset (description in Sect. 2.3.2) …

Deleted: the

Deleted: ¶

WRF has different schemes or options to represent physics and dynamics processes. Three PBL schemes were used to investigate the effect of different parameterizations of heat, moisture, and momentum exchange between the surface and PBL on the simulated PBL structure and height. They are the local closure Mellor-Yamada-Nakanishi-Niino (MYNN) scheme (Nakanishi and Niino, 2009), the non-local closure Yonsei University (YSU) scheme (Hong et al., 2006), and the hybrid local-nonlocal Asymmetric Convective Model version 2 (ACM2) scheme (Pleim, 2007). Details of the PBL schemes are in Sect. 2.3.3. Two microphysics schemes were used: the Morrison double moment (2M) scheme (Morrison et al., 2009) and the single-moment 6-class (WSM6) scheme (Hong and Lim, 2006). Other schemes adopted in this paper were the Monin-Obukhov Similarity surface layer, the Noah land surface scheme (Chen and Dudhia, 2001), the Rapid Radiative Transfer Model (RRTM) longwave and shortwave radiation schemes (Iacono et al., 2008), and the New Tiedtke cumulus scheme (Zhang et al., 2011; Tiedtke, 1989).

To select the best model configuration to represent meteorology during the 2021 TRACER-AQ campaign, we designed eight model experiments with different physics options, boundary meteorology, data assimilation, and reinitializing option, as listed in Table S2. First, [Base] is the baseline configuration: MYNN for PBL, 2M for microphysics, NCEP FNL for boundary conditions, no nudging for assimilation, and no reinitialization. Second, [YSU] and [ACM2] experiments used the YSU and ACM2 PBL schemes, respectively, while keeping other options the same as [Base]. Differences between [Base], [YSU], and [ACM2] show the effects of different PBL parameterizations. Third, the [WSM6] experiment differs from [Base] by replacing the 2M microphysics scheme with WSM6. Differences between [Base] and [WSM6] show the effects of different microphysics schemes. Next, [ERA5] and [HRRR] were designed to show the effects of different meteorological initial and boundary conditions on the WRF performance by using ERA5 and HRRR instead of NCEP FNL, respectively. We examined the effects of data assimilation options in [Nudged]. [Nudged] adopted observation nudging and surface analysis nudging to assimilate both onshore and offshore measurements from multiple platforms, including the TCEQ CAMS, boats, and the NCEP surface and upper air measurements into WRF meteorology (see Text S2 for details). Differences between [Base] and [Nudged] show the effects of assimilation. Last, [Reinit] used daily reinitialization where the simulation was broken into many 30-hour segments with the first 6 hours of each segment (18:00-23:00 Coordinated Universal Time (UTC) of a previous day) as spin-up and the subsequent 24 hours (0:00-23:00 UTC of the following day) used for analysis (Yahya et al., 2015; Otte et al., 2008). Differences between [Base] and [Reinit] show the effects of a free-running option versus model reinitialization.

The WRF model generally reproduces observed temporal variability and spatial distribution in key meteorological parameters with a correlation coefficient higher than 0.5 in most cases. However, the model, regardless of configuration settings, shows persistent low biases in PBL heights, low biases in air temperatures, high biases in relative humidity, and high biases in wind speed (see Text S3 for details). While different WRF configuration has its own advantage in reducing model biases, [HRRR], [Nudged], and [Reinit] configurations stand out as the three best simulations based on campaign-wide statistics (see Text S3 for details). Considering that [Nudged] requires additional efforts to prepare observational datasets and [Reinit] needs to automate the model running process, [HRRR] is the easiest and the most effective option to

Deleted: ¶

reproduce meteorology for computationally expensive chemistry simulations and was thus selected to be presented in the analysis below. Meanwhile, the three simulations with different PBL schemes (i.e., [Base], [YSU], and [ACM2]) were also selected because the choice of the PBL scheme is crucial in determining PBL heights (Section 2.3.3), which is one of the major interests of this study. Therefore, we chose four simulations, that is [HRRR], [Base], [YSU], and [ACM2], in Table 1. The surface layer, land surface, longwave and shortwave radiation, and Tiedtke cumulus schemes remain unchanged in all simulations.

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

2.3.3 Determination of PBL height in different schemes

Atmospheric models adopt the thermodynamic concept and rely on parameterization schemes to define the structure and the height of the PBL. The heights of the PBL are determined differently among different PBL schemes in the WRF model. The intra-scheme differences can originate from (1) the vertical profile of thermodynamic quantities simulated with different assumptions of the vertical exchange of heat, moisture, and momentum and (2) the diagnosis of the PBL height from these thermodynamic quantities. The PBL heights determined by different schemes can differ by 20-30% (Hu et al., 2010; Xie et al., 2013).

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

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

Deleted: during the TRACER-AQ 2021 campaign

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Default Paragraph Font, Font: (Asian) Times New Roman, Font color: Auto

Formatted: Font: (Default) Times New Roman, Font color: Text 1, Pattern: Clear

Formatted: Font: (Default) Times New Roman, Font color: Text 1, Pattern: Clear

Formatted: Font: (Default) Times New Roman, Font color: Text 1, Pattern: Clear

Deleted: ¶
Among the above simulations, we chose four simulations (Table 1), including the three simulations with different PBL schemes …

Deleted: and the best simulation [HRRR] determined by campaign-wide statistics (see Text S3 for details) in the analysis below.

Deleted: ¶
¶
¶

Deleted: N

Deleted: Y

PBL height is below ~0.5 km, while it weights the TKE-based definition negligible when the BRN-based PBL height is above ~1 km.

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

2.4 Performance metrics for wind

The mean of wind speed and direction is calculated using the vector notation approach, a commonly used method in wind evaluations, as described in Yu et al. (2023). This method treats wind as vectors with their u (eastward) and v (northward) wind components. First, the mean u and v wind components are found by averaging all u and v wind values over a given time period. Then, the resultant vector is determined by taking the square root of the sum of the squares of the mean u and mean v wind components. The magnitude of the resultant vector represents the mean wind speed, and the angle of the resultant vector represents the mean wind direction.

The difference between observed and modeled wind direction was calculated as below.

$$\Delta = \begin{cases} M - O, & \text{when } |M - O| \le 180^\circ \\ (M - O)\left(1 - \dfrac{360}{|M - O|}\right), & \text{when } |M - O| > 180^\circ \end{cases}$$

where M is the model output, and O is the observation. The correlation between observed and modeled wind direction was determined by a circular correlation coefficient as below.

$$R = \frac{\sum_{i=1}^{N} \sin\,(M_i - \bar{M}) \sin\,(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{N} \sin^2(M_i - \bar{M})} \sqrt{\sum_{i=1}^{N} \sin^2(O_i - \bar{O})}}$$

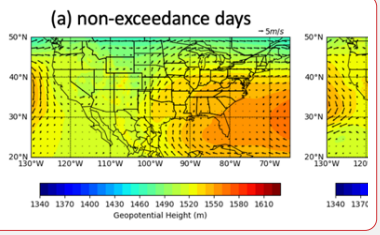
3. Lower tropospheric layering for ozone-episode and non-ozone-episode days

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 temperatures (Fig. 3a) and relative humidity (Fig. 3b) than non-episode days.

Formatted: Highlight
Formatted: Highlight
Formatted: Font: (Default) Times New Roman, Font color: Text 1, Pattern: Clear
Formatted: Font: (Default) Times New Roman, Font color: Text 1, Pattern: Clear
Formatted: Font: (Default) Times New Roman, Font color: Text 1, Pattern: Clear
Formatted: Font: (Default) Times New Roman, Font color: Text 1, Pattern: Clear
Formatted: Font: (Default) Times New Roman, Font color: Text 1, Pattern: Clear
Formatted: Font: (Default) Times New Roman, Font color: Text 1, Pattern: Clear
Formatted: Font: (Default) Times New Roman, Font color: Text 1, Pattern: Clear
Formatted: Font: Bold, Highlight
Formatted: Indent: First line: 0 cm

Deleted: temperature
Deleted: b
Deleted: d
Deleted: (Fig. 3a; Fig. 3c)

Deleted:



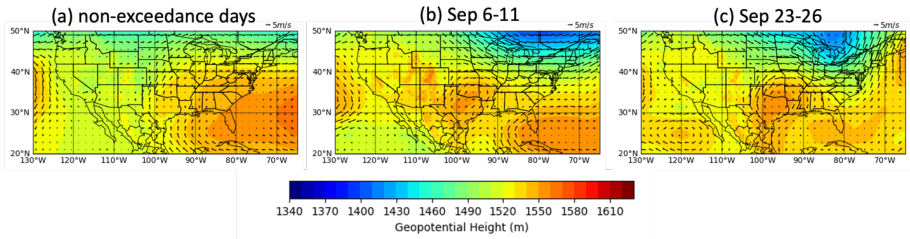


Figure 2. Synoptic conditions denoted by geopotential height at 850 hPa and the associated winds for (a) the non-episode days and the two ozone episodes of (b) September 6-11 and (c) September 23-26, 2021.

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 and ozone-episode 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 (Fig. 3e). Such differences can also be seen from the direct measurements of the ceilometer, e.g. atmospheric backscatter profiles. During ozone episodes, the high-pressure center creates favorable meteorological conditions 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. 3f). 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 as observed CBL, SBL, or RL hereafter in a manner consistent with the modeled equivalents.

4. PBL height evaluation

Based on the observed differences in diurnal PBL variations between non-episode days and ozone episodes in Section 3, we first assessed the observation-model differences in diurnal variation of PBL heights and other meteorological variables in Section 4.1. The ground-based ceilometer at the La Porte site is used to evaluate the diurnal variation due to its capability for continuous measurements throughout the day. Meanwhile, the HSRL-2 instrument provides data covering a significant portion of the urban Houston region and adjacent waters and is thus used to evaluate spatial and temporal (daytime) variations of the PBL heights in Section 4.2.

Deleted: (Fig. 3i)

Deleted: (Fig. 3j)

Deleted: i

Deleted: e

Deleted: , in Figures 3k and 3l

Deleted: condition

Deleted: l

Deleted: k

Deleted: The next section evaluates the observed and the modeled PBL heights.

Formatted: Tab stops: 7.06 cm, Left

Deleted: In this section, we evaluate the modeled PBL height with two types of independent field measurements. The ground-based ceilometer at the La Porte site is used to evaluate the diurnal variation, given its continuous measurements throughout the day.

Deleted: acquired

Deleted: d

Deleted: over much of

Formatted: Font: (Default) Times New Roman, Font color: Text 1

4.1 Evaluation with ceilometer

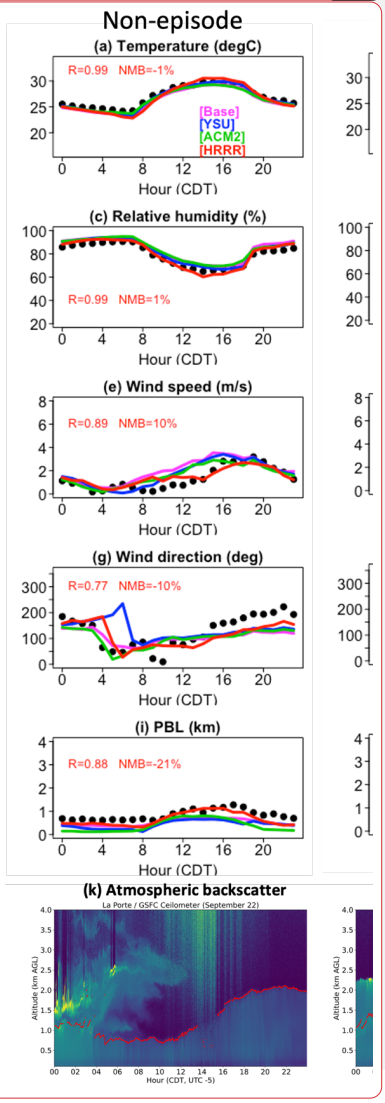
Figure 3 compares the diurnal variations of the PBL height between non-episode days and ozone-episode days. During the daytime, the observations represent the CBL height for both types of days. This aligns with the standard model output for the PBL height during the daytime, which is the CBL height. At night, the observations represent the SBL height on non-episode days whereas they represent the RL height on ozone-episode days. However, the model only provides the SBL as the standard output for nighttime PBL, lacking information on other nocturnal layers such as RL. As a result, meaningful comparisons between the observed RL height and the modeled SBL height during ozone episodes become challenging. Therefore, the modeled RL needs to be extracted for a fair comparison against the observed RL during ozone episodes.

Deleted: D…urnal variations of the PBL heights…averaged …etween non-episode days and ozone-episode daysduring non-episode periods and ozone episodes are separately evaluated in Figure 3… During the daytime, tT…e observations represent the daytime …BL height on …or both types of days. This aligns with the standard model output for the PBL height during the daytime, which is the CBL height. At night, the observations represent the SBL height on non-episode days whereas they represent but ... [2]

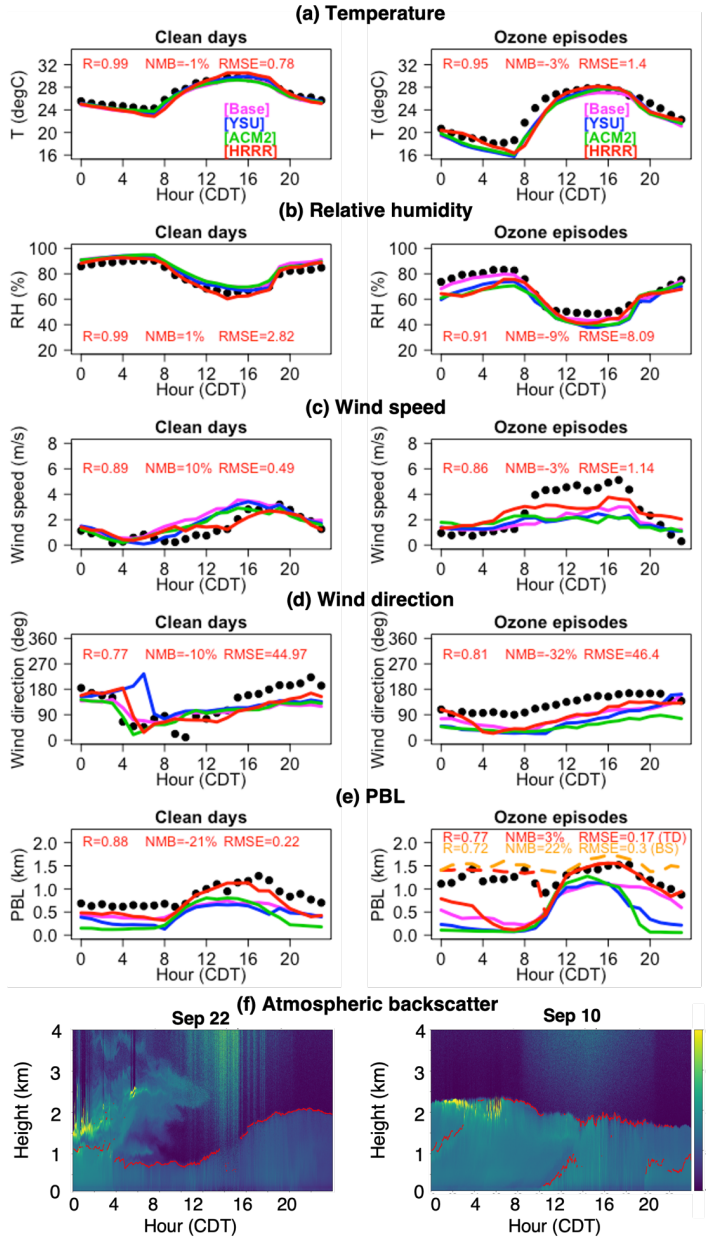
Formatted: Font: (Default) Times New Roman, Font color: Text 1, Pattern: Clear

Formatted: Font: (Default) Times New Roman, Font color: Text 1, Pattern: Clear

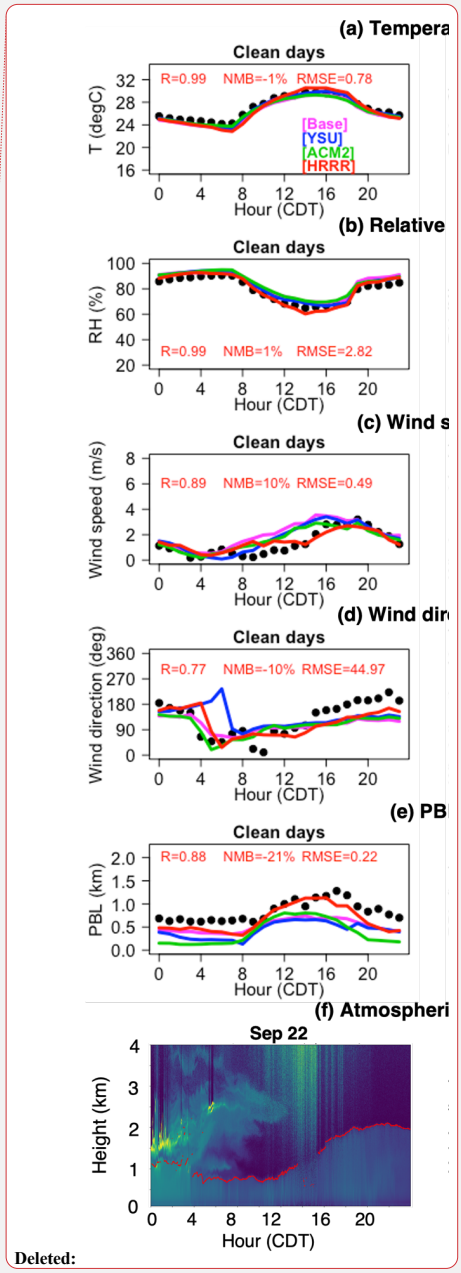
Deleted: The modeled equivalents are needed to yield equal comparisons between the models and observations. The model diagnoses the CBL height as a standard output for the PBL height during the daytime. However, the model only diagnoses the SBL as the standard output for nighttime PBL rather than other nocturnal layers such as RL (Fig. 3j). …herefore, the modeled RL needs to be extracted for a valid …air comparison with ... [3]



Deleted:


(a) Temperature
Clean days
R=0.99 NMB=-1% RMSE=0.78
Ozone episodes
R=0.95 NMB=-3% RMSE=1.4
[Base]
[YSU]
[ACM2]
[HRRR]
T (degC)
Hour (CDT)
(b) Relative humidity
R=0.99 NMB=1% RMSE=2.82
R=0.91 NMB=-9% RMSE=8.09
RH (%)
(c) Wind speed
R=0.89 NMB=10% RMSE=0.49
R=0.86 NMB=-3% RMSE=1.14
Wind speed (m/s)
(d) Wind direction
R=0.77 NMB=-10% RMSE=44.97
R=0.81 NMB=-32% RMSE=46.4
Wind direction (deg)
(e) PBL
R=0.88 NMB=-21% RMSE=0.22
R=0.77 NMB=3% RMSE=0.17 (TD)
R=0.72 NMB=22% RMSE=0.3 (BS)
PBL (km)
(f) Atmospheric backscatter
Sep 22
Sep 10
Height (km)
Normalized Range Corrected Signal (Log10 AU)

1


(a) Tempera
Clean days
R=0.99 NMB=-1% RMSE=0.78
[Base]
[YSU]
[ACM2]
[HRRR]
T (degC)
Hour (CDT)
(b) Relative
R=0.99 NMB=1% RMSE=2.82
RH (%)
(c) Wind s
R=0.89 NMB=10% RMSE=0.49
Wind speed (m/s)
(d) Wind dir
R=0.77 NMB=-10% RMSE=44.97
Wind direction (deg)
(e) PB
R=0.88 NMB=-21% RMSE=0.22
PBL (km)
(f) Atmospheri
Sep 22
Height (km)

Deleted:

Figure 3. Diurnal variations in observed versus modeled surface meteorology and PBL height averaged over non-episode periods (left) and ozone episodes (right) in September 2021. In (a)-(e), black dots are NASA GSFC observations at the La Porte site, while lines are model equivalents of different configurations of WRF-GC. In (e), dashed lines represent residual layers identified by aerosol backscatter (BS) versus thermodynamically by potential temperature (TD) from the [HRRR] configuration. (f) shows the ceilometer-measured atmospheric backscatter profiles overlaid with mixed layer heights of a non-episode day of September 22 and an ozone-episode day of September 10.

We first selected [HRRR] as the best simulation among the four simulations according to their daytime performances for both types of days. On non-episode days, the model simulations show the diurnal mean and standard deviation of the PBL height of 0.52±0.14 km for [Base], 0.43±0.17 km for [YSU], 0.39±0.27 km for [ACM2], and 0.66±0.28 km for [HRRR] in comparison with the observed value of 0.83±0.22 km (left panel of Fig. 3e). On ozone-episode days, the model simulations show the CBL height variation of 0.96±0.18 km for [Base], 0.60±0.37 km for [YSU], 0.50±0.5 km for [ACM2], and 1.25±0.29 km for [HRRR] in comparison with the observed value of 1.26±0.24 km during the afternoon and evening hours of 15:00-23:00 CDT (right panel of Fig. 3e). During the same period, the model simulations show the PBL decay rates of 53 m h^{-1} for [Base], 102 m h^{-1} for [YSU], 135 m h^{-1} for [ACM2], and 59 m h^{-1} for [HRRR] in comparison with the observed 60 m h^{-1}. These comparisons demonstrate that the four model simulations generally underestimate the PBL height by 180–450 m throughout the day on non-episode days and by 10–760 m during the daytime on ozone-episode days. Among the four simulations, [HRRR] best captures the observed mean height and decay rate during the daytime. Therefore, [HRRR] is selected to display its aerosol backscatter and potential temperature profiles in Figures 4 and 5, enabling further examination of its representation of the nighttime RL.

Next, the simulated aerosol backscatter and potential lapse rates of the [HRRR] simulation are used to extract the modeled RL heights. The modeled aerosol backscatter shows the volume of the atmosphere in which aerosol species are mixed and dispersed. Substantially stronger backscatter signals are found within the first ~2 km than the free troposphere at 3-4 km aloft (Fig. 4b, 5b). Therefore, we take the sharpest vertical gradient in the backscatter signal (i.e., the largest first derivative of backscatter) to estimate the modeled mixed layer height. The extracted layers have daytime variations of 1.58±0.13 km and nighttime variations of 1.50±0.06 km during ozone episodes. Such nighttime variations are representative of the RL top in the model. It is worth noting that the modeled aerosol backscatter in Figures 4b and 5b is not equivalent to the ceilometer-measured atmospheric backscatter in Figures 4a and 5a which includes both aerosol and molecular backscatter signals. Yet this modeled aerosol backscatter is the closest product from the model and denotes relatively consistent backscatter differences above and below the RL.

Apart from aerosol backscatter, the potential lapse rate is also commonly used to distinguish atmospheric layers according to their instability. The potential lapse rate or potential temperature gradient $\left(\frac{d\theta}{dz}\right)$ is thermodynamically defined as the changes of potential temperature (θ) with height (z). Distinguished by the values of the potential lapse rate, the modeled nocturnal PBL consists of a stable SBL, a neutrally stratified RL, and a CI layer during most ozone-episode days

Deleted: different days during

Deleted: For the first five rows

Deleted: data simulated by the

Deleted: from

Deleted: model… Different colors denote different model configurations. …n (e? ... [4]

Deleted: j…, dashed lines represent residual layers identified by aerosol backscatter (BS) and ... [5]

Deleted: (TD)

Deleted: () …rom the [HRRR] configuration. The last row…f) shows the ceilometer-measured atmospheric backscatter profiles overlaid with mixed layer heights of two typical days; … non-episode day of September 22 (k) …nd an ozone-episode day of September 10 (l) ... [6]

Deleted: Before extracting RL characteristics in the model, …w… first selected [HRRR] as the best one …imulation among the four simulations with the best…ccording to their daytime performances for both types of days among the four simulations and examined its nighttime performance ... [7]

Deleted: (Fig. 3i)

Deleted: [Base],…[YSU], [ACM2], and [HRRR] respectively ... [8]

Deleted: (Fig.3j)

Deleted: (…Base])… 0.60±0.37 km for (…YSU])… 0.50±0.5 km for (…ACM2])… and 1.25±0.29 km for (…HRRR]) ... [9]

Deleted: (…5:00-23:00 CDT) ... [10]

Deleted: (…Base])… 102 m h^{-1} for (…YSU])… 135 m h^{-1} for (…ACM2])… and 59 m h^{-1} for (…HRRR]) ... [11]

Moved (insertion) [4]

Deleted: All…model simulations generally underestimate the PBL height by: ... [12]

Moved up [4]: All model simulations generally underestimate the PBL height: 180–450 m throughout the day on non-episode days and 10–760 m during the daytime on ozone-episode days. The model underestimations are relative to the observed mixed layer height. The actual PBL height biases in the model can be larger or smaller than those underestimations depending on the relationship between the backscatter-defined mixed layer and the thermodynamically defined CBL. Among the four simulations, the [HRRR] best

Deleted: All model simulations generally underestima ... [13]

Deleted: Second…ext, the simulated aerosol backscat ... [14]

Moved (insertion) [5]

Deleted: T…ise…modeled aerosol backscatter is the ... [15]

Moved up [5]: The modeled aerosol backscatter is the closest

Deleted: P…tential lapse rate or potential temperature ... [16]

Deleted: Figures 4c and 5c show that

(Fig.4c, 5c). The modeled top of the RL can be identified from the height where the RL (with little or low temperature increases at 0-3 °C km^{-1}) transitions to the CI layer (with drastic temperature increases at 8-14 °C km^{-1}). Therefore, we use the sharpest gradient in the potential lapse rate, which is 6.6 °C km^{-1} on average, to identify the top of the modeled RL. This thermodynamically identified top of the RL exhibits a variation of 1.39±0.03 km during ozone episodes, which is slightly lower than the 1.50±0.06 km identified by the aerosol backscatter in the previous paragraph.

The modeled PBL heights identified by the two methods above show reasonable agreement with the observed PBL heights measured by the ceilometer. Such comparison can be found in Figure 3e. The thermodynamically-identified layer exhibits a slightly better agreement with the observation, exhibiting a correlation coefficient (R) of 0.77, a normalized mean bias (NMB) of 3%, and a root mean square error (RMSE) of 0.17 km. In contrast, the backscatter-identified layer shows slightly lower correlation and larger biases (R=0.72; NMB=22%; RMSE=0.30 km) during ozone episodes.

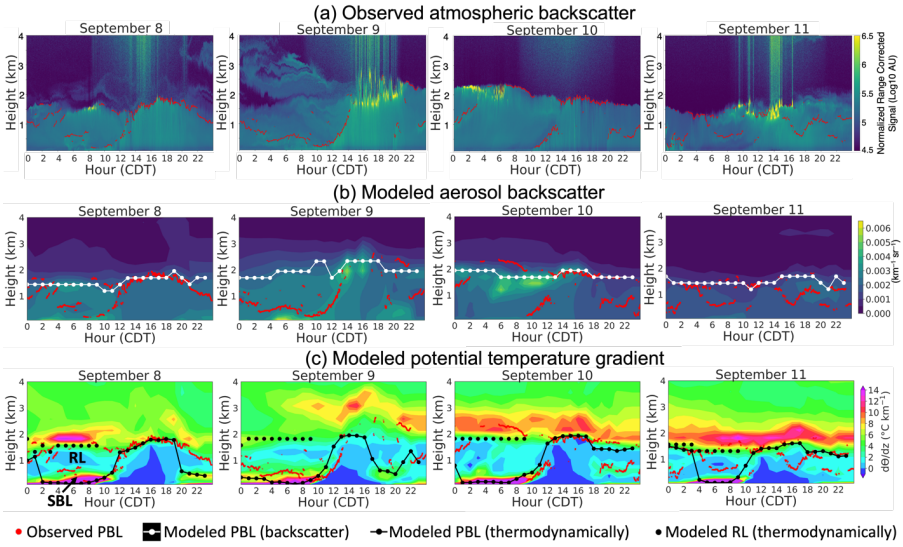
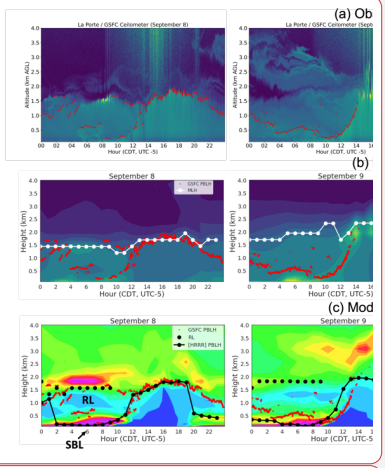


Figure 4. Observed and modeled heights of lower tropospheric layers at the La Porte site during the ozone episode of September 8-11. The contours show (a) ceilometer-observed attenuated atmospheric backscatter signal produced by aerosols and molecules combined at 1064 nm, (b) modeled unattenuated backscatter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

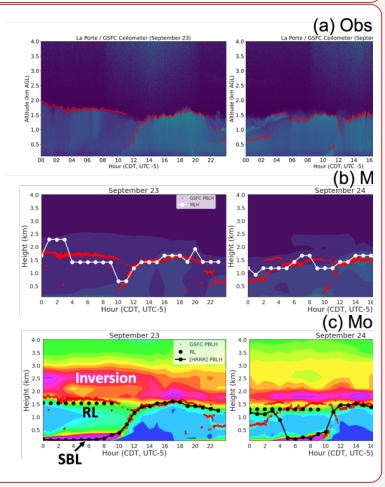
Deleted: top is…an be identified from the height where the RL (with little or low temperature increases at 0-3 °C km^{-1}) shifts …ransitions to the CI layer (with drastic temperature increases at 8-14 °C km^{-1}). Therefore, we use the sharpest gradient in the potential lapse rate, which is 6.6 °C km^{-1} on average, to it can be identified …dentify by…he top of the modeled RL the sharpest gradient in the potential lapse rate, which is 6.6 °C km^{-1} on average… This thermodynamically The modeled RL top …dentified top of the RL here ha… [17]

Deleted: The observed RL validates the backscatter-identified and thermodynamically-identified layers from the model…uch comparison can be found in Figure 3ej… The thermodynamically-identified layer Model results have…xhibits a slightly better agreement with the ceilometer defined MLH for the thermodynamically-identified layer, …bservation, with …xhibiting a correlation coefficient (R) of 0.77 and… a normalized mean bias (NMB) of 3%, and a root mean square error (RMSE) of 0.17 km. In contrast, , than for …he backscatter-identified layer shows slight… [18]



Deleted:

Deleted: layer



Deleted:

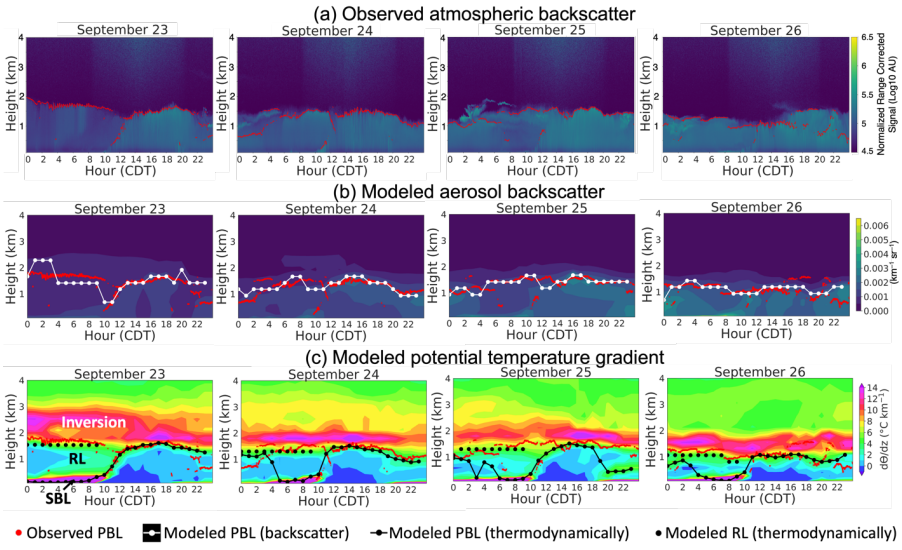


Figure 5. Same as Figure 4 but for the ozone episode of September 23-26.

4.2 Evaluation with HSRL-2

Section 4.1 above evaluated the model performance in continuous temporal variations using a ground-based ceilometer at the La Porte coastal site. In this section, our focus shifts to assessing the spatial variations of the modeled PBL heights by comparing them with the airborne HSRL-2 mixed layer heights. The HSRL-2 conducted measurements over the Houston region and the adjacent Galveston Bay, typically three times per day on eight high ozone days and two clean days in September 2021. As mentioned in the previous sections, the mixed layer can represent the PBL under various conditions. Therefore, from this point onwards, we refer to the observed mixed layer heights as observed PBL heights to maintain consistency with the modeled equivalents.

During ozone episodes, over land in the urban Houston region, the observed PBL heights gradually increase from 0.63±0.25 km in the morning (8:00-10:00 CDT), to 1.27±0.38 km at noon (11:00-13:00 CDT), and further to 1.69±0.23 km in the afternoon (14:00-16:00 CDT). Compared to land, the higher heat capacity in water leads to slower heating and cooling, resulting in a more stable atmosphere and shallower PBL. Over Galveston Bay, the observed heights are consistently lower by around 0.13-0.26 km during the three measured time periods. Such daytime variation and land-water differences can be observed on a specific high ozone day of September 9 in Figure 6. In comparison with these observations above, the four model simulations underestimate the PBL heights to different extents (NMB from -3% to -62%; RMSE from 0.22 km to 0.81 km) in Table 2. The model exhibits notably lower performance on land in the morning than under other conditions, showing less correlation and larger biases. This is because the morning mixed layer heights on land during ozone episodes can be difficult to

Deleted: This section
Deleted: we
Deleted: evaluates
Deleted: and temporal (daytime)
Deleted: collected
Deleted: for ten days
Deleted: during
Deleted: episodes
Deleted: stated
Deleted: above
Deleted: different
Formatted: Font: (Default) Times New Roman, Font color: Text 1, Pattern: Clear
Deleted: and we refer to the observed mixed layer heights as observed PBL hereafter in a manner consistent with the modeled equivalents.¶
Deleted: The four model simulations underestimate the observed PBL heights under most conditions.
Deleted: s
Deleted: the
Deleted: of
Formatted: Font: (Default) Times New Roman, Font color: Text 1, Pattern: Clear
Deleted: T
Formatted: Font: (Default) Times New Roman, Font color: Text 1, Pattern: Clear
Deleted: over water with values of 0.44±0.34 km, 1.14±0.45 km, and 1.43±0.24 km for the three time periods, respectively
Deleted: particularly show
Deleted: ess

retrieve with the influences from multiple layers (e.g., SBL and RL), and they can differ substantially from the thermodynamically defined PBL from the model. Therefore, we do not expect the model to capture the spatial patterns of mixed layer heights on land in the morning. Excluding this special case of morning PBL on land, we found that [HRRR] exhibits higher correlation (R=0.26-0.68) and lower biases (NMB from -3% to -21%; RMSE from 0.22 km to 0.46 km) in most case during ozone episodes.

During non-episode days, the observed PBL height increases from 0.78±0.14 km in the morning to 1.07±0.24 km at noon over land, and slightly from 0.57±0.28 km in the morning to 0.65±0.34 km at noon over water. The model captures such variations during clean days less effectively, resulting in lower correlation and larger biases compared to ozone episodes (Table 2). One important reason for the lower model performance during clean days compared to ozone episodes is the substantial difference in the number of data points collected. There are significantly fewer data points available during the two clean days compared to the eight high ozone days (Table 2).

The above analysis indicates that under both high ozone and clean days, the four model simulations consistently underestimate the observed PBL heights under most conditions. Next, we will investigate the model performances in simulating land-water differences.
The observed land-water differences in PBL heights are larger in the afternoon than in the morning during both ozone-episode and non-episode days. During noon and afternoon hours of ozone episodes, the model exhibits better performance in capturing the heights of the PBL on land than over water. This is expected as the model's physics parameterization is likely better tuned and calibrated for land surfaces, where more observations are available, compared to water. During the same period, the observed mean land-water differences of 0.13 km (noon) and 0.26 km (afternoon) are predicted to be larger in the model with values of 0.32-0.52 km (noon) and 0.44-0.56 km (afternoon), respectively. This is because the model shows consistently smaller underestimations over land than water during this period. During non-episode days, the observed land-water gradients of 0.21 km (morning) and 0.42 km (noon) are simulated to be 0.14-0.22 km (morning) and 0.36-0.76 km (noon), respectively. The [ACM2] and the [HRRR] slightly outperform the other two simulations for land-water differences (Table 2).

Deleted: Among the four simulations,

Deleted: show

Deleted: the best

Deleted: spatial

Deleted: s

Deleted: under

Deleted: conditions

Deleted: ¶
¶
¶
The four model simulations underestimate the observed PBL heights under most conditions (e.g., ozone episodes versus clean days, land versus water).¶
¶
Compared to observations, the four model simulations underestimate the heights 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 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 km in the afternoon). The model performs better in capturing the heights of the PBL on land than over water during noon and afternoon hours. This is expected as the model's physics and parameterization schemes are likely better tuned and calibrated for land surfaces, where more observations are available, compared to water.¶
¶

Formatted: Font: (Default) Times New Roman, Font color: Text 1, Pattern: Clear

Formatted: Font: (Default) Times New Roman, Font color: Text 1, Pattern: Clear

Formatted: Font: (Default) Times New Roman, Font color: Text 1, Pattern: Clear

Deleted: non-episodes

Formatted: Font: (Default) Times New Roman, Font color: Text 1, Pattern: Clear

Deleted: During non-episode days, the observed PBL height over land is 0.78±0.14 km in the morning and 1.07±0.24 km at noon. The morning heights are underestimated by 0.10-0.34 km in the four model simulations, while the noon heights can be either underestimated by 0.25-0.37 km or ... [19]

Formatted: Tab stops: 4.97 cm, Left

Formatted: Font: (Default) Times New Roman, Font color: Text 1, Pattern: Clear

Formatted: Font: (Default) Times New Roman, Font color: Text 1, Pattern: Clear

Moved up [7]: The observed land-water differences in PBL heights are larger in the afternoon than in the morning during

Moved (insertion) [7]

Deleted: The observed land-water differences in PBL heights are larger in the afternoon than in the morning during ... [20]

Deleted: One ozone-episode day, September 9, is selected to show the spatial characteristics of different simulation ... [21]

Table 2. Statistics of the observed HSRL-2 mixed layer height and the modeled thermodynamic PBL height during ozone-episode days (September 8-10 and 23-26) and non-episode days (September 1 and 3). Morning, noon, and afternoon are defined as 8:00-10:00 CDT, 11:00-13:00 CDT, and 14:00-16:00 CDT. Land and water are defined by the gray boxes in Figure 6. The correlation coefficient (R) and normalized mean bias (NMB) are unitless. The root mean square error (RMSE) has the same unit as PBL height in km.

	Simulations	Statistics	Morning		Noon		Afternoon	
			Land	Water	Land	Water	Land	Water
Ozone-episode		No. of points	1884	534	2273	570	2036	515
	[Base]	R	0.01	0.43	0.42	0.34	0.50	0.32
		NMB	-38%	-19%	-19%	-44%	-22%	-40%
		RMSE	0.38	0.32	0.44	0.67	0.46	0.70
	[YSU]	R	-0.02	0.29	0.47	0.21	0.51	0.12
		NMB	-62%	-56%	-15%	-52%	-18%	-43%
		RMSE	0.49	0.41	0.41	0.79	0.41	0.81
	[ACM2]	R	0.07	0.22	0.46	0.27	0.37	0.17
		NMB	-47%	-38%	-6%	-41%	-16%	-32%
		RMSE	0.43	0.38	0.38	0.68	0.48	0.67
	[HRRR]	R	0.05	0.40	0.68	0.50	0.58	0.26
		NMB	-41%	-6%	-3%	-20%	-4%	-21%
		RMSE	0.38	0.31	0.29	0.46	0.22	0.46
	Simulations	Statistics	Morning		Noon		Afternoon	
			Land	Water	Land	Water	Land	Water
Non-episode		No. of points	333	90	224	117		
	[Base]	R	0.17	0.05	-0.11	0.05		
		NMB	-27%	-38%	-23%	-43%		
		RMSE	0.31	0.35	0.43	0.44		
	[YSU]	R	0.19	0.06	-0.03	-0.20		
		NMB	-44%	-50%	-34%	-47%		
		RMSE	0.43	0.40	0.51	0.47		
	[ACM2]	R	0.2	-0.13	0.12	0.04		
		NMB	-30%	-42%	5%	-40%		
		RMSE	0.37	0.38	0.46	0.44		
	[HRRR]	R	0.23	-0.02	-0.02	0.21		
		NMB	-13%	-5%	22%	-16%		
		RMSE	0.24	0.30	0.38	0.36		

Deleted: Differences

Deleted: WRF-GC

Deleted: Moring

Formatted: Font: 12 pt

Deleted: bias

Deleted: difference

Deleted: and t

Deleted: difference are calculated by model simulations minus HSRL-2, with the unit of kilometers.

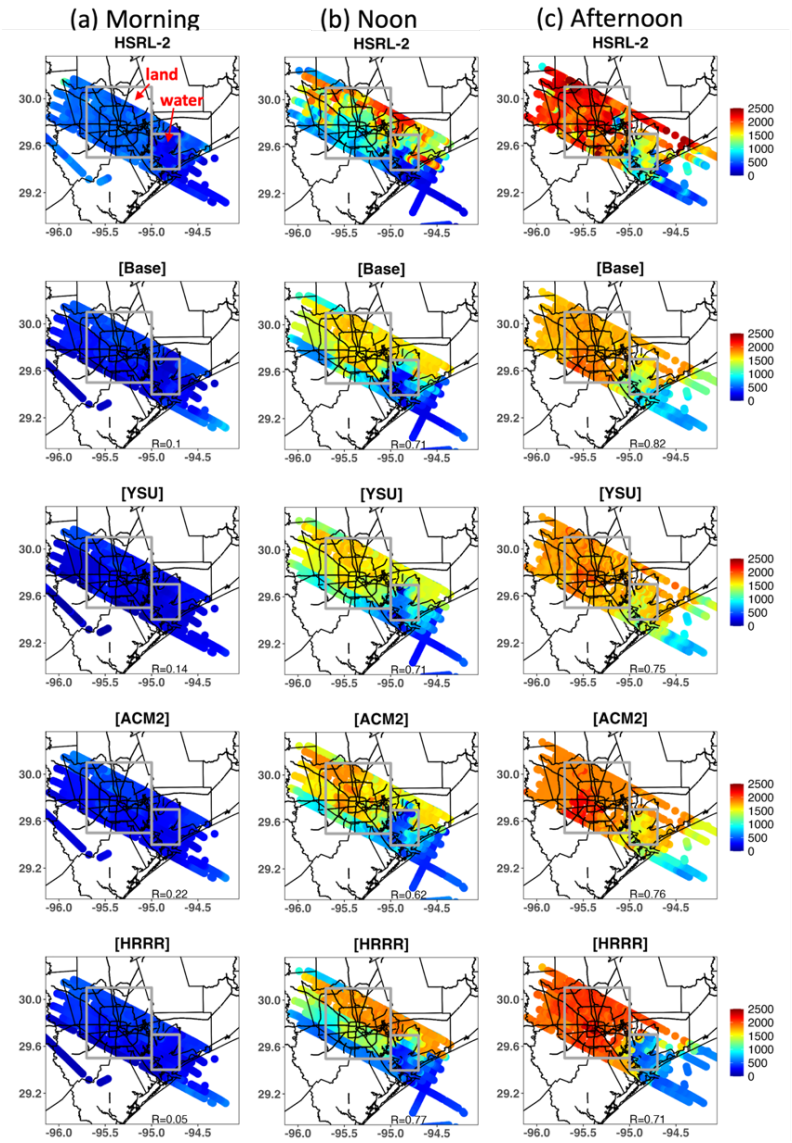


Figure 6. Spatial variabilities of the PBL heights (in meters) from the HSRL-2 and different WRF-GC simulations (a) in the morning (8:00-10:00 CDT), (b) at noon (11:00-13:00 CDT), and (c) in the afternoon (14:00-16:00 CDT) of September 9, 2021.

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 heights 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 in Fig. 4c and Fig. 5c, and then calculate the temperature differences in the model between the top and bottom of the CI layers on 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 (NMB from 0% to -2%; RMSE from 0.92 °C to 1.30 °C).

Combining the inversion strengths (Fig. 4c; Fig. 5c) and the vertical ozone distributions from the TROPOZ lidar (Fig. 8a; Fig. 8d) 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.

Deleted: variations

Deleted: i.e.

Deleted: (

Deleted:)

Deleted: in

Deleted: = -0.64 °C ~ -0.17 °C

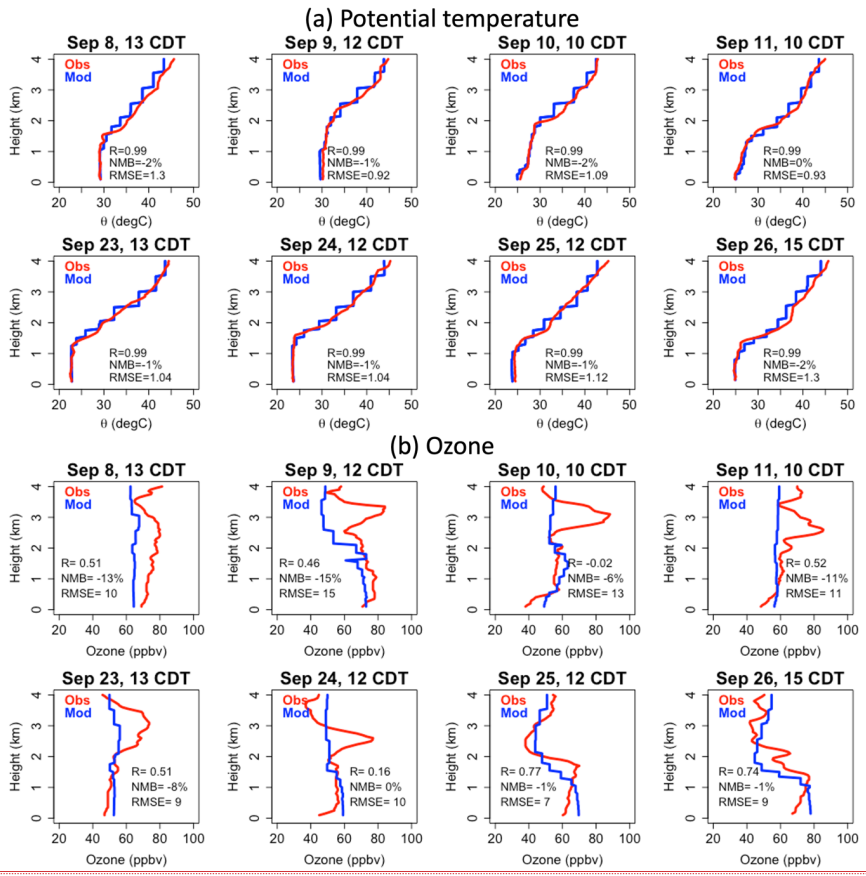
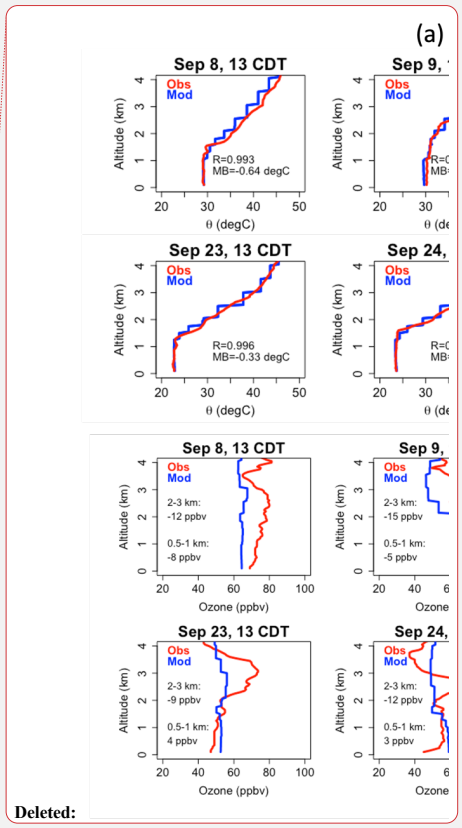


Figure 7. Vertical profiles of (a) potential temperature and (b) ozone from ozonesonde measurements and the WRF-GC [HRRR] simulation at La Porte during September 8-11 and September 23-26.

Deleted:



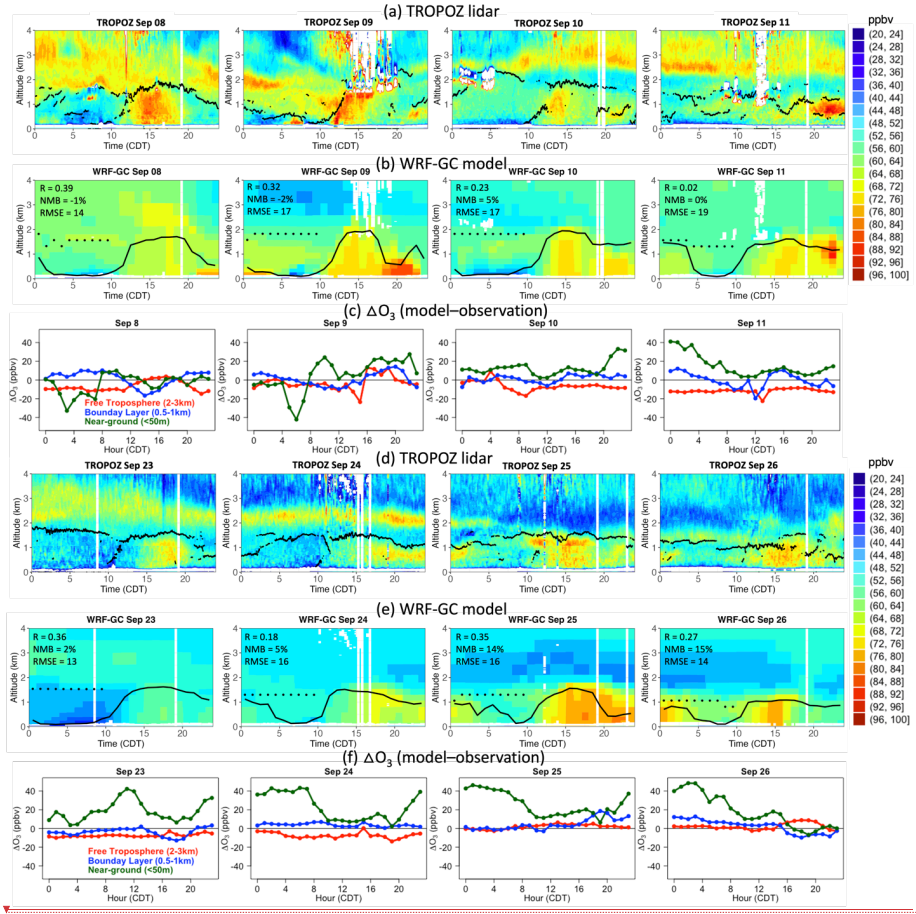
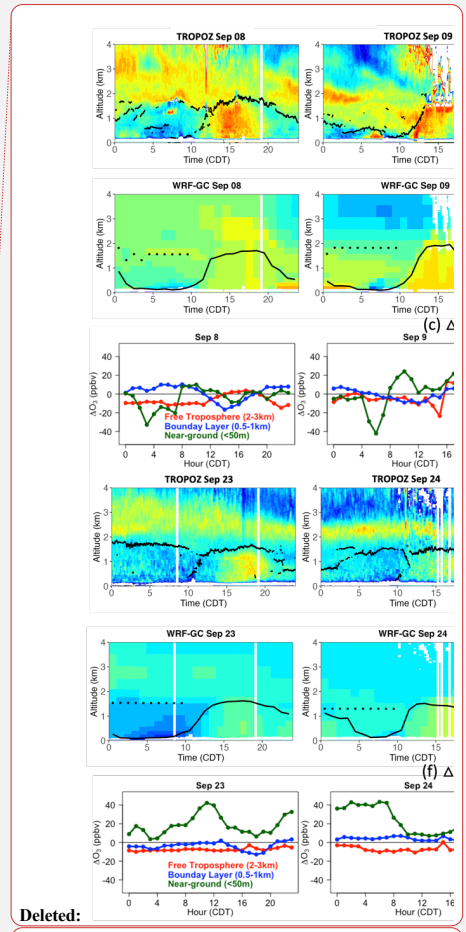
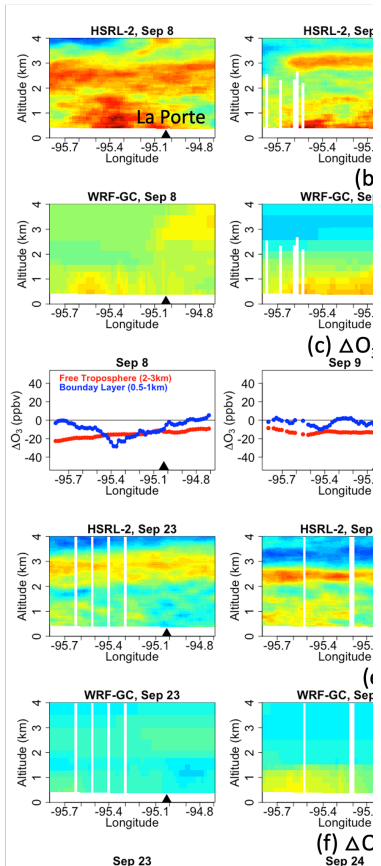


Figure 8. Time series of the vertical ozone profile from the TROPOZ ozone lidar (a, d) and the WRF-GC [HRRR] simulation (b, e) at La Porte. Observed and modeled boundary layer heights are inserted, respectively. Dots represent the modeled residual layer identified in this study. Line plots (c, f) show 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.

Deleted:





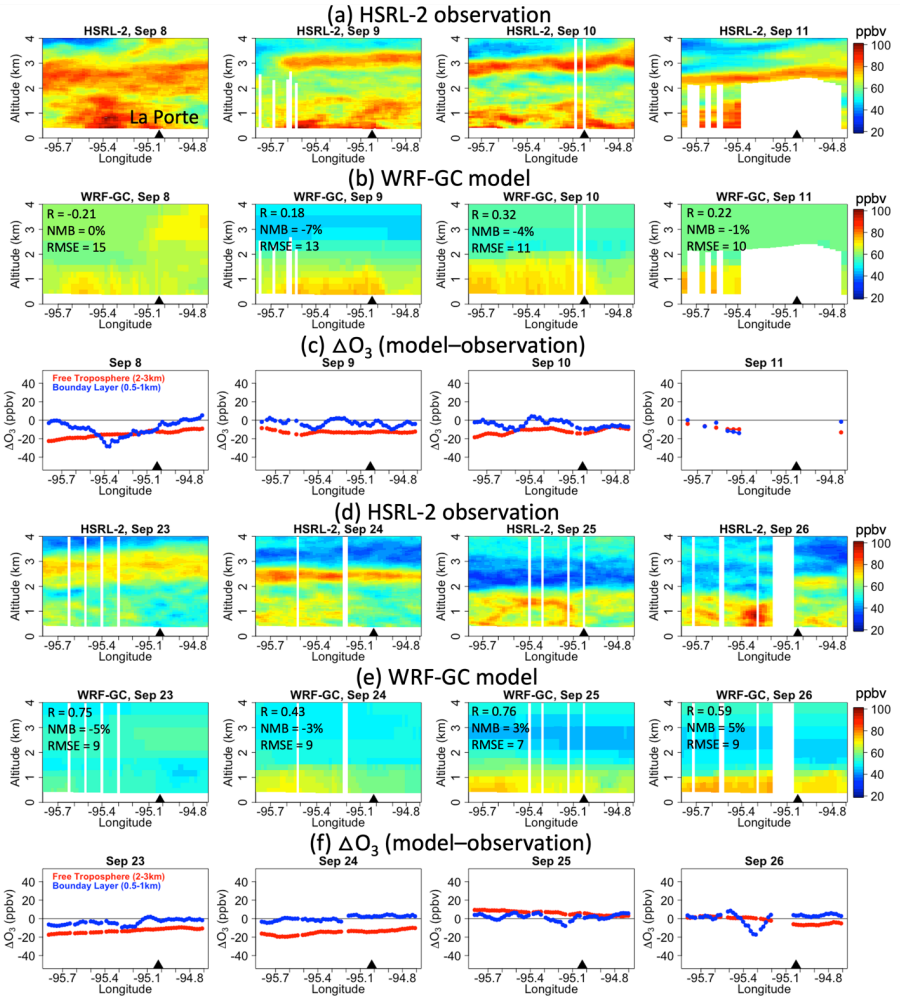


Figure 9. Vertical ozone profiles from (a, d) the HSRL-2 and (b, e) the WRF-GC [HRRR] simulation. The profiles are taken from a flight track (Fig.1) over urban Houston and Galveston Bay at around 11:00-13:00 CDT each day. Line plots (c, f) show ozone differences (model minus observation) at the free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

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 (< 50 m). 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).

To assess vertical variations below the first 4 km, we present performance metrics in Figure 7 for ozonesondes, Figure 8 for TROPOZ, and Figure 9 for HSRL-2. Different comparisons between observations and the model reflect distinct aspects. For instance, comparisons with ozonesonde pertain to vertical variations at a fixed location and time (R=0.46-0.77; NMB from -1% to -15%; RMSE=7-15 ppbv). This emphasis on a specific aspect explains why the correlation is higher

Deleted: cross compare

Deleted:

Deleted: Correlation

Deleted: ¶

Formatted: Font: (Default) Times New Roman, Font color: Text 1, Do not check spelling or grammar, Pattern: Clear

Formatted: Font: (Default) Times New Roman, Font color: Text 1, Do not check spelling or grammar, Pattern: Clear

Formatted: Font: (Default) Times New Roman, Font color: Text 1, Do not check spelling or grammar, Pattern: Clear

compared to TROPOZ and HSRL-2, which encompass a broader range of variations. Comparisons with TROPOZ relate to vertical and temporal variations at a fixed location (R=0.18-0.39; NMB from -2% to 15%; RMSE=13-17 ppbv). Comparisons with HSRL-2 represent a combination of vertical, temporal, and spatial variations (R=0.18-0.76; NMB from -7% to 5%; RMSE=7-13 ppbv). The above statistics exclude one or two extreme cases in each observation. Despite the differences in correlation resulting from the diverse representations of variations, biases are similar when compared to the three different observations.

Formatted: Font: (Default) Times New Roman, Font color: Text 1, Do not check spelling or grammar, Pattern: Clear

Formatted: Font: (Default) Times New Roman, Font color: Text 1, Do not check spelling or grammar, Pattern: Clear

Formatted: Font: (Default) Times New Roman, Font color: Text 1, Do not check spelling or grammar, Pattern: Clear

Deleted: ¶

Table 3. Absolute (abs.) and percentage (pct.) ozone differences between field measurements and the model at the 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%

Table 4. Absolute (abs.) and percentage (pct.) ozone differences between field measurements and 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. 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 is critical in thoroughly evaluating the spatial and temporal variations of the PBL heights 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 both observations, the four model simulations of [Base], [YSU], [ACM2], and [HRRR] generally underestimate the PBL heights. When compared with the ceilometer, the [HRRR] captures the diurnal variations during non-episode days (R=0.88; NMB=-21%; RMSE=0.22 km). 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%; RMSE=0.17 km) compares slightly better than that of the backscatter-identified layer (R=0.72; NMB=22%; RMSE=0.30 km) during ozone episodes. Meanwhile, when compared with the HSRL-2, the [HRRR] exhibits higher correlation (R=0.26-0.68) and lower biases (NMB from -3% to -21%; RMSE=0.22-0.46 km) than other simulations during noon and afternoon hours during ozone episodes. For land-water differences in PBL heights, the water has shallower PBL heights compared to land. The model predicts larger land-water differences than observations because the model consistently underestimates PBL heights over land compared to water.

Deleted: are

Deleted: ¶

Deleted: Compared with the ceilometer,

Deleted: 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,

Deleted: 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.

Deleted: 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, ozonesondes, a model 49i ozone analyzer, and a CAMS site named La Porte Sylvan Beach. First, individual evaluations were conducted 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). Results show that 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. Second, we assessed vertical variations of ozone below the first 4 km through comparisons with TROPOZ, HSRL-2, and ozonesondes. Correlation is higher when compared to ozonesondes than TROPOZ and HSRL-2, as ozonesondes emphasize a specific aspect of vertical variations at a fixed location and time, while the other two encompass a broader range of variations with temporal and spatial variations included. Despite the differences in correlation resulting from the diverse representations of variations, biases shown by these three comparisons are similar.

Based on these evaluations, we summarized model limitations that prevent a 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 night. Further efforts are needed to identify and incorporate the RL into the model's standard outputs. Alternative modules aimed at identifying the PBL using simulated vertical backscatter gradients can also enhance the validation of PBL heights with backscatter-derived observations. The second limitation is the underestimation of the layer of enhanced ozone 2-3 km aloft in the free troposphere that was often present on ozone-episode days during the campaign. Given its height of 2-3 km and a lifetime of around a week, the layer of enhanced ozone was likely transported into Houston by synoptic flows of cold fronts from the north. The underrepresentation of the synoptic layer of enhanced ozone affects model representations across regions horizontally and atmospheric layers vertically, making it particularly important to model vertical ozone distributions and the effects of entrainment accurately.

Our findings have implications for the predictivity of ozone's vertical mixing and distribution across different modeling systems. For example, WRF is widely used in various meteorology-chemistry coupling systems with different treatments of boundary layer mixing. In WRF-Chem, boundary layer mixing in the chemistry part uses a mixing coefficient originating in WRF such that the boundary layer mixing calculations in the meteorology and chemistry parts share the same set of coefficients. In WRF-GC, the chemistry part from GEOS-Chem only takes the PBL height from WRF as the maximum height for boundary layer mixing but conducts independent calculations of boundary layer mixing using its own internal methods, which are not reliant on WRF. Unlike online coupled WRF-Chem and WRF-GC, WRF is offline coupled to CAMx in the WRF-CAMx system, and the boundary layer mixing in the chemistry part of CAMx is subject to WRF output frequency instead of the native transport time step in WRF. Considering these

Deleted: E

Deleted: done

Deleted: As a result,

Deleted: of the model limitations

Deleted: and

Deleted: from

Formatted: Font: (Default) Times New Roman, (Asian) +Body Asian (DengXian), Font color: Text 1, Pattern: Clear

distinct treatments of boundary layer mixing in models, the single-layer PBL representation can have varying impacts on the simulation of vertical mixing and, consequently, the vertical distribution of ozone and other air pollutants. Thus, it is essential to understand the differences in boundary layer mixing among different meteorology-chemistry coupling systems. Follow-up studies to this work will address these aspects with a detailed analysis of vertical mixing processes in various models.

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

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

Author contributions. XL and YW conceived the research idea. XL wrote the initial draft of the paper and performed the analyses and model simulations. JF, TG, and SA provided the shipborne data. JS, MR, and LT provided the TROPOZ and ceilometer data. PW and JS provided the ozonesonde data. JH, TS, AJS, and MF provided the HSRL-2 data. All authors contributed to the interpretation of the results and the preparation of the paper.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Acknowledgements. The authors acknowledge TCEQ for providing the hourly wind, temperature, relative humidity, and MDA8 ozone data, and NASA Langley Atmospheric Science Data Center for providing the TRACER-AQ data archive. We thank Richard Ferrare for helpful suggestions on this paper.

Financial support. This research was supported by the Texas Commission on Environmental Quality (TCEQ) (Grant No. 582-22-31544-019) and by a grant from the Texas Air Quality Research Program (AQRP) (22-008) at The University of Texas at Austin through the Texas Emission Reduction Program (TERP) and the TCEQ. The findings, opinions, and conclusions are the work of the author(s) and do not necessarily represent the findings, opinions, or conclusions of the AQRP or the TCEQ.

References

Formatted: Font: (Default) Times New Roman, (Asian) +Body Asian (DengXian), Font color: Text 1, Pattern: Clear

Deleted: how

Formatted: Font: (Default) Times New Roman, (Asian) +Body Asian (DengXian), Font color: Text 1, Pattern: Clear

Deleted: the model limitation of a single-layer PBL representation affects

Deleted: in chemical simulations

Deleted: intercomparison

Formatted: Font: (Asian) +Body Asian (DengXian)

Deleted: open source

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Deleted: GG and TB provided the LMOL data.

Deleted: ¶

Angevine, W., Senff, C., and Westwater, E.: Boundary layers/Observational Techniques-Remote, in: Encyclopedia of Atmospheric Sciences, edited by: Holton, J. R., Academic Press, Oxford, 271– 279, https://doi.org/10.1016/B0-12-227090-8/00089-0, 2003.

Banta, R. M., Senff, C. J., Nielsen-Gammon, J., Darby, L. S., Ryerson, T. B., Alvarez, R. J., Sandberg, S. R., Williams, E. J., and Trainer, M.: A bad air day in Houston, B. Am. Meteorol. Soc., 86, 657–669, doi:10.1175/bams-86-5-657, 2005.

Banta, R. M., Senff, C. J., Alvarez, R. J., Langford, A. O., Parrish, D. D., Trainer, M. K., Darby, L. S., Hardesty, R. M., Lambeth, B., Neuman, J. A., Angevine, W. M., Nielsen-Gammon, J., Sandberg, S. P., and White, A. B.: Dependence of daily peak O3 concentrations near Houston, Texas on environmental factors: wind speed, temperature, and boundary-layer depth, Atmos. Environ., 45, 162–173, doi:10.1016/j.atmosenv.2010.09.030, 2011.

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

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

Bodini, N., Lundquist, J. K., and Newsom, R. K.: Estimation of turbulence dissipation rate and its variability from sonic anemometer and wind Doppler lidar during the XPIA field campaign, At- mos. Meas. Tech., 11, 4291–4308, https://doi.org/10.5194/amt- 11-4291-2018, 2018.

Bonin, T. A., Newman, J. F., Klein, P. M., Chilson, P. B., and Wharton, S.: Improvement of vertical velocity statistics measured by a Doppler lidar through comparison with sonic anemometer observations, Atmos. Meas. Tech., 9, 5833–5852, https://doi.org/10.5194/amt-9-5833-2016, 2016.

Bonin, T. A., Carroll, B. J., Hardesty, R. M., Brewer, W. A., Hajny, K., Salmon, O. E., and Shepson, P. B.: Doppler Lidar Observations of the Mixing Height in Indianapolis Using an Automated Composite Fuzzy Logic Approach, J. Atmos. Ocean. Tech., 35, 473–490, https://doi.org/10.1175/JTECH-D-17-0159.1, 2018.

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

Caicedo, V., Rappenglück, B., Lefer, B., Morris, G., Toledo, D., and Delgado, R.: Comparison of aerosol lidar retrieval methods for boundary layer height detection using ceilometer aerosol

Deleted: ¶

Formatted: Font: Times New Roman, 12 pt

Formatted: Font: Times New Roman, 12 pt, Font color: Text 1

Formatted: Pattern: Clear

Formatted: Font: Times New Roman, 12 pt, Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font: Times New Roman, 12 pt, Font color: Text 1

Formatted: Pattern: Clear

backscatter data, Atmos. Meas. Tech., 10, 1609–1622, https://doi.org/10.5194/amt-10-1609-2017, 2017.

Caicedo, V., Rappenglück, B., Cuchiara, G., Flynn, J., Ferrare, R., Scarino, A., Berkoff, T., Senff, C., Langford, A., and Lefer, B.: Bay breeze and sea breeze circulation impacts on the planetary boundary layer and air quality from an observed and modeled DISCOVER-AQ Texas case study, J. Geophys. Res.-Atmos., 124, 7359–7378, https://doi.org/10.1029/2019JD030523, 2019.

Caicedo, V., Delgado, R., Sakai, R., Knepp, T., Williams, D., Cavender, K., Lefer, B., and Szykman, J.: An Automated Common Algorithm for Planetary Boundary Layer Retrievals Using Aerosol Lidars in Support of the U.S. EPA Photochemical Assessment Monitoring Stations Program, J. Atmos. Oceanic Technol., 37, 1847–1864, https://doi.org/10.1175/JTECH-D-20-0050.1, 2020.

Caputi, D. J., Faloona, I., Trousdell, J., Smoot, J., Falk, N., and Conley, S.: Residual layer ozone, mixing, and the nocturnal jet in California's San Joaquin Valley, Atmos. Chem. Phys., 19, 4721–4740, https://doi.org/10.5194/acp-19-4721-2019, 2019.

Chen, F. and Dudhia, J.: Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part II: Preliminary model validation, Monthly Weather Review, 129(4), 587-604, 2001.

Cimini, D., De Angelis, F., Dupont, J.-C., Pal, S., and Haeffelin, M.: Mixing layer height retrievals by multichannel microwave radiometer observations, Atmos. Meas. Tech., 6, 2941–2951, https://doi.org/10.5194/amt-6-2941-2013, 2013.

Cuchiara, G. C., Li, X., Carvalho, J., and Rappenglück, B.: Intercomparison of planetary boundary layer parameterization and its impacts on surface ozone concentration in the WRF/Chem model for a case study in Houston/Texas, Atmos. Environ., 96, 175– 185, doi:10.1016/j.atmosenv.2014.07.013, 2014.

Dacic, N., Sullivan, J. T., Knowland, K. E., Wolfe, G. M., Oman, L. D., Berkoff, T. A., and Gronoff, G. P.: Evaluation of NASA's high-resolution global composition simulations: Understanding a pollution event in the Chesapeake Bay during the summer 2017 OWLETS campaign, Atmos. Environ., 222, 117133, https://doi.org/10.1016/j.atmosenv.2019.117133, 2020.

Dreessen, J., Sullivan, J., and Delgado, R.: Observations and impacts of transported Canadian wildfire smoke on ozone and aerosol air quality in the Maryland region on June 9–12, 2015, J. Air Waste Manage., 66, 842–862, https://doi.org/10.1080/10962247.2016.1161674, 2016.

Soleimanian, E., Wang, Y., Li, W., Liu, X., Griggs, T., Flynn, J., Walter, P.J. and Estes, M.J.: Understanding ozone episodes during the TRACER-AQ campaign in Houston, Texas: The role of transport and ozone production sensitivity to precursors, Science of The Total Environment, 2023.

Formatted: Font: Times New Roman, 12 pt, Font color: Text 1

Formatted: Normal, Don't adjust right indent when grid is defined, Don't adjust space between Latin and Asian text, Don't adjust space between Asian text and numbers

Formatted: Font color: Text 1

Deleted: ¶

Formatted: Font: (Default) Times New Roman, 12 pt, Font color: Text 1, Pattern: Clear

Formatted: Font: (Default) Times New Roman, 12 pt, Font color: Text 1, Pattern: Clear

Formatted: Font: (Default) Times New Roman, 12 pt, Font color: Text 1, Pattern: Clear

Formatted: Font: (Default) Times New Roman, 12 pt, Font color: Text 1, Pattern: Clear

Feng, X., Lin, H., Fu, T.-M., Sulprizio, M. P., Zhuang, J., Jacob, D. J., Tian, H., Ma, Y., Zhang, L., Wang, X., Chen, Q., and Han, Z.: WRF-GC (v2.0): online two-way coupling of WRF (v3.9.1.1) and GEOS-Chem (v12.7.2) for modeling regional atmospheric chemistry–meteorology interactions, Geosci. Model Dev., 14, 3741–3768, https://doi.org/10.5194/gmd-14-3741-2021, 2021.

Geiß, A., Wiegner, M., Bonn, B., Schäfer, K., Forkel, R., von Schneidemesser, E., Münkel, C., Chan, K. L., and Nothard, R.: Mixing layer height as an indicator for urban air quality?, Atmos. Meas. Tech., 10, 2969–2988, https://doi.org/10.5194/amt10-2969-2017, 2017.

Gronoff, G., Robinson, J., Berkoff, T., Swap, R., Farris, B., Schroeder, J., Halliday, H. S., Knepp, T., Spinei, E., Car- rion, W., Adcock, E. E., Johns, Z., Allen, D., and Pip- pin, M.: A Method for Quantifying near Range Point Source Induced O3 Titration Events Using Co-Located Li- dar and Pandora Measurements, Atmos. Environ., 204, 43–52, https://doi.org/10.1016/j.atmosenv.2019.01.052, 2019.

Gronoff, G., Berkoff, T., Knowland, K. E., Lei, L., Shook, M., Fab- bri, B., Carrion, W., and Langford, A. O.: Case study of strato- spheric Intrusion above Hampton, Virginia: lidar-observation and modeling analysis, Atmos. Environ., 259, 1352–2310, https://doi.org/10.1016/j.atmosenv.2021.118498, 2021.

Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., and Wang, X.: The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions, Geosci. Model Dev., 5, 1471–1492, doi:10.5194/gmd-5- 1471-2012, 2012.

Haman, C. L., Couzo, E., Flynn, J. H., Vizuete, W., Heffron, B., and Lefer, B. L.: Relationship between boundary layer heights and growth rates with ground-level ozone in Houston, Texas, J. Geophys. Res.-Atmos., 119, 6230–6245, https://doi.org/10.1002/2013jd020473, 2014.

Hair, J. W., Hostetler, C. A., Cook, A. L., Harper, D. B., Ferrare, R. A., Mack, T. L., Welch, W., Izquierdo, L. R., and Hovis, F. E.: Airborne High Spectral Resolution lidar for profiling aerosol optical properties, Appl. Opt., 47, 6734–6752, 2008.

Hair, J., Hostetler, C., Cook, A., Harper, D., Notari, A., Fenn, M., Newchurch, M., Wang, L., Kuang, S., Knepp, T., Burton, S., Ferrare, R., Butler, C., Collins, J., and Nehrir, A.: New capability for ozone dial profiling measurements in the troposphere and lower stratosphere from aircraft, EPJ Web Conf., 176, 01006, https://doi.org/10.1051/epjconf/201817601006, 2018.

Hong, S.-Y., Noh, Y., and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes, Mon. Weather Rev., 134, 2318–2341, https://doi.org/10.1175/MWR3199.1, 2006.

Hong, S. Y. and Lim, J. O. J.: The WRF Single-Moment 6-Class Microphysics Scheme (WSM6), J. Korean Meteor. Soc., 42, 129– 151, 2006.

Hu, X. M., Nielsen-Gammon, J. W., and Zhang, F.: Evaluation of three planetary boundary layer schemes in the WRF model, J. Appl. Meteorol. Clim., 49, 1831–1844, doi:10.1175/2010JAMC2432.1, 2010.

Hudman, R. C., Moore, N. E., Mebust, A. K., Martin, R. V., Russell, A. R., Valin, L. C., and Cohen, R. C.: Steps towards a mechanistic model of global soil nitric oxide emissions: implementation and space based-constraints, Atmos. Chem. Phys., 12, 7779–7795, https://doi.org/10.5194/acp-12-7779-2012, 2012.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.: Radiative forcing by longlived greenhouse gases: Calculations with the AER radiative transfer models, J. Geophys. Res.-Atmos., 113, D13103, https://doi.org/10.1029/2008JD009944, 2008.

Jaffe, D.: Relationship between Surface and Free Tropospheric Ozone in the Western U.S, Environ. Sci. Technol, 45, 432–438, doi:10.1021/es1028102, 2011.

Jaffe, D. A., Cooper, O. R., Fiore, A. M., Henderson, B. H., Tonnesen, G. S., Russell, A. G., Henze, D. K., Langford, A. O., Lin, M., and Moore, T.: Scientific assessment of background ozone over the U.S.: Implications for air quality management, Elem. Sci. Anth., 6, p. 56, https://doi.org/10.1525/elementa.309, 2018.

Jensen, M. P., Flynn, J. H., Judd, L. M., Kollias, P., Kuang, C., Mcfarquhar, G., Nadkarni, R., Powers, H., and Sullivan, J.: A Succession of Cloud, Precipitation, Aerosol, and Air Quality Field Experiments in the Coastal Urban Environment, B. Am. Meteorol. Soc., 103, 103–105, 2022.

Johnson, M., Kuang, S., Wang, L., and Newchurch, M.: Evaluating summer-time ozone enhancement events in the southeast United States, Atmosphere, 7, 108, https://doi.org/10.3390/atmos7080108, 2016.

Kaser, L., Patton, E. G., Pfister, G. G., Weinheimer, A. J., Montzka, D. D., Flocke, F., Thompson, A. M., Stauffer, R. M., and Halliday, H. S.: The effect of entrainment through atmospheric boundary layer growth on observed and modeled surface ozone in the Colorado Front Range, J. Geophys. Res.-Atmos., 122, 6075– 6093, https://doi.org/10.1002/2016JD026245, 2017.

Knepp, T. N., Szykman, J. J., Long, R., Duvall, R. M., Krug, J., Beaver, M., Cavender, K., Kronmiller, K., Wheeler, M., Delgado, R., Hoff, R., Berkoff, T., Olson, E., Clark, R., Wolfe, D., Van Gilst, D., and Neil, D.: Assessment of mixed-layer height estimation from single-wavelength ceilometer profiles, Atmos. Meas. Tech., 10, 3963–3983, https://doi.org/10.5194/amt-10-3963-2017, 2017.

Knote, C., Tuccella, P., Curci, G., Emmons, L., Orlando, J. J., Madronich, S., Baró, R., Jiménez-Guerrero, P., Luecken, D., Hogrefe, C., Forkel, R., Werhahne, J., Hirtl, M., Pérez, J., José, R.,

Moved (insertion) [6]

Giordano, L., Brunner, D., Yahya, K., and Zhang, Y.: Influence of the choice of gas-phase mechanism on predictions of key gaseous pollutants during the AQMEII phase-2 intercomparison, Atmos. Environ., 115, 553– 568, https://doi.org/10.1016/j.atmosenv.2014.11.066, 2015.

Kotthaus, S., Bravo-Aranda, J. A., Collaud Coen, M., Guerrero-Rascado, J. L., Costa, M. J., Cimini, D., O'Connor, E. J., Hervo, M., Alados-Arboledas, L., Jiménez-Portaz, M., Mona, L., Ruffieux, D., Illingworth, A., and Haeffelin, M.: Atmospheric boundary layer height from ground-based remote sensing: a review of capabilities and limitations, Atmos. Meas. Tech., 16, 433–479, https://doi.org/10.5194/amt-16-433-2023, 2023.

Kotsakis, A., Sullivan, J.T., Hanisco, T.F., Swap, R.J., Caicedo, V., Berkoff, T.A., Gronoff, G., Loughner, C.P., Ren, X., Luke, W.T. and Kelley, P.: Sensitivity of total column NO2 at a marine site within the Chesapeake Bay during OWLETS-2, Atmos. Environ., 277, 119063, 2022.

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, W., Wang, Y., Liu, X., Soleimanian, E., Griggs, T., Flynn, J., and Walter, P.: Understanding offshore high-ozone events during TRACER-AQ 2021 in Houston: Insights from WRF-CAMx photochemical modeling, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2023-1117, 2023.

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.

Formatted: Font: (Default) Times New Roman, 12 pt, Font color: Text 1, Pattern: Clear

Moved up [6]: Jensen, M. P., Flynn, J. H., Judd, L. M., Kollias, P., Kuang, C., Mcfarquhar, G., Nadkarni, R., Powers, H., and Sullivan, J.: A Succession of Cloud, Precipitation, Aerosol, and Air Quality Field Experiments in the Coastal Urban Environment, B. Am. Meteorol. Soc., 103, 103–105, 2022.¶

Deleted: ¶

Formatted: Font color: Auto, Do not check spelling or grammar

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.

Nakanishi, M. and Niino, H.: Development of an improved turbulence closure model for the atmospheric boundary layer, J. Meteorol. Soc. Jpn., 87, 895–912, https://doi.org/10.2151/jmsj.87.895, 2009.

Otte, T. L.: The impact of nudging in the meteorological model for retrospective air quality simulations. Part I: Evaluation against national observation networks, J. Appl. Meteor. Climatol., 47, 1853–1867, 2008.

Parrish, D. D., Aikin, K. C., Oltmans, S. J., Johnson, B. J., Ives, M., and Sweeny, C.: Impact of transported background ozone inflow on summertime air quality in a California ozone exceedance area, Atmos. Chem. Phys., 10, 10093–10109, doi:10.5194/acp10-10093-2010, 2010.

Pleim, J. E.: A combined local and nonlocal closure model for the atmospheric boundary layer, Part I: model description and testing, J. Appl. Meteor. Clim., 46, 1383–1395, 2007a.

Pleim, J. E.: A combined local and nonlocal closure model for the atmospheric boundary layer, Part II: application and evaluation in a mesoscale meteorological model, J. Appl. Meteor. Clim., 46, 1396–1409, 2007b.

Rappenglück, B., Perna, R., Zhong, S., and Morris, G. A.: An analysis of the vertical structure of the atmosphere and the upper-level meteorology and their impact on surface ozone levels in Houston, Texas, J. Geophys. Res.-Atmos., 113, D17315, doi:10.1029/2007JD009745, 2008.

Rizza, U., Barnaba, F., Miglietta, M. M., Mangia, C., Di Liberto, L., Dionisi, D., Costabile, F., Grasso, F., and Gobbi, G. P.: WRF-Chem model simulations of a dust outbreak over the central Mediterranean and comparison with multi-sensor desert dust observations, Atmos. Chem. Phys., 17, 93–115, https://doi.org/10.5194/acp-17-93-2017, 2017.

Scarino, A. J., Obland, M. D., Fast, J. D., Burton, S. P., Ferrare, R. A., Hostetler, C. A., Berg, L. K., Lefer, B., Haman, C., Hair, J. W., Rogers, R. R., Butler, C., Cook, A. L., and Harper, D. B.: Comparison of mixed layer heights from airborne high spectral resolution lidar, ground-based measurements, and the WRFChem model during CalNex and CARES, Atmos. Chem. Phys., 14, 5547–5560, doi:10.5194/acp-14-5547-2014, 2014.

Schnell, J. L., Prather, M. J., Josse, B., Naik, V., Horowitz, L. W., Cameron-Smith, P., Bergmann, D., Zeng, G., Plummer, D. A., Sudo, K., Nagashima, T., Shindell, D. T., Faluvegi, G., and Strode, S. A.: Use of North American and European air quality networks to evaluate global chemistry–climate modeling of surface ozone, Atmos. Chem. Phys., 15, 10581–10596, doi:10.5194/acp-15-10581-2015, 2015.

Soleimanian, E., Wang, Y., Li, W., Liu, X., Wasti, S., Griggs, T., Flynn J., Walter P., and Estes M.: Understanding ozone episodes during TRACER-AQ campaign in Houston, Texas: the role of transport and ozone production sensitivity to precursors, Science of the Total Environment, https://doi.org/10.1016/j.scitotenv.2023.165881, 2023.

Stull, R.: An introduction to Atmospheric Boundary Layer Meteorology, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1988.

Sullivan, J. T., McGee, T. J., Sumnicht, G. K., Twigg, L. W., and Hoff, R. M.: A mobile differential absorption lidar to measure sub-hourly fluctuation of tropospheric ozone profiles in the Baltimore–Washington, D.C. region, Atmos. Meas. Tech., 7, 3529–3548, https://doi.org/10.5194/amt-7-3529-2014, 2014.

Sullivan, J. T., McGee, T. J., Thompson, A. M., Pierce, R. B., Sumnicht, G. K., Twigg, L. W., Eloranta, E., and Hoff, R. M.: Characterizing the lifetime and occurrence of stratospheric-tropospheric exchange events in the rocky mountain region using high-resolution ozone measurements, J. Geophys. Res.-Atmos., 120, 12410–12424, 2015.

Sullivan, J. T., Rabenhorst, S. D., Dreessen, J., McGee, T. J., Delgado, R., Twigg, L., and Sumnicht, G.: Lidar observations revealing transport of O3 in the presence of a nocturnal low-level jet: Regional implications for “next-day” pollution, Atmos. Environ., 158, 160–171, 2017.

Sullivan, J. T., Berkoff, T., Gronoff, G., Knepp, T., Pippin, M., Allen, D., Twigg, L., Swap, R., Tzortziou, M., Thompson, A. M., and Stauffer, R. M.: The ozone water–land environmental transition study: An innovative strategy for understanding Chesapeake Bay pollution events, B. Am. Meteorol. Soc., 100, 291–306, 2019.

Sullivan, J.T., Apituley, A., Mettig, N., Kreher, K., Knowland, K.E., Allaart, M., Piters, A., Van Roozendael, M., Veefkind, P., Ziemke, J.R. and Kramarova, N.: Tropospheric and stratospheric ozone profiles during the 2019 TROpomi vaLIdation eXperiment (TROLIX-19), Atmos. Chem. Phys., 22(17), 11137-11153, 2022.

Deleted: Soleimanian, E., ¶

Formatted: Font: (Asian) Times New Roman, Font color: Auto

Formatted: Font: (Asian) Times New Roman, Not Bold, Font color: Auto

Formatted: Font: (Asian) Times New Roman, Font color: Auto

Formatted: Font: (Asian) Times New Roman, Not Italic, Font color: Auto

Formatted: Font: (Asian) Times New Roman, Font color: Auto

Formatted: Default Paragraph Font

Formatted: Font: (Asian) Times New Roman, Font color: Auto

Tangborn, A., Demoz, B., Carroll, B. J., Santanello, J., and Anderson, J. L.: Assimilation of lidar planetary boundary layer height observations, Atmos. Meas. Tech., 14, 1099–1110, https://doi.org/10.5194/amt-14-1099-2021, 2021.

Tiedtke, M.: A Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale Models, Monthly Weather Review, 117, 1779–1800, 1989.

Tucker, S. C., Banta, R. M., Langford, A. O., Senff, C. J., Brewer, W. A., Williams, E. J., Lerner, B. M., Osthoff, H., and Hardesty, R. M.: Relationships of coastal nocturnal boundary layer winds and turbulence to Houston ozone concentrations during TexAQS 2006, J. Geophys. Res., 115, D10304, doi:10.1029/2009JD013169, 2010.

TCEQ: Texas Commission on Environmental Quality (2022). Houston-Galveston-Brazoria: Current attainment status, available at: https://www.tceq.texas.gov/airquality/sip/hgb/hgb-status (last access: 12 Jan 2023)

Travis, K. R. and Jacob, D. J.: Systematic bias in evaluating chemical transport models with maximum daily 8 h average (MDA8) surface ozone for air quality applications: a case study with GEOS-Chem v9.02, Geosci. Model Dev., 12, 3641–3648, https://doi.org/10.5194/gmd-12-3641-2019, 2019.

Travis, K. R., Jacob, D. J., Fisher, J. A., Kim, P. S., Marais, E. A., Zhu, L., Yu, K., Miller, C. C., Yantosca, R. M., Sulprizio, M. P., Thompson, A. M., Wennberg, P. O., Crounse, J. D., St. Clair, 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.

Wulfmeyer, V., Pal, S., Turner, D. D., and Wagner, E.: Can Water Vapour Raman Lidar Resolve Profiles of Turbulent Variables in the Convective Boundary Layer?, Bound.-Lay. Meteorol., 136, 253–284, https://doi.org/10.1007/s10546-010-9494-z, 2010.

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.

Formatted: Font: Times New Roman, 12 pt, Font color: Text 1

Formatted: Normal

Formatted: Font color: Text 1

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.

Yu, E., Bai, R., Chen, X., and Shao, L.: Impact of physical parameterizations on wind simulation with WRF V3.9.1.1 under stable conditions at planetary boundary layer gray-zone resolution: a case study over the coastal regions of North China, Geosci. Model Dev., 15, 8111–8134, https://doi.org/10.5194/gmd-15-8111-2022, 2022.

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.

Zhao, D., Xin, J., Wang, W., Jia, D., Wang, Z., Xiao, H., Liu, C., Zhou, J., Tong, L., Ma, Y. and Wen, T.: Effects of the sea-land breeze on coastal ozone pollution in the Yangtze River Delta, China, Science of The Total Environment, 807, 150306, 2022.

Zhou, W., Cohan, D. S., and Henderson, B. H.: Slower ozone production in Houston, Texas following emission reductions: evidence from Texas Air Quality Studies in 2000 and 2006, Atmos. Chem. Phys., 14, 2777–2788, doi:10.5194/acp-14-2777- 2014, 2014.

Formatted: Font: (Default) Times New Roman, 12 pt, Font color: Text 1, Pattern: Clear

Page 14: [18] Deleted **Liu, Xueying** **8/3/23 10:15:00 PM**

Page 14: [18] Deleted **Liu, Xueying** **8/3/23 10:15:00 PM**

Page 14: [18] Deleted **Liu, Xueying** **8/3/23 10:15:00 PM**

Page 14: [18] Deleted **Liu, Xueying** **8/3/23 10:15:00 PM**

Page 14: [18] Deleted **Liu, Xueying** **8/3/23 10:15:00 PM**

Page 14: [18] Deleted **Liu, Xueying** **8/3/23 10:15:00 PM**

Page 14: [18] Deleted **Liu, Xueying** **8/3/23 10:15:00 PM**

Page 14: [18] Deleted **Liu, Xueying** **8/3/23 10:15:00 PM**

Page 16: [19] Deleted **Liu, Xueying** **8/4/23 9:57:00 PM**

Page 16: [20] Deleted **Liu, Xueying** **8/4/23 10:27:00 PM**

Page 16: [21] Deleted **Liu, Xueying** **8/4/23 10:39:00 PM**

Text S1. Identification of ozone episodes

Ozone exceedance days were identified according to surface measurements from the TCEQ CAMS (onshore) and the boats operating in Galveston Bay during the field campaign (offshore). Full description of the boat observations is given in Li et al. (2023). The criteria used in this study are (1) any onshore site from the CAMS network in Houston and Galveston or (2) offshore boat ozone observations that registered daily maximum 8-hour average (MDA8) ozone in exceedance of 70 ppbv, the current air quality standard for ozone. In total, six ozone episodes were identified during the whole campaign period over July-Oct 2021 (Table S1). Among these, three ozone episodes are in the extensive operating period of September 2021.

Table S1. Dates of ozone episodes and the associated MDA8 O_3 maximum.

Episodes	Highest MDA8 O_3 (ppbv)	
	Onshore (CAMS)	Offshore (Boat)
07/26 - 07/28	97	72
08/25	78	50
09/06 - 09/11	89	79
09/17 - 09/19	75	80
09/23 - 09/26	81	65
10/06 - 10/09	92	96

Formatted
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font: Times New Roman, 12 pt, Font color: Text 1
Deleted: The highest MDA8 ozone of each ozone episode is shown in Table S1.
Deleted: selected
Deleted: ¶
Formatted: Space After: 12 pt
Formatted: Font: 12 pt
Formatted: Font: 12 pt
Formatted Table
Formatted: Font: 12 pt
Formatted: Font: 12 pt
Formatted: Font: 12 pt
Formatted: Font: 12 pt
Formatted: Font: 12 pt
Formatted: Font: 12 pt
Formatted: Font: 12 pt
Formatted: Font: 12 pt
Formatted: Font: 12 pt
Formatted: Font: 12 pt
Formatted: Font: 12 pt
Formatted: Font: 12 pt
Formatted: Font: 12 pt
Formatted: Font: 12 pt
Formatted: Font: 12 pt
Formatted: Font: 12 pt
Formatted: Font: 12 pt
Formatted: Font: 12 pt
Formatted: Font: 12 pt

Text S2. Description of WRF nudging

We used observation nudging together with surface analysis nudging (also known as surface grid nudging) in WRF as the data assimilation method. In observation nudging, the modeled fields are nudged to match better with observations at individual locations with a radius of influence. The data used for observation nudging are ground-based hourly measurements of temperature, relative humidity as well as wind speed and direction from the Texas Commission on Environmental Quality (TCEQ) continuous ambient monitoring stations (CAMS). Based on site elevations, most nudging is performed within 500 m above sea level in eastern Texas, as shown in Figure S1. There are around 155, 98, and 49 observations ingested into WRF domains 1, 2 and 3, respectively. In analysis nudging, temperature, moisture and wind fields are nudged toward gridded analysis above the PBL (~1 km). The OBSGRID program was used for both observation nudging and surface analysis nudging. The program generated merged input files so that observation nudging and surface analysis nudging were conducted simultaneously when running the model. In addition to data assimilation, we adopted objective analysis in OBSGRID to provide better initial and boundary conditions, where first-guess meteorological fields are updated by incorporating observational data. The combined adoption of observation nudging, surface analysis nudging, and objective analysis in the [Nudged] simulation was to maximize the benefits of assimilating observations, as recommended by Chapter 7 of the WRF user guide.

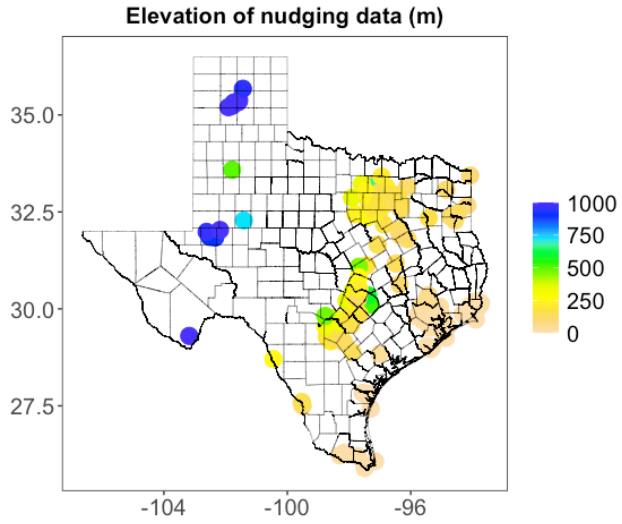
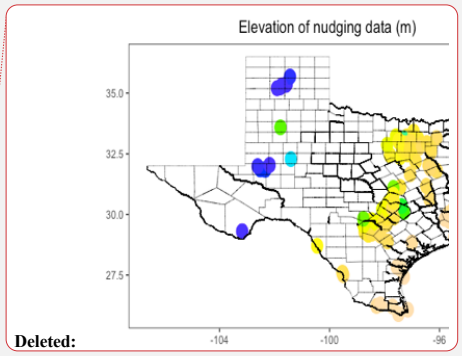


Figure S1. Elevation of the Texas Commission on Environmental Quality (TCEQ) continuous ambient monitoring stations (CAMS) used as observational data for WRF nudging methods.

Deleted:



Text S3. Evaluation of all model experiments

All WRF simulations are shown in Table S2. We evaluated the spatial and temporal variabilities of all simulations against the onshore TCEQ CAMS (Figure S2; Figure S3; Table S3) and the offshore boat measurements (Figure S4; Figure S5; Table S4). The WRF model generally reproduces observed temporal variability and spatial distribution in key meteorological parameters with a correlation coefficient higher than 0.5 in most cases. However, the model, regardless of configuration settings, shows persistent low biases in PBL heights, low biases in air temperatures, high biases in relative humidity, and high biases in wind speed. While different WRF configuration has its own advantage in reducing model biases, [HRRR], [Nudged] and [Reinit] configurations stand out as the three best simulations based on campaign-wide statistics. Considering that [Nudged] requires additional efforts to prepare observational datasets and [Reinit] needs to automate the model running process, [HRRR] is the easiest and the most effective option to reproduce meteorology for computationally expensive chemistry simulations and was thus selected to be presented in the main text.

Table S2. List of model experiments.

Simulations	BC Meteorology	PBL	Microphysics	Nudging	Reinitializing
[Base]	NCEP FNL	MYNN	2M	No	No
[WSM6]	NCEP FNL	MYNN	WSM6	No	No
[YSU]	NCEP FNL	YSU	2M	No	No
[ACM2]	NCEP FNL	ACM2	2M	No	No
[ERA5]	ECMWF ERA5	MYNN	2M	No	No
[HRRR]	HRRR	MYNN	2M	No	No
[Nudged]	NCEP FNL	MYNN	2M	Yes	No
[Reinit]	NCEP FNL	MYNN	2M	No	Yes

The mean of wind speed and direction is calculated using the vector notation approach, a commonly used method in wind evaluations, as described in Yu et al. (2023). This method treats wind as vectors with their u (eastward) and v (northward) wind components. First, the mean u and v wind components are found by averaging all u and v wind values over a given time period. Then, the resultant vector is determined by taking the square root of the sum of the squares of the mean u and mean v wind components. The magnitude of resultant vector represents the mean wind speed, and the angle of the resultant vector represents the mean wind direction.

The difference between observed and modeled wind direction was calculated as below.

$$\Delta = \begin{cases} M - O, & \text{when } |M - O| \le 180^\circ \\ (M - O)\left(1 - \dfrac{360}{|M - O|}\right), & \text{when } |M - O| > 180^\circ \end{cases}$$

where M is the model output, and O is the observation. The correlation between observed and modeled wind direction was determined by a circular correlation coefficient as below.

$$R = \frac{\sum_{i=1}^{N} \sin\,(M_i - \bar{M}) \sin\,(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{N} \sin^2(M_i - \bar{M})} \sqrt{\sum_{i=1}^{N} \sin^2(O_i - \bar{O})}}$$

Formatted: Text

Deleted: [HRRR] is the easiest and the most effective option to reproduce meteorology during the TRACER-AQ 2021 campaign and Considering both onshore and offshore evaluations, [HRRR] was selected

Deleted: as the best simulation to represent campaign-wide statistics

Deleted: 2

Deleted: ¶

Deleted: ¶

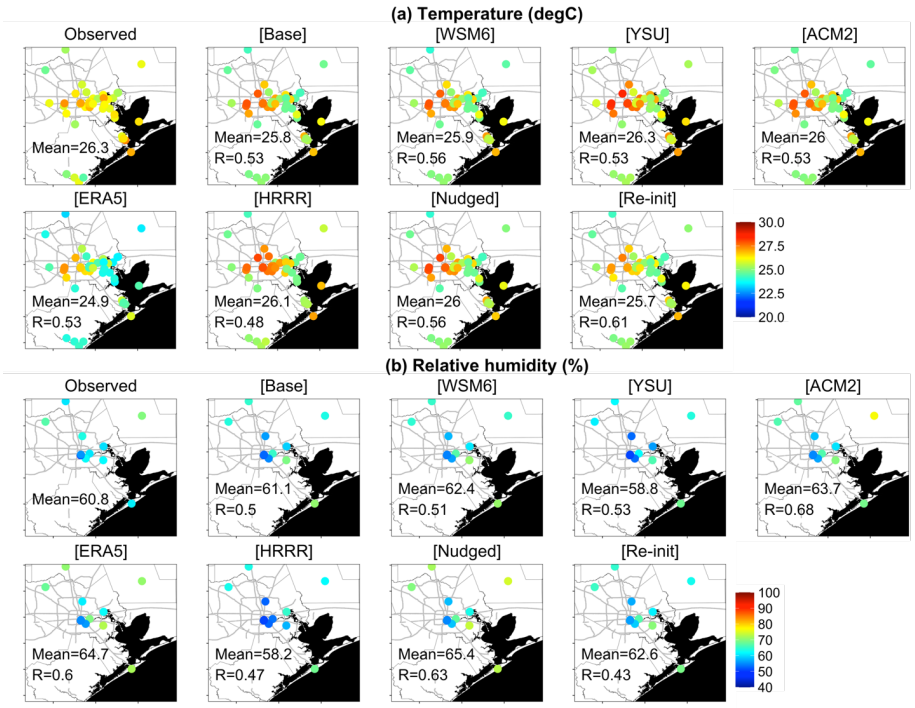
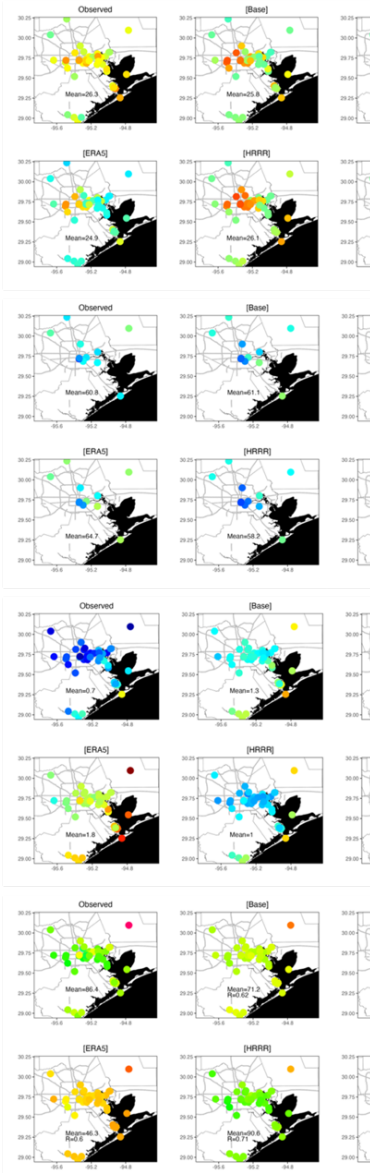


Figure S2. Spatial distribution of temporal averages of CAMS-observed and modeled mean meteorology during ozone episodes. The averages of wind speed and directions are calculated using the directional and vector mean approach. First, the mean of the u-wind (and v-wind) component is computed by averaging all u-wind (and v-wind) values at each station over the given period of time. Then, the mean wind speed and direction are calculated based on these mean u and v wind components at each station.



Deleted:

Formatted: Normal

Deleted: temporal

Formatted: Font color: Text 1

Formatted: Font: (Default) Times New Roman, Font color: Text 1, Pattern: Clear

Deleted: at the same station… Similarly, the mean of the v-wind component is determined separately by averaging all v-wind values at the same station. …henThen ... [1]

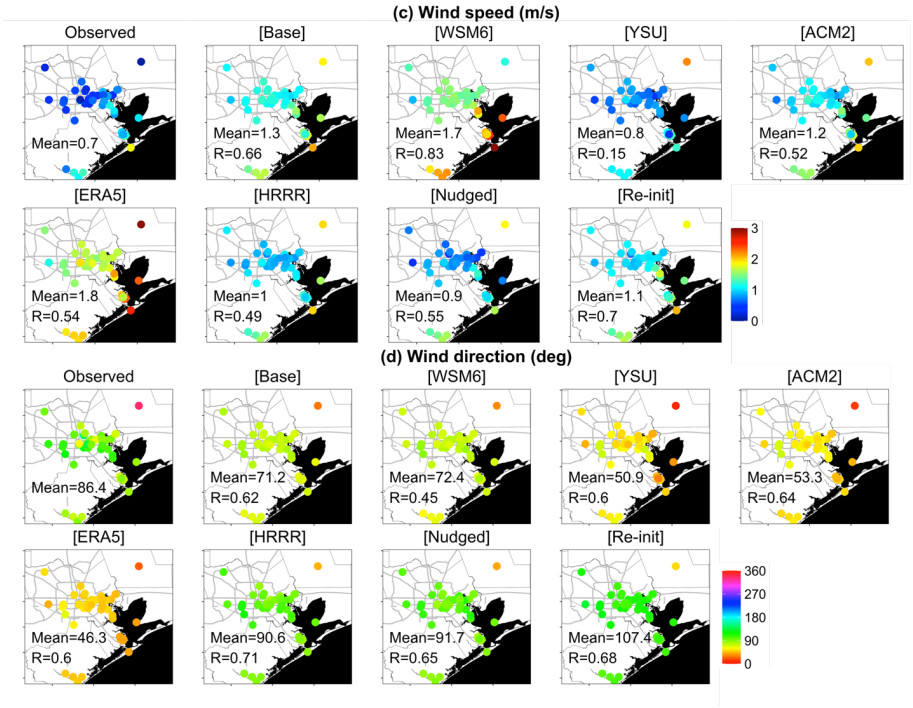
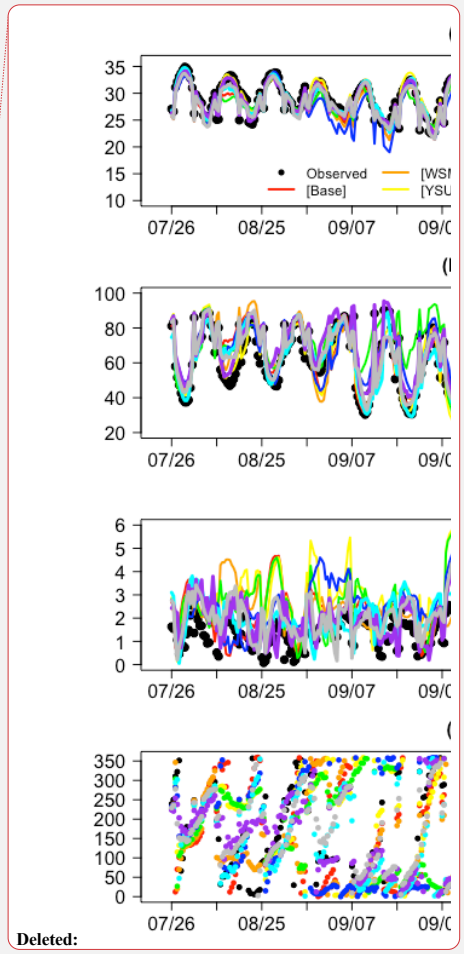


Figure S2 (continued). Spatial distribution of temporal averages of CAMS-observed and modeled mean meteorology during ozone episodes. The averages of wind speed and directions are calculated using the directional and vector mean approach. First, the mean of the u-wind (and v-wind) component is computed by averaging all u-wind (and v-wind) values at each station over the given period of time. Then, the mean wind speed and direction are calculated based on these mean u and v wind components at each station.

Deleted:



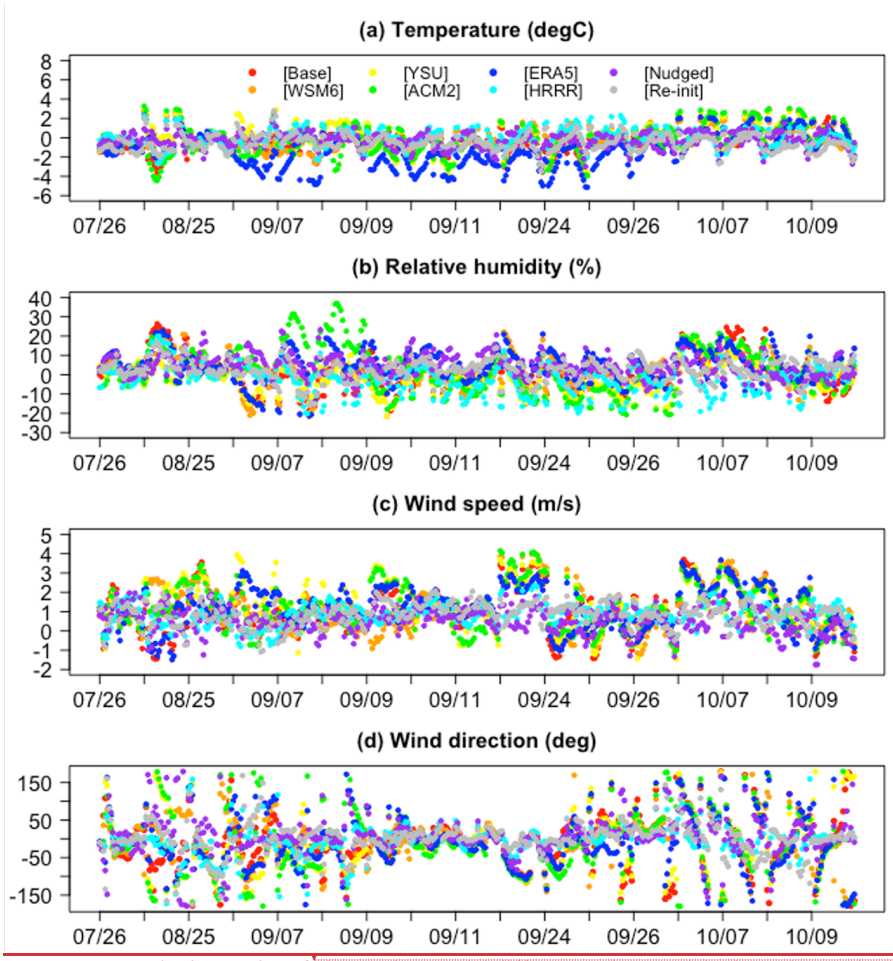


Figure S3. Hourly time series of observation-model differences (i.e., model minus observation) are shown for (a) air temperature, (b) relative humidity, (c) wind speed and (d) wind direction. The differences are spatial averages across all CAMS stations and the WRF model equivalents during ozone episodes. Refer to Text S3 for the calculations of spatial averages of wind speed and directions, as well as the differences between observed and modeled wind directions.

Deleted: Hourly time series of

Deleted: spatial averages of

Formatted: Font color: Text 1

Deleted: S

Deleted:

Deleted: between

Deleted: at

Deleted: observations

Deleted: simulations

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Deleted: are computed

Formatted: Font: (Default) Times New Roman, Font color: Text 1, Pattern: Clear

Formatted: Font: (Default) Times New Roman, Font color: Text 1, Pattern: Clear

Deleted: the mean of the u-wind component is computed by averaging all u-wind values from all stations. Similarly, the mean of the v-wind component is determined separately by averaging all v-wind values from all stations. T,mean seSpatial averaged of wind speed and directions are calculated by vector mean.

Deleted: ¶
¶

Deleted: ¶

Table S3. Performance metrics of spatiotemporal variability between CAMS-observed and WRF-modeled meteorology during ozone episodes. Hourly meteorology at all stations is used for the calculation of performance metrics below. All metrics have the same unit as meteorological variables, except that the correlation coefficient (R) and normal mean bias (NMB) are unitless. OBS and MOD represent the spatial and temporal averages of observations and model equivalents, respectively.

Variables	Simulation	OBS	MOD	R	NMB	MB	MAE	RMSE
Temperature (°C)	[Base]	26.18	25.82	0.88	-0.01	-0.36	1.69	2.15
	[WSM6]		25.84	0.89	-0.01	-0.35	1.57	1.99
	[YSU]		26.29	0.89	0.00	0.11	1.65	2.11
	[ACM2]		25.95	0.86	-0.01	-0.23	1.76	2.23
	[ERA5]		24.91	0.85	-0.05	-1.28	2.17	2.71
	[HRRR]		26.12	0.89	0.00	-0.06	1.59	2.05
	[Nudged]		25.92	0.92	-0.01	-0.26	1.43	1.84
	[Re-init]		25.69	0.92	-0.02	-0.49	1.41	1.77
Relative humidity (%)	[Base]	60.12	60.94	0.76	0.01	0.82	10.25	13.04
	[WSM6]		62.21	0.78	0.03	2.09	9.85	12.28
	[YSU]		58.45	0.80	-0.03	-1.68	9.54	12.31
	[ACM2]		62.73	0.71	0.04	2.60	11.40	14.71
	[ERA5]		64.21	0.77	0.07	4.08	10.55	12.76
	[HRRR]		57.82	0.79	-0.04	-2.30	9.13	12.13
	[Nudged]		64.63	0.82	0.08	4.51	9.54	12.05
	[Re-init]		62.57	0.84	0.04	2.45	8.37	10.66
Wind speed (m/s)	[Base]	0.67	1.29	0.35	0.59	1.01	1.40	1.70
	[WSM6]		1.67	0.37	0.61	1.04	1.39	1.72
	[YSU]		0.80	0.39	0.75	1.29	1.55	1.87
	[ACM2]		1.16	0.38	0.66	1.12	1.44	1.77
	[ERA5]		1.76	0.43	0.64	1.09	1.38	1.66
	[HRRR]		1.00	0.54	0.49	0.83	1.12	1.36
	[Nudged]		0.89	0.55	0.30	0.51	0.96	1.20
	[Re-init]		1.14	0.61	0.48	0.82	1.07	1.31
Wind direction (deg)	[Base]	87.76	72.32	0.43	-0.05	-7.67	56.5	73.36
	[WSM6]		72.56	0.38	-0.04	-5.51	56.41	72.93
	[YSU]		53.26	0.41	-0.08	-12.14	60.30	77.29
	[ACM2]		54.87	0.37	-0.07	-10.64	64.15	81.29
	[ERA5]		47.32	0.43	-0.07	-10.92	58.05	74.83
	[HRRR]		92.51	0.61	-0.02	-3.43	40.16	57.55
	[Nudged]		93.29	0.48	0.02	3.00	46.05	64.70
	[Re-init]		109.03	0.47	0.00	-0.32	39.99	57.67

Deleted: for

Deleted: five

Formatted: Font: Times New Roman, 12 pt, Font color: Text 1

Formatted: Font: Times New Roman, 12 pt, Font color: Text 1

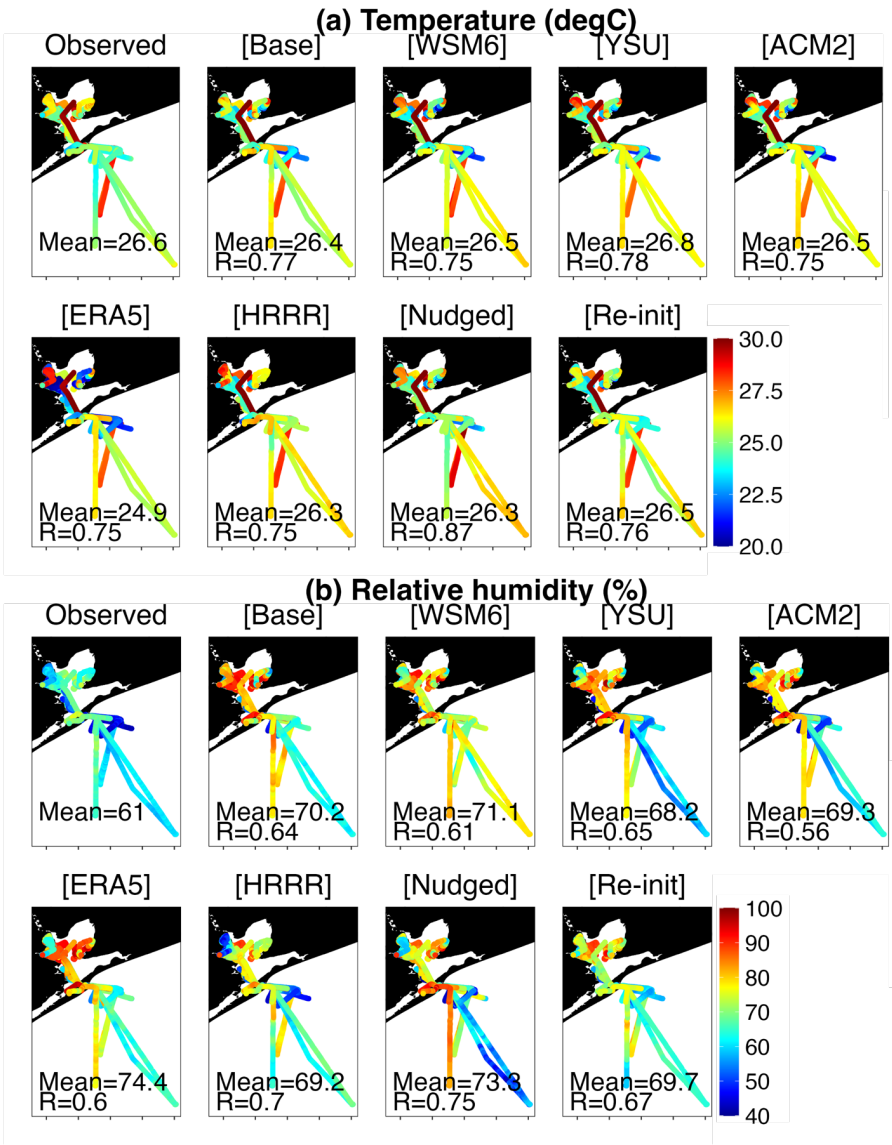
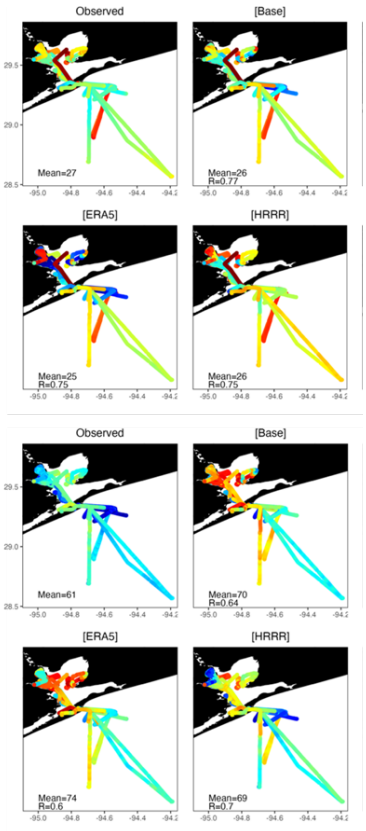


Figure S4. Spatial distribution of boat-observed and modeled meteorology during ozone episodes.

Deleted:



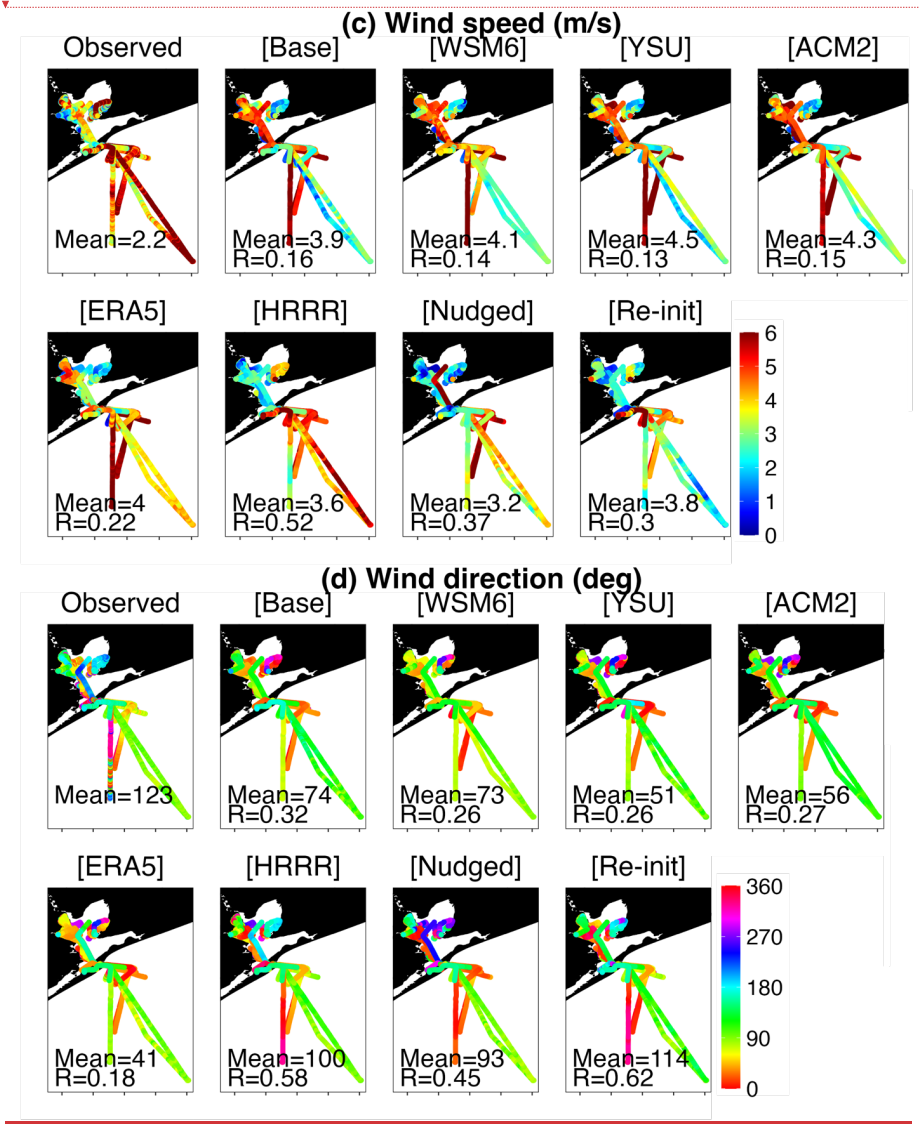
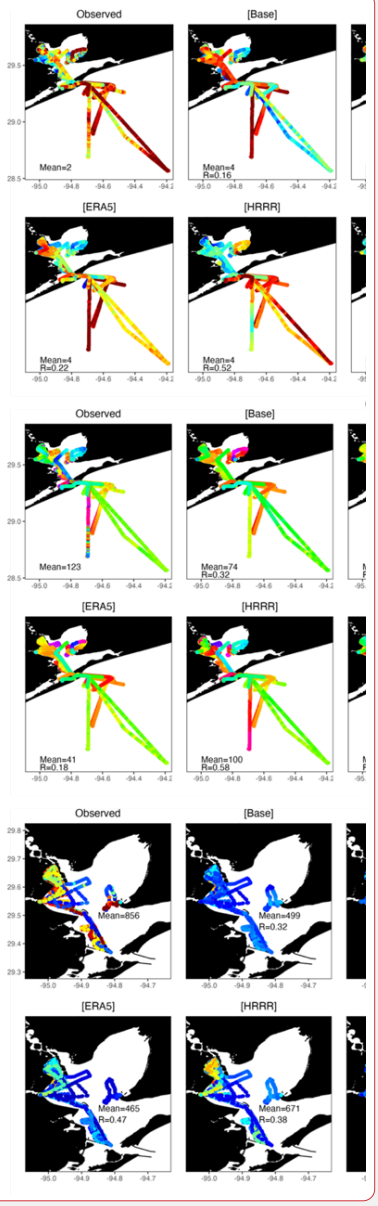


Figure S4 (continued). Spatial distribution of boat-observed and modeled meteorology during ozone episodes.

Deleted:



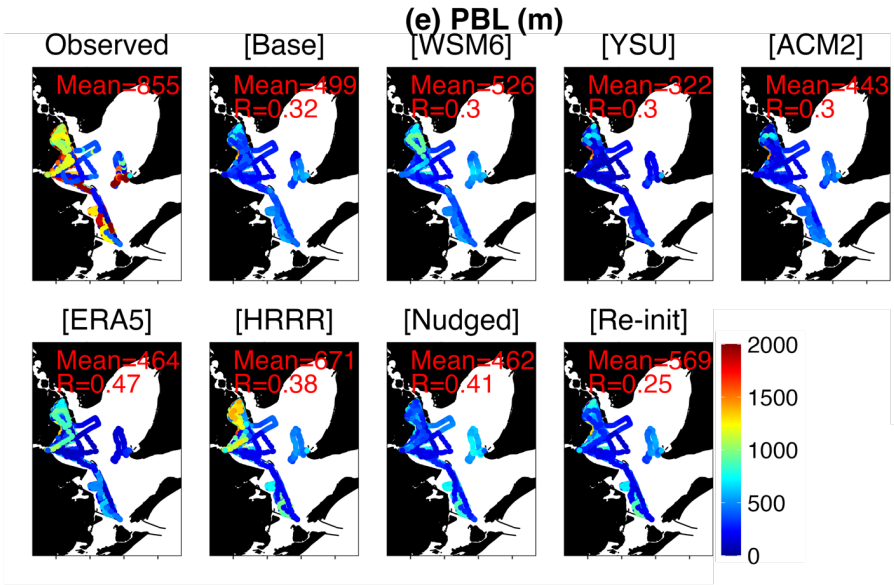
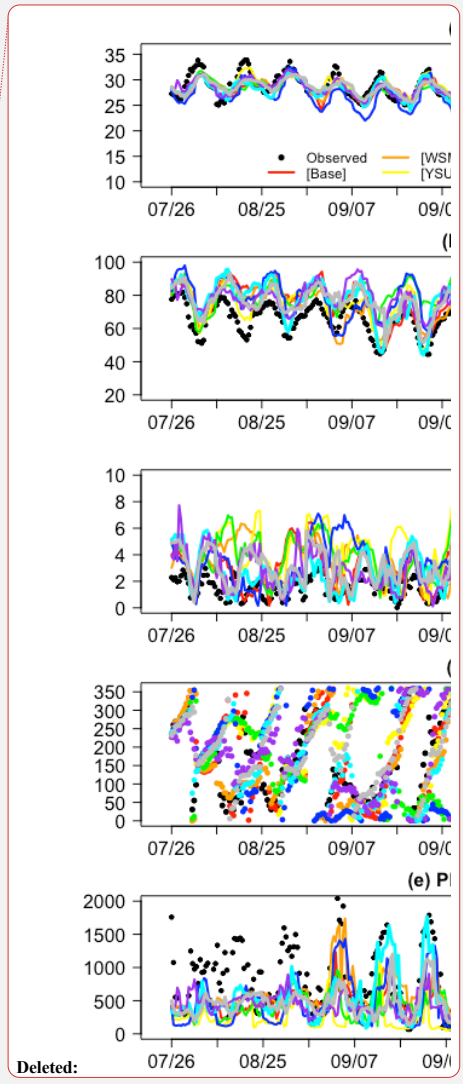


Figure S4 (continued). Spatial distribution of boat-observed and modeled meteorology during ozone episodes.



Deleted:

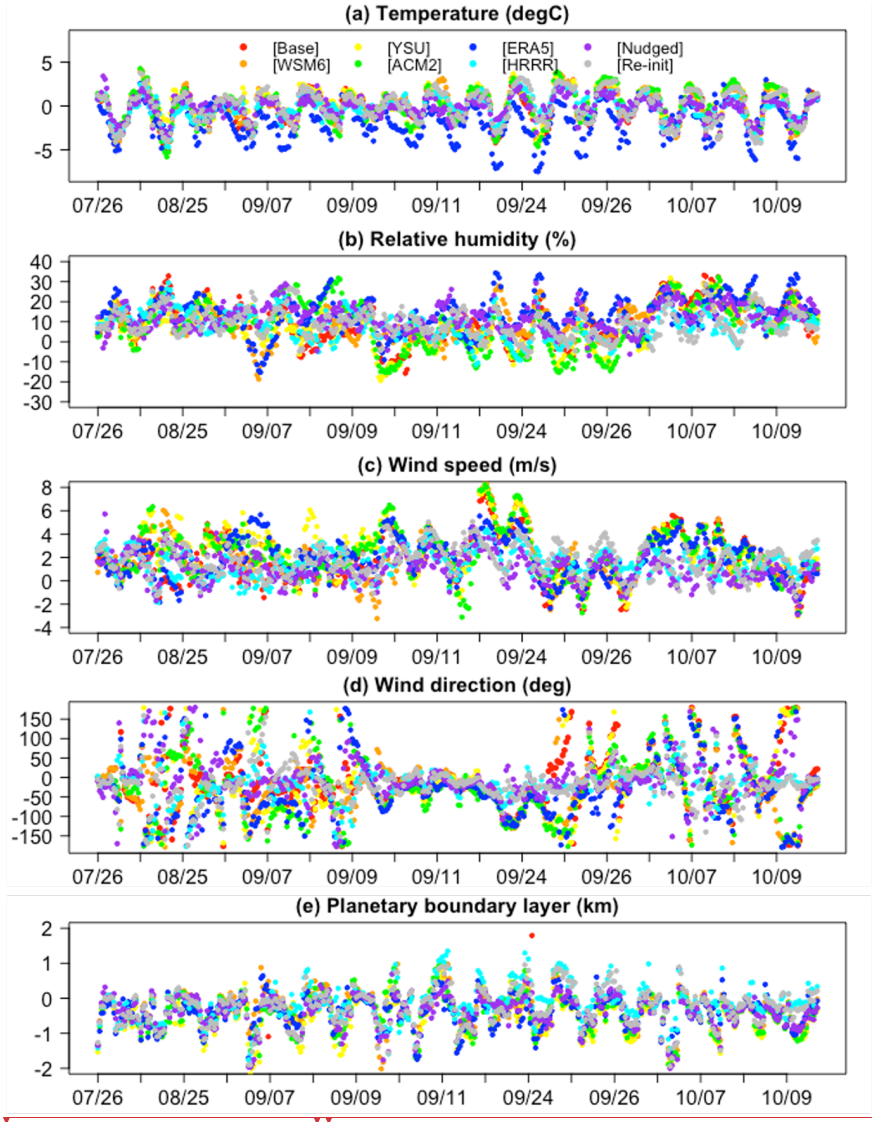


Figure S5. Hourly time series of observation-model differences (i.e., model minus observation) are shown for (a) air temperature, (b) relative humidity, (c) wind speed, (d) wind direction and (e) boundary layer height during ozone episodes. Refer to Text S3 for the calculations of averages of wind speed and directions, as well as the differences between observed and modeled wind directions.

Deleted: ¶

Deleted: (a) air temperature, (b) relative humidity, (c) wind speed, (d) wind direction and (e) boundary layer height between boat observations and WRF model simulations during ozone episodes.¶

Deleted: O

Deleted: spatial

Table S4. Performance metrics of spatiotemporal variability between boat-observed and WRF-modeled meteorology during ozone episodes. 1-minute meteorology is used for the calculation of performance metrics below. All metrics have the same unit as meteorological variables, except that the correlation coefficient (R) and normal mean bias (NMB) are unitless. OBS and MOD represent the spatial and temporal averages of observations and model equivalents, respectively.

Variables	Simulation	OBS	MOD	R	NMB	MB	MAE	RMSE
Temperature (°C)	[Base]	26.55	26.45	0.77	0.00	-0.11	1.71	2.14
	[WSM6]		26.50	0.75	0.00	-0.05	1.77	2.20
	[YSU]		26.78	0.78	0.01	0.22	1.71	2.10
	[ACM2]		26.51	0.75	0.00	-0.04	1.78	2.21
	[ERA5]		24.85	0.75	-0.06	-1.70	2.21	3.00
	[HRRR]		26.30	0.75	-0.01	-0.25	1.89	2.29
	[Nudged]		26.30	0.87	-0.01	-0.25	1.26	1.65
	[Re-init]		26.53	0.76	0.00	-0.02	1.71	2.15
Relative humidity (%)	[Base]	60.96	70.24	0.64	0.15	9.28	11.95	14.59
	[WSM6]		71.09	0.61	0.17	10.14	11.76	14.38
	[YSU]		68.20	0.65	0.12	7.24	10.96	13.29
	[ACM2]		69.35	0.56	0.14	8.40	12.75	15.33
	[ERA5]		74.38	0.60	0.22	13.42	14.66	17.23
	[HRRR]		69.20	0.70	0.14	8.24	10.38	12.68
	[Nudged]		73.35	0.75	0.20	12.39	12.87	14.92
	[Re-init]		69.68	0.67	0.14	8.72	10.25	12.42
Wind speed (m/s)	[Base]	0.73	2.47	0.16	0.74	1.67	2.20	2.78
	[WSM6]		2.62	0.14	0.82	1.85	2.33	2.92
	[YSU]		2.17	0.13	0.99	2.22	2.63	3.19
	[ACM2]		1.99	0.15	0.92	2.07	2.49	3.09
	[ERA5]		1.89	0.22	0.78	1.74	2.21	2.72
	[HRRR]		1.68	0.52	0.59	1.32	1.69	2.05
	[Nudged]		1.75	0.37	0.41	0.92	1.57	1.96
	[Re-init]		2.02	0.30	0.69	1.55	2.00	2.41
Wind direction (deg)	[Base]	144.15	118.78	0.32	-0.08	-11.45	57.74	75.38
	[WSM6]		113.5	0.26	-0.13	-19.10	60.40	77.29
	[YSU]		135.77	0.26	-0.11	-16.44	63.52	81.13
	[ACM2]		125.25	0.27	-0.11	-17.20	68.93	85.92
	[ERA5]		96.69	0.18	-0.17	-25.20	69.00	85.30
	[HRRR]		137.93	0.58	-0.08	-12.53	41.54	58.16
	[Nudged]		146.95	0.45	-0.05	-7.68	47.87	65.51
	[Re-init]		146.96	0.62	-0.10	-14.98	42.98	59.66

Deleted: ¶
¶
¶

Deleted: for

Deleted: five

Boundary layer height (m)	[Base]	855.58	499.27	0.32	-0.42	-356.30	529.63	699.67
	[WSM6]		526.69	0.30	-0.38	-328.88	526.38	691.82
	[YSU]		322.22	0.30	-0.62	-533.36	612.29	817.16
	[ACM2]		443.60	0.30	-0.48	-411.97	562.12	747.06
	[ERA5]		464.75	0.47	-0.46	-390.83	507.51	680.30
	[HRRR]		671.27	0.38	-0.22	-184.31	461.30	637.68
	[Nudged]		462.09	0.41	-0.46	-393.48	516.18	696.37
	[Re-init]		569.57	0.25	-0.33	-286.00	518.21	689.22

References:

Li, W., Wang, Y., Liu, X., Soleimanian, E., Griggs, T., Flynn, J., and Walter, P.: Understanding offshore high-ozone events during TRACER-AQ 2021 in Houston: Insights from WRF-CAMx photochemical modeling, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2023-1117, 2023.

Yu, E., Bai, R., Chen, X., and Shao, L.: Impact of physical parameterizations on wind simulation with WRF V3.9.1.1 under stable conditions at planetary boundary layer gray-zone resolution: a case study over the coastal regions of North China, Geosci. Model Dev., 15, 8111–8134, https://doi.org/10.5194/gmd-15-8111-2022, 2022.

Deleted: Text S4. Vertical ozone distribution at University of Houston Similar to the TROPOZ lidar in the main text, the Langley Mobile Ozone Lidar (LMOL) is part of the ground-based Tropospheric Ozone Lidar Network (TOLNet, https://www-air.larc.nasa.gov/missions/TOLNet/). The LMOL has been used to provide continuous, high-resolution profile measurements of ozone during various campaigns. Figure S6 shows the vertical ozone distribution measured by the LMOL at University of Houston during the two ozone episodes. ¶

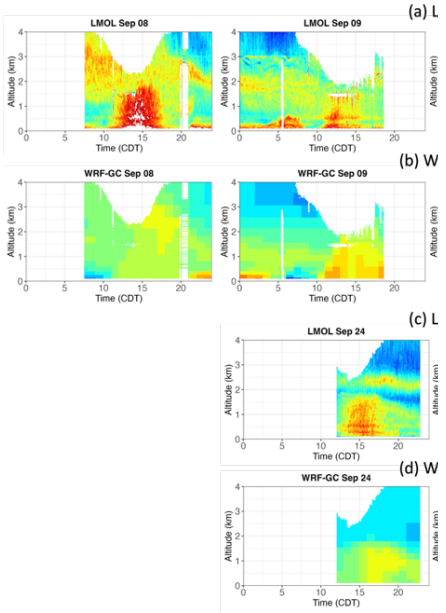


Figure S6. Time series of the vertical ozone profile from (a, c) the LMOL ozone lidar and (b, d) the WRF-GC [HRRR] simulation at the University of Houston site during September 8–11 and September 23–26 of 2021. ¶

Formatted: Font: (Default) Times New Roman, 12 pt, Font color: Auto, Do not check spelling or grammar, Pattern: Clear

Page 4: [1] Deleted **Liu, Xueying** **8/1/23 7:46:00 PM**

Page 4: [1] Deleted **Liu, Xueying** **8/1/23 7:46:00 PM**

Page 4: [1] Deleted **Liu, Xueying** **8/1/23 7:46:00 PM**