

AQRP Project 22-008

Modeling analysis of the Tracking Aerosol Convection Experiment Air Quality (TRACER-AQ) and over-water measurements to improve prediction of on-land and offshore ozone

Final Report

August 31, 2023

Principal Investigators

Yuxuan Wang, University of Houston

Contributors

James Flynn, University of Houston

Paul Walter, St. Edwards University

Xueying Liu, University of Houston

Shailaja Wasti, University of Houston

Wei Li, University of Houston

Claudia Bernier, University of Houston

Evelyn Martinez Sabari, University of Houston

Yongcheol Jeong, University of Houston

Geoffrey Roberts, University of Houston

Acknowledgement

The preparation of this report was funded by a grant from the Texas Air Quality Research Program (AQRP) at The University of Texas at Austin through the Texas Emission Reduction Program (TERP) and the Texas Commission on Environmental Quality (TCEQ). The findings, opinions and conclusions are the work of the author(s) and do not necessarily represent findings, opinions, or conclusions of the AQRP or the TCEQ.

Executive Summary

This project used observations from the 2021 Galveston Offshore Ozone Observations (GO3) and Tracking Aerosol Convection Experiment/Air Quality (TRACER-AQ) field campaigns to evaluate and improve meteorological and photochemical models and identify the sources and influences of high ozone over adjacent waters near Houston.

First, the project evaluated ten WRF configurations of different meteorological inputs, physics options, and data assimilation options against both onshore and offshore observations. 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], [Nudged2] and [Reinit] configurations stand out as the three best simulations based on campaign-wide statistics. Considering that [Nudged2] 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 during the TRACER-AQ 2021 campaign.

Second, we compare inter-model differences in ozone and precursor gases among three photochemical models, i.e., CAMx, WRF-GC, and WRF-Chem, using observations from multi-platforms (e.g., surface and remote sensing measurements). In comparison with surface measurements, CAMx best captures ozone concentrations, while WRF-GC and WRF-Chem show significantly higher positive and negative biases, respectively. One of the major reasons for such differences originates from the ozone episode of September 17-19 caused by the passing of Hurricane Nicolas. Models have different extents of difficulties in representing the meteorology and chemistry after the hurricane. The second major reason is associated with clean days. All three models show substantially larger biases and less correlation on clean days than on ozone episode days. In comparison with column concentrations measured by remote sensing instruments, all models underestimate ozone at 0-1 km aloft and NO_x and HCHO at 0-8 km aloft, but WRF-GC shows the smallest bias compared with the other two models. Ozone sensitivity derived from column concentrations of NO_x and HCHO is best captured by WRF-GC (diurnal variations) and CAMx (spatial patterns). WRF-Chem has the best performance in simulating long-range transported ozone plumes above the boundary layer as observed by the two ozone lidars, which may be attributed to the fact that only WRF-Chem implemented 2021 fire emissions.

Last, through zeroing-out local anthropogenic emissions in two models (WRF-GC and CAMx), we found local anthropogenic emissions contribute to less than 20% of surface ozone in Houston during two high-ozone periods in September 2021. Over-water ozone in both models is more responsive to reductions in land emissions than to over-water emissions, although offshore ozone does not decrease much (less than 0.5%) in response to 10% emission reductions of land-based emissions because of the large background in the presence or it even increases because of the offset effect of increasing the lifetime of primary pollutants leading to more in situ ozone production over the waters. Indeed, CAMx process analysis revealed that in situ production of ozone over the waters is the leading process responsible for net ozone increase at offshore locations during episode days relative to clean days. This increase is likely resulted from enhanced transport of ozone precursors originating from outside Houston to offshore locations. The contributions from direct advection of ozone can be high at offshore locations in some

specific cases, but its mean contributions over multiple days are averaged out and present a small sink for offshore ozone. These modeling experiments suggest (1) the resiliency of high ozone over water to small changes in local emissions, making it difficult to control, and (2) a need to better model natural emissions as well as regional transport dynamics for improving offshore ozone simulation.

Table of Contents

1. Introduction	5
2. Model and Data	6
2.1 Meteorological Model Description.....	6
2.2 Photochemical Model Description	7
2.2.1 CAMx.....	7
2.2.2 WRF-GC.....	8
2.2.3 WRF-Chem	8
2.3 Observational data	9
2.3.1 Meteorological Observations	9
2.3.2 Chemical Observations.....	9
2.3.3 Identification of Ozone Exceedance Days	10
2.4 Performance Metrics.....	10
3. Meteorological Model Evaluation and Improvement	11
3.1. Experimental Setup.....	11
3.2. PBL Evaluation	12
3.3. Evaluation of Other Meteorological Variables	16
3.3.1 CAMS sites.....	16
3.3.2 Boat observations	20
3.3.3 La Porte surface measurements.....	24
3.4. Summary.....	27
4. Photochemical Model Evaluation and Model Intercomparison	27
4.1 Evaluation of Photochemical Models with TRACER-AQ Observations	27
4.1.1. CAMx Evaluation.....	28
4.1.2. WRF-GC Evaluation	31
4.1.3. WRF-Chem Evaluation	34
4.2 Model Intercomparison.....	37
4.2.1. Mean state of ozone and precursors.....	37
4.2.2. O ₃ diurnal cycle.....	41
4.2.3. Column distribution of O ₃ and precursors	43
4.2.4. Vertical distribution of ozone	48
4.2.5. VOC and NO _x sensitivity.....	50
4.3. Summary.....	51
5. Investigation of Elevated Offshore Ozone's Sources	52
5.1 Regional Background Ozone	53
5.2 Emission Perturbation Experiments.....	61
5.3 CAMx Process Analysis.....	70
5.4 Summary.....	72
6. Summary and Future Directions	73
References	75

1. Introduction

In the urban coastal environment of the Houston-Galveston-Brazoria (HGB), ozone (O_3) pollution can be driven by diverse mechanisms that influence ozone development. Due in part to different regional background O_3 , high O_3 events in the HGB are most associated with continental outflow, while the lowest O_3 levels are from onshore winds (Berlin et al., 2013; Wang et al., 2016; Li et al., 2020). When a sea breeze recirculation occurs, ozone concentrations initially over the ocean can be recirculated throughout the coast and further inland leading to high ozone levels (Banta et al., 2005; Rappenglück et al., 2008; Caicedo et al., 2019). While photochemical models occasionally predict elevated ozone concentrations offshore, there is limited confidence in model predictions because the offshore environment of the HGB has been historically under-monitored until recently.

The Tracking Aerosol Convection Experiment-Air Quality (TRACER-AQ) study, including the Galveston Offshore Ozone Observations (GO3) field campaign, was carried out during July – October 2021. TRACER-AQ has a wide range of measurement platforms covering both offshore and onshore locations (**Figure 1**), including ozonesondes, boats, ground-based ozone lidars, mobile labs, stationary sites, and aircraft remote sensing. During the TRACER-AQ period, there were six multi-day ozone episodes, resulting in over 20 days during which at least one on-land monitor or ship-based measurement with Maximum Daily 8-hour Average (MDA8) ozone concentrations exceeded the current National Ambient Air Quality Standard (NAAQS) of 70 ppbv. Therefore, TRACER-AQ provided unprecedentedly rich observations of ozone air pollution that are needed to validate current air quality models.

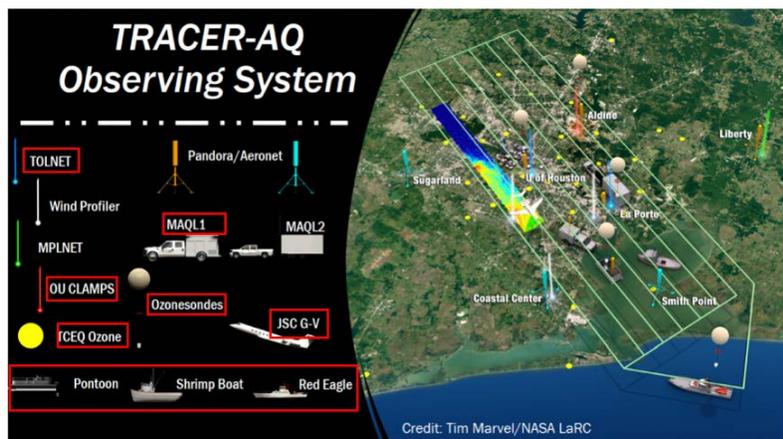


Figure 1. Schematic summary of the observing systems in the TRACER-AQ campaign. Red squares show the observations to be used in the project.

This AQRP project aims to use the TRACER-AQ observations from both onshore and offshore monitors to evaluate and improve meteorological and photochemical models, and identify the sources and influences of high ozone over waters. First, the project evaluates the meteorological Weather Research and Forecasting (WRF) model and identifies model settings that best represent the TRACER-AQ observations in **Section 3**. Second, the project compares three different photochemical models (CAMx, WRF-GC, and WRF-Chem) that are driven by the same WRF outputs in **Section 4**. The model intercomparison characterizes the strengths and weaknesses of the regulatory model, CAMx, in the context of other air quality models. Finally, the project uses the models to identify the sources and influences of high ozone over Galveston Bay and the Gulf of Mexico in **Section 5**.

2. Model and Data

2.1 Meteorological Model Description

The WRF model (v3.9.1.1) was used in this 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 2**. The corresponding horizontal resolutions for domains 1–3 are 12 kilometers (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 up to 10 hectopascals (hPa).

Global meteorological fields were used to provide boundary and initial conditions for the mesoscale WRF model. This study employed three kinds of global meteorological data as alternatives. 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 National Oceanic and Atmospheric Administration (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.

WRF has different schemes or options to represent physics and dynamics processes. Three planetary boundary layer (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 were 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). Detailed formulation of PBL height diagnosis is shown in Sect. 2.2.3. 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 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.

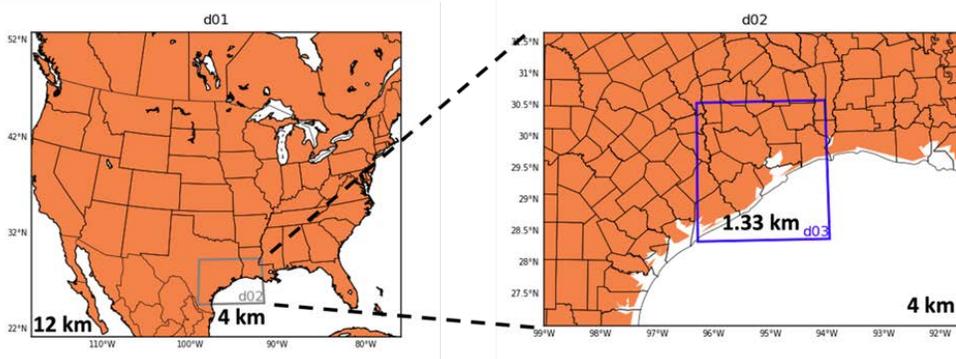


Figure 2. Model domains and horizontal resolutions.

2.2 Photochemical Model Description

Table 1 summarizes model settings for three photochemical models. Detailed settings are described in Section 2.2.1 for CAMx, Section 2.2.2 for WRF-GC, and Section 2.2.3 for WRF-Chem.

Table 1. Summary of photochemical model settings. The year of the emissions is listed in parentheses. Online calculation means that emissions are calculated online using model-generated meteorology. N/A represents not available.

	CAMx	WRF-GC	WRF-Chem
Domain	d01 (12 km), d02 (4 km), d03 (1 km)	d02 (4 km), d03 (1 km)	d02 (4 km), d03 (1 km)
Simulation period	Jul 20–30, Aug 20 – Oct 13 of 2021	Sep 1– Oct 1 of 2021	Sep 1–30 of 2021
Meteorological IC/BC	HRRR	HRRR	HRRR
Chemical IC/BC	GEOS-Chem (2021)	GEOS-Chem (2021)	WACCM (2021)
Gas-phase chemistry	CB6r5	Full O _x -NO _x -VOC-halogen-aerosol chemistry in GEOS-Chem	MOZART
Aerosol chemistry	N/A	Full O _x -NO _x -VOC-halogen-aerosol chemistry in GEOS-Chem	MOSAIC
Anthropogenic emission	TCEQ SIP (2019)	TCEQ SIP (2019)	TCEQ SIP (2019)
Fire emission	FINN v1.5 (2019)	GFED (2019)	FINN v2.5 (2021)
Soil NO _x	BEIS v5 (2019)	Hudman et al. (online calculation)	MEGAN (online calculation)
Lightning NO _x	N/A	Murray et al. (online calculation)	N/A
Biogenic	BEIS v5 (2019)	MEGAN (online calculation)	MEGAN (online calculation)

2.2.1 CAMx

The project used the CAMx model v7.10. The three CAMx domains aligned with the WRF grids but had smaller spatial coverage. The corresponding horizontal resolutions and grid numbers for domains 1–3 are 12 km × 12 km, 4 km × 4 km, and 1.33 km × 1.33 km, respectively. All domains have identical vertical resolutions with 30 vertical levels from the surface to ~100 hPa. The IC/BC inputs for the outmost domain are from the GEOS-Chem (v12.2.1) global simulation with NEI 2011 nitrogen oxides (NO_x) emissions scaled down to 2021. The Carbon Bond version 6 revision 5 (CB6r5) was used for gas-phase chemistry, including the inorganic iodine depletion of O₃ over oceanic water (Burkholder et al., 2019). The first-order eddy viscosity (K-theory) diffusion scheme was selected for vertical mixing within the

PBL, in which the vertical diffusion coefficients (K_v) were supplied from WRF outputs. Dry deposition is based on the Wesely scheme (Wesely, 1989).

Emission files with 12 km and 4 km spatial resolutions from the 2019 SIP modeling platform provided by TCEQ are used in the simulation. Since our domains are smaller than those in the SIP modeling, the original emission files were cropped to match the grid boundaries for CAMx to read properly. In addition, we redistributed the on-road emissions from 4 km to 1.33 km over the Houston area. The 4 km emission fluxes were first disaggregated evenly to the 1.33 km grids and then collected onto major roads using a 1 km rasterized road shapefile to produce on-major-road 1.33 km emissions. Some 1.33 km grid points off the major roads had missing values, which were filled using a smoothing method that averaged eight nearby grid points. The scaling factors for on- and off-major-road emissions were kept in order to maintain the on-road emission budget consistent before and after the spatial redistribution. Finally, total emissions were calculated by adding the 1.33 km on- and off-major-road emissions. The emissions for other sectors were also similarly interpolated to 1.33 km without separating into non- or off-major-road temporary emissions. The redistributed emissions were tested to perform better in capturing the on-road distributions than using the Flexi-nesting function in CAMx, which can regrid the emissions on the fly.

2.2.2 WRF-GC

The project used the WRF-GC v2.0, 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). We set up two domains at horizontal resolutions of 4 km and 1.33 km to cover the same d02 (for Southeast Texas) and d03 (for the Houston-Galveston region) domains of CAMx. To maintain consistency with the CAMx domain abbreviations, we refer to the two WRF-GC domains as d02 and d03 for the following of the report. 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 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.

WRF-GC obtained meteorological boundary and initial conditions from the High-Resolution Rapid Refresh (HRRR) from NOAA Amazon Web Service (<https://registry.opendata.aws/noaa-hrrr-pds>), which is at hourly temporal and 3 km horizontal resolutions.

WRF-GC used the most updated full O_x - NO_x -VOC-halogen-aerosol chemistry from GEOS-Chem. For anthropogenic emissions, we used the same TCEQ SIP emission inventory used in CAMx for southeastern Texas at 4 km horizontal resolution (1.33 km for on-road emissions) and the 2013 National Emission Inventory (NEI) for the rest of the US at 0.1° horizontal resolution. Fire 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.

2.2.3 WRF-Chem

The project used the WRF-Chem model v4.2.2. We set up two domains at horizontal resolutions of 4 km and 1.33 km to cover the same d02 (for Southeast Texas) and d03 (for the Houston-Galveston region) domains of CAMx. To maintain consistency with the CAMx domain

abbreviations, we refer to the two WRF-Chem domains as d02 and d03 for the following of the report. All domains have identical vertical resolutions with 50 vertical levels from the surface to ~10 hPa. IC/BC for meteorology are provided by the High-Resolution Rapid Refresh (HRRR) from NOAA Amazon Web Service (<https://registry.opendata.aws/noaa-hrrr-pds>), which is at hourly temporal and 3 km horizontal resolutions. IC/BC for the gas-phase species and aerosols are provided by the Whole Atmosphere Community Climate Model (WACCM; Gettelman et al. (2019)) of year 2021. The WACCM model output datasets are available on a horizontal grid resolution of $1^\circ \times 1^\circ$ and interpolated in space every six hours and are produced daily by the National Center for Atmospheric Research (NCAR). The gas-phase chemical mechanism used is the Model of Ozone and Related Chemical Tracers (MOZART; Emmons et al., 2020). MOZART is coupled with the aerosol scheme, the Model for Simulating Aerosol Interactions and Chemistry (MOSAIC; Zaveri et al., 2008), using four discrete sectional size bins (MOZART-MOSAIC4b). Fire emissions were provided by the Fire INventory from NCAR version 2.5 (FINNv2.5; (Wiedinmyer and Emmons, 2022)). Dry deposition is based on the Wesely scheme (Wesely, 1989).

For anthropogenic emissions, we used the same TCEQ SIP emission inventory used in CAMx and WRF-GC. The SIP modeling emissions were cropped using the same method described in Section 1 for CAMx. The regridded emissions at 4 km and 1.3 km domains were fitted to the modeling domain used for the WRF-Chem simulation, for which the emissions were interpolated to the new meshes using a conservative flow method ([GitHub - JoseAgustin/interpola: Interpola emissions to a new mesh based in wrfchemin and wrfinput files](#)). The rest of the emission species needed to complete MOZART-MOSAIC4b chemical mechanism mapping was obtained from EPA NEI 2017, using the anthro_emis tool publicly available by NCAR (<https://www2.acom.ucar.edu/wrf-chem/wrf-chem-tools-community>).

2.3 Observational data

2.3.1 Meteorological Observations

Meteorological variables including air temperature, relative humidity, wind, and PBL measured during the TRACER-AQ 2021 campaign are used to validate the modeled meteorology. The PBL is derived from aerosol backscatter signals from (1) the High Spectral Resolution Lidar-2 (HSRL2) flying over urban Houston and the Galveston Bay, and (2) two ceilometers respectively at a ground-based La Porte site and on boats operating in the Galveston Bay. Surface meteorological measurements are from (1) the TCEQ continuous ambient monitoring stations (CAMS) across the greater Houston area, (2) boats operating in Galveston Bay, and (3) the La Porte site.

2.3.2 Chemical Observations

Synchronized observations from onshore monitoring stations, mobile measurements, boats, and ground-based and airborne remote sensing during TRACER-AQ were used to validate ozone and precursor simulations from the three photochemical models. All data are available from the TRACER-AQ website (<https://www-air.larc.nasa.gov/missions/tracer-aq/>). Surface ozone measurements are from the TCEQ continuous ambient monitoring stations (CAMS), the mobile lab operating in urban Houston, and the boats operating between Galveston Bay and the Gulf. Ozone vertical profiles are from (1) Tropospheric Ozone (TROPOZ) lidar located at La Porte, (2) the High Spectral Resolution Lidar-2 (HSRL-2) onboard the NASA aircraft, and (3) ozonesonde. The tropospheric NO₂ and formaldehyde (HCHO) column measurements are from the GEOCAPE Airborne Simulator (GCAS) onboard the NASA aircraft, which measures three

times a day between 8:00-11:00 (raster 1), 11:00-14:00 (raster 2), and 14:00-17:00 (raster 3) over the Houston area.

2.3.3 Identification of Ozone Exceedance Days

Both onshore (CAM5) and offshore (boat) ozone observations were used to select ozone exceedance days. Ozone exceedance days were identified when (1) any onshore site from the CAM5 network in Houston and Galveston or (2) offshore boat ozone observations registered daily maximum 8-hour average (MDA8) ozone in exceedance of 70 ppbv, the current air quality standard for ozone. Six high ozone episodes were identified based on the above criteria; they are July 26-28, August 25, September 6-11, September 17-19, September 23-26, and October 6-9 in 2021. These episode periods were the focus of model evaluation and intercomparison. September 17-19 episode is discarded for further analysis due to the influence of Tropical Cyclone Nicolas.

2.4 Performance Metrics

Table 2 shows the performance metrics used in this study. As wind direction is circular, the difference between observed and modeled angles of wind was calculated as below.

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

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

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

Table 2. Model performance metrics used in this study. M is the model output, O is the observation, N is the number of samples, and $\bar{M} = 1/N \sum_{i=1}^N M_i$, $\bar{O} = 1/N \sum_{i=1}^N O_i$.

Performance Metrics	Formulas
Mean Bias (MB)	$MB = 1/N \sum_{i=1}^N (M_i - O_i)$
Mean Absolute Error (MAE)	$MAE = 1/N \sum_{i=1}^N M_i - O_i $
Normalized Mean Bias (NMB)	$NMB = \frac{\sum_{i=1}^N (M_i - O_i)}{\sum_{i=1}^N O_i} \times 100\%$
Correlation Coefficient (R)	$Corr.R = \frac{\sum_{i=1}^N (M_i - \bar{M})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^N (M_i - \bar{M})^2} \sqrt{\sum_{i=1}^N (O_i - \bar{O})^2}}$
Root Mean Square Error (RMSE)	$RMSE = \sqrt{1/N \sum_{i=1}^N (M_i - O_i)^2}$

3. Meteorological Model Evaluation and Improvement

3.1. Experimental Setup

To select model configurations that best represent the 2021 offshore monitoring data, we previously designed seven model experiments for a TCEQ-funded project titled ‘Analysis of 2021 Offshore Monitoring’ ([Base], [YSU], [ACM2], [WSM6], [ERA5], [Nudged], and [SIP] in **Table 3**). For this AQRP project, we additionally added three model experiments to improve model performance based on our previous findings from the TCEQ-funded project ([HRRR], [Nudged2], and [Reinit] in **Table 3**). In all, ten model experiments with different boundary and initial conditions, physics options, and data assimilation options were presented in the following analysis.

First, [Base] is the baseline configuration that represents our initial guess of the best model 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. Thus, 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 WRF performance by using the ERA5 and HRRR instead of NCEP FNL, respectively. Further, we examined the effects of data assimilation options in [Nudged] and [Nudged2]. [Nudged] adopted observation nudging and surface analysis nudging to assimilate TCEQ Continuous Ambient Monitoring Station (CAMS) only, while [Nudged2] is an updated version with more observations from multiple platforms, including TCEQ CAMS, boat, and NCEP surface and upper air, assimilated into the model. Differences between [Base] and two nudged simulations show the effects of assimilation. Next, [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. Differences between [Base] and [Reinit] show the effects of model reinitializing over free-running simulation. Last, [SIP] adopts the WRF settings from TCEQ 2019 SIP and serves as a reference comparison for other model experiments.

Each simulation listed in **Table 3** was performed for two periods: July 20-30 and August 20-October 13. These periods cover five high ozone episodes: July 26-28, August 25, September 6-11, September 23-26, and October 6-9.

Table 3. List of model experiments. [Base], [YSU], [ACM2], [WSM6], [ERA5], [Nudged], and [SIP] are the seven model experiments for a previous TCEQ-funded project titled ‘Analysis of 2021 Offshore Monitoring’. [HRRR], [Nudged2] and [Reinit] are the three model experiments for this AQRP project.

Simulations	BC Meteorology	PBL	Microphysics	Nudging	Reinitializing
[Base]	NCEP FNL	MYNN	2M	No	No
[YSU]	NCEP FNL	YSU	2M	No	No
[ACM2]	NCEP FNL	ACM2	2M	No	No
[WSM6]	NCEP FNL	MYNN	WSM6	No	No
[ERA5]	ECMWF ERA5	MYNN	2M	No	No
[HRRR]	HRRR	MYNN	2M	No	No
[Nudged]	NCEP FNL	MYNN	2M	Yes	No
[Nudged2]	NCEP FNL	MYNN	2M	Yes	No
[Reinit]	NCEP FNL	MYNN	2M	No	Yes
[SIP]	ECMWF ERA5	YSU	WSM6	No	No

3.2. PBL Evaluation

This section evaluates modeled PBL height with two types of observations. The first type is PBL derived from the HSRL2 flying over urban Houston and Galveston Bay for ten days (on September 1, 3, 8-11, and 23-26 of 2021), and the second type is PBL derived from the ceilometers at a ground-based La Porte site (for the whole September) and on mobile boats operating in the Galveston Bay (from late July to early October). For evaluation of spatial variabilities, the HSRL2 and boat measurements are used due to their mobile tracks. For temporal variabilities, the boat and La Porte measurements are shown for the whole September of 2021 due to their long records, while the HSRL2 is shown for diurnal patterns because of short records of only ten non-consecutive days in September. Among all model simulations, [HRRR], [Nudged2], and [Reinit] are the three best simulations based on campaign-wide statistics (**Tables 4–7**) and are thus selected to show detailed model performances in the figures in this and the following sections.

The HSRL2 scanned the greater Houston region three times a day for ten non-consecutive days in September of 2021, September 1, 3, 8-11, and 23-26. Among these days, one ozone exceedance day, September 9, is selected for the evaluation of PBL spatial variabilities in **Figure 3**. The model simulations show the mean morning, noon, and afternoon PBL heights of 275-324 meters (m), 886-1128 m, and 1408-1555 m, in comparison with the observed values of 471 m, 983 m, and 1675 m, respectively. The model simulations capture spatial variabilities at noon ($R=0.62-0.77$) and in the afternoon ($R=0.71-0.76$) but have difficulties capturing that in the morning ($R=-0.1-0.14$). Relatively less correlation in the morning is probably related to the presence of complex nocturnal PBL structure with both residual layer and stable surface layer. Despite less spatial correlation in the morning, land-water differences are well represented by the model throughout the day, with lower PBL heights over water than on land. Meanwhile, diurnal variations of PBL heights over the ten days are shown in **Figure 4**. Compared to water, land has a relatively stronger increase in PBL heights from early morning to afternoon. The model simulations capture the strong PBL variations on land with consistent minor underestimation throughout all hours, but the model has difficulty in capturing PBL variations over water particularly in the afternoon hours despite different configurations.

In addition to the HSRL2 observations, two ceilometers, that is one on a mobile boat and the other one at a ground-based stationary site named La Porte, also observe PBL heights. **Figure 5** shows the spatial variabilities of modeled and observed PBL over Galveston Bay. The model simulations show the mean PBL height of 462–671 m in comparison with the observed value of 856 m over water. The model captures the low PBL in the morning but has difficulties capturing the high PBL in the afternoon, leading to low correlation among different configurations ($R=0.25\text{--}0.41$). Meanwhile, **Figure 6** shows hourly time series of PBL heights measured by boat and at La Porte for temporal variability evaluation. Corresponding statistics of the two comparisons are shown in **Table 4** and **Table 5**, respectively. All configurations underestimate both PBL on land and over water with a correlation coefficient ranging from 0.25 to 0.47.

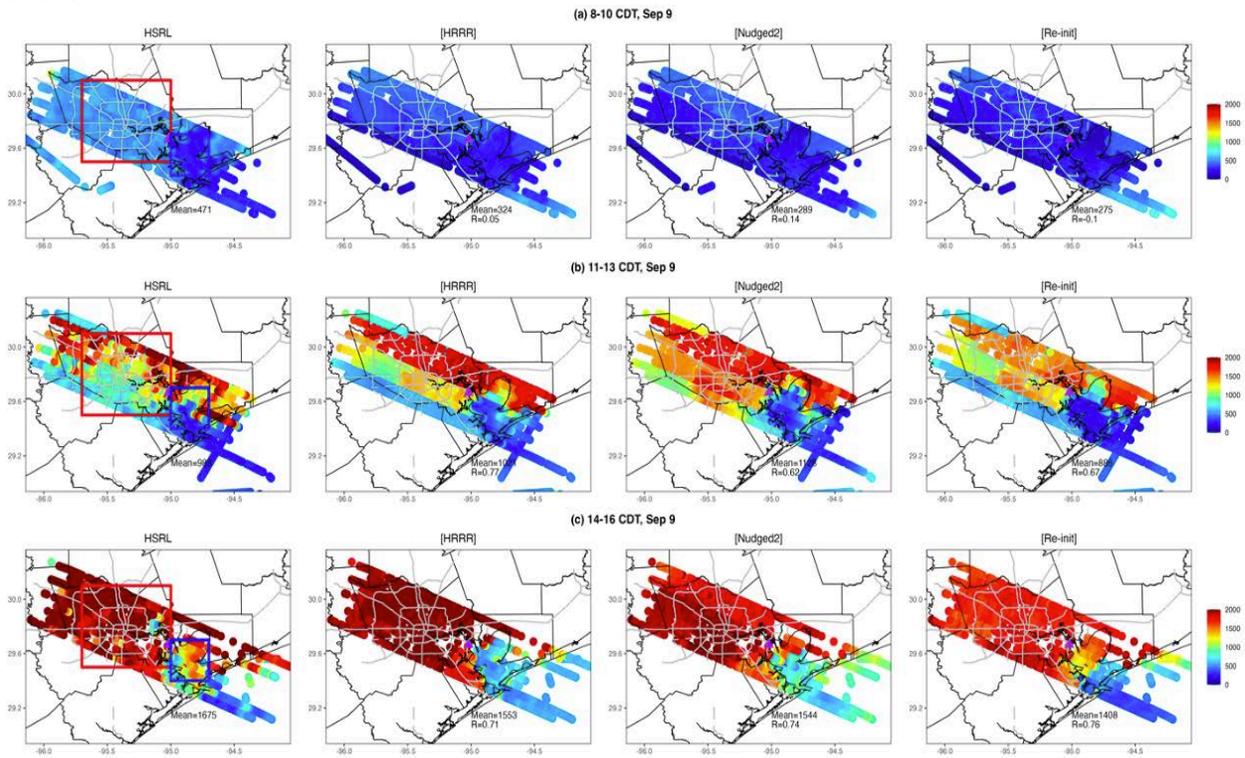


Figure 3. Spatial distribution of HSRL2-observed and modeled PBL (a) in the morning, (b) at noon, and (c) in the afternoon of September 9, 2021. Red and blue boxes denote land and water, respectively. Time used here is local Central Daylight Time (CDT), 5 hours behind UTC.

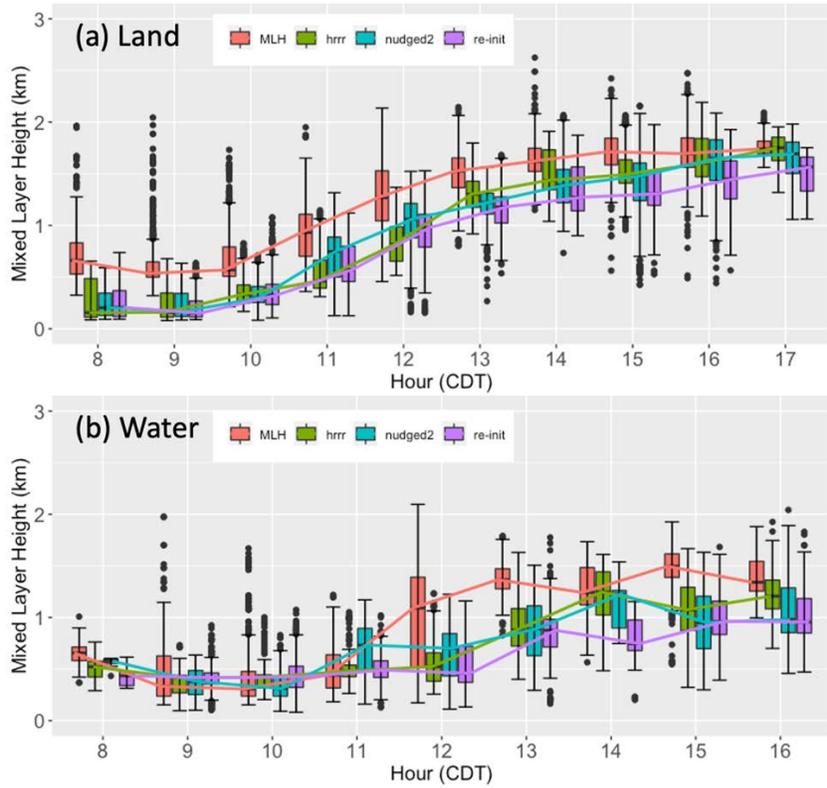


Figure 4. Diurnal variabilities of HSRL2-observed and modeled PBL (a) on land and (b) over water. Land and water are respectively defined as red and blue boxes in Figure 3.

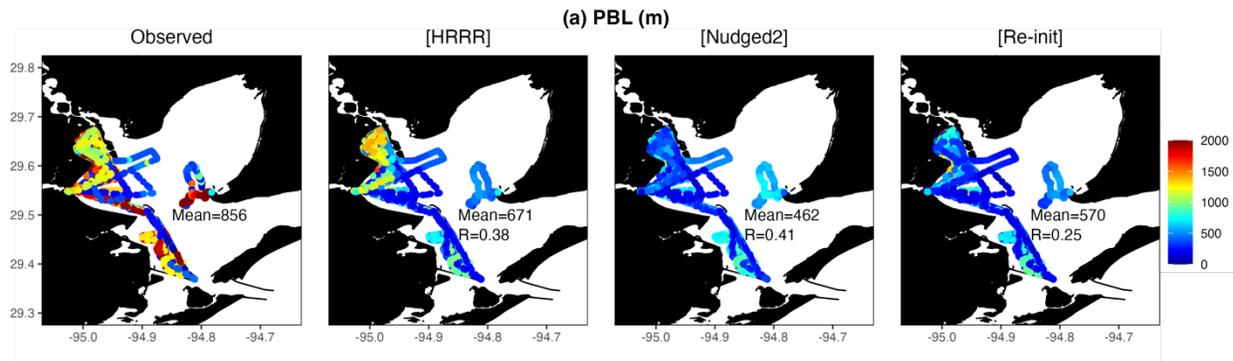


Figure 5. Spatial distribution of boat-observed and modeled PBL during ozone episodes.

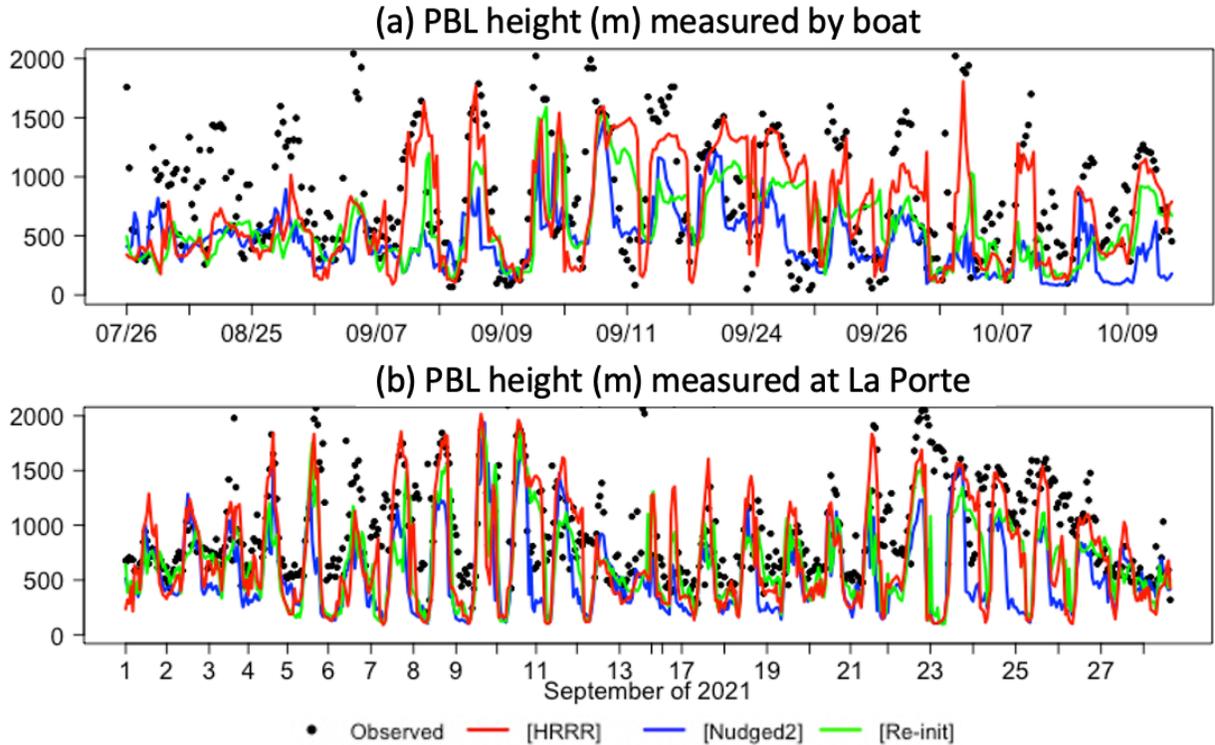


Figure 6. Hourly time series of modeled PBL against (a) boat observations and (b) ground-based observations at the La Porte site. Corresponding statistics are shown in Table 4 and Table 5, respectively.

Table 4. Performance metrics of spatiotemporal variability of PBL height (m) between the observations on boat and the WRF model for five ozone episodes. OBS and MOD stand for observation and model, respectively.

Variables	Simulation	OBS	MOD	R	NMB	MB	MAE	RMSE
Boundary layer	[Base]	855.58	499.27	0.32	-0.42	-356.30	529.63	699.67
	[WSM6]	855.58	526.69	0.30	-0.38	-328.88	526.38	691.82
	[YSU]	855.58	322.22	0.30	-0.62	-533.36	612.29	817.16
	[ACM2]	855.58	443.60	0.30	-0.48	-411.97	562.12	747.06
	[ERA5]	855.58	464.75	0.47	-0.46	-390.83	507.51	680.30
	[HRRR]	855.58	671.27	0.38	-0.22	-184.31	461.30	637.68
	[Nudged]	855.58	462.20	0.47	-0.46	-393.37	497.27	680.35
	[Nudged2]	855.58	462.09	0.41	-0.46	-393.48	516.18	696.37
	[Reinit]	855.58	569.57	0.25	-0.33	-286.00	518.21	689.22
	[SIP]	855.58	399.24	0.57	-0.53	-456.33	514.01	697.39

Table 5. Performance metrics of temporal variability of PBL height (m) between the observations at La Porte and the WRF model for September 2021. OBS and MOD stand for observation and model, respectively.

Variables	Simulation	OBS	MOD	R	NMB	MB	MAE	RMSE
Boundary layer	[Base]	981.26	602.49	0.38	-0.39	-378.77	435.35	602.42
	[WSM6]	981.26	609.48	0.43	-0.38	-371.78	444.37	602.36
	[YSU]	981.26	453.32	0.34	-0.54	-527.94	553.26	721.32
	[ACM2]	981.26	410.16	0.37	-0.58	-571.10	609.09	761.76
	[ERA5]	981.26	563.55	0.33	-0.43	-417.70	468.69	636.41
	[HRRR]	981.26	721.95	0.42	-0.26	-259.31	393.02	572.44
	[Nudged]	981.26	610.54	0.36	-0.38	-370.72	440.08	619.22
	[Nudged2]	981.26	555.91	0.38	-0.43	-425.35	463.91	632.75
	[Reinit]	981.26	665.43	0.41	-0.32	-315.83	387.92	562.18
	[SIP]	981.26	462.69	0.32	-0.53	-518.57	586.24	745.60

3.3. Evaluation of Other Meteorological Variables

In this section, we evaluated WRF simulations against three independent observations including TCEQ CAMS sites (Section 3.3.1), boat observations (Section 3.3.2), and surface measurements at the La Porte site (Section 3.3.3). Among all model simulations, [HRRR], [Nudged2], and [Reinit] are the three best simulations based on campaign-wide statistics (Tables 4–7) and are thus selected to show detailed model performances in the figures in this section.

3.3.1 CAMS sites

Figure 7 shows the evaluation of spatial variabilities between CAMS-observed and WRF-modeled meteorology. The mean air temperature of 26.3°C averaged over five ozone episodes is slightly underestimated by the modeled value of 25.7–26.1°C. The model shows a larger land-water gradient than the observations, with overestimation in urban Houston but underestimation near the coast of Galveston Bay. This probably suggests a stronger heat island effect in the model than the reality. As temperature and humidity are negatively correlated, the modeled humidity shows underestimation at the urban core but overestimation near the coast. For wind estimates, the modeled mean wind speed of 0.9–1.1 m s⁻¹ is higher than the observed value of 0.7 m s⁻¹, but the averaged wind direction, which is northeasterly, is well captured by all three model simulations. Meanwhile, **Figure 8** shows temporal variabilities, and the corresponding statistics are shown in **Table 6**. Compared to observations, the modeled mean temperatures are generally less than 1 °C lower with correlation coefficients in the ranges of 0.85-0.92. Modeled mean relative humidity is 1-4% higher than observations with correlation coefficients in the ranges of 0.71-0.80. Modeled wind speed is 0.1-1.0 m s⁻¹ higher than

observations with correlation coefficients in the ranges of 0.33-0.50. Observed and modeled winds are both northeasterlies.

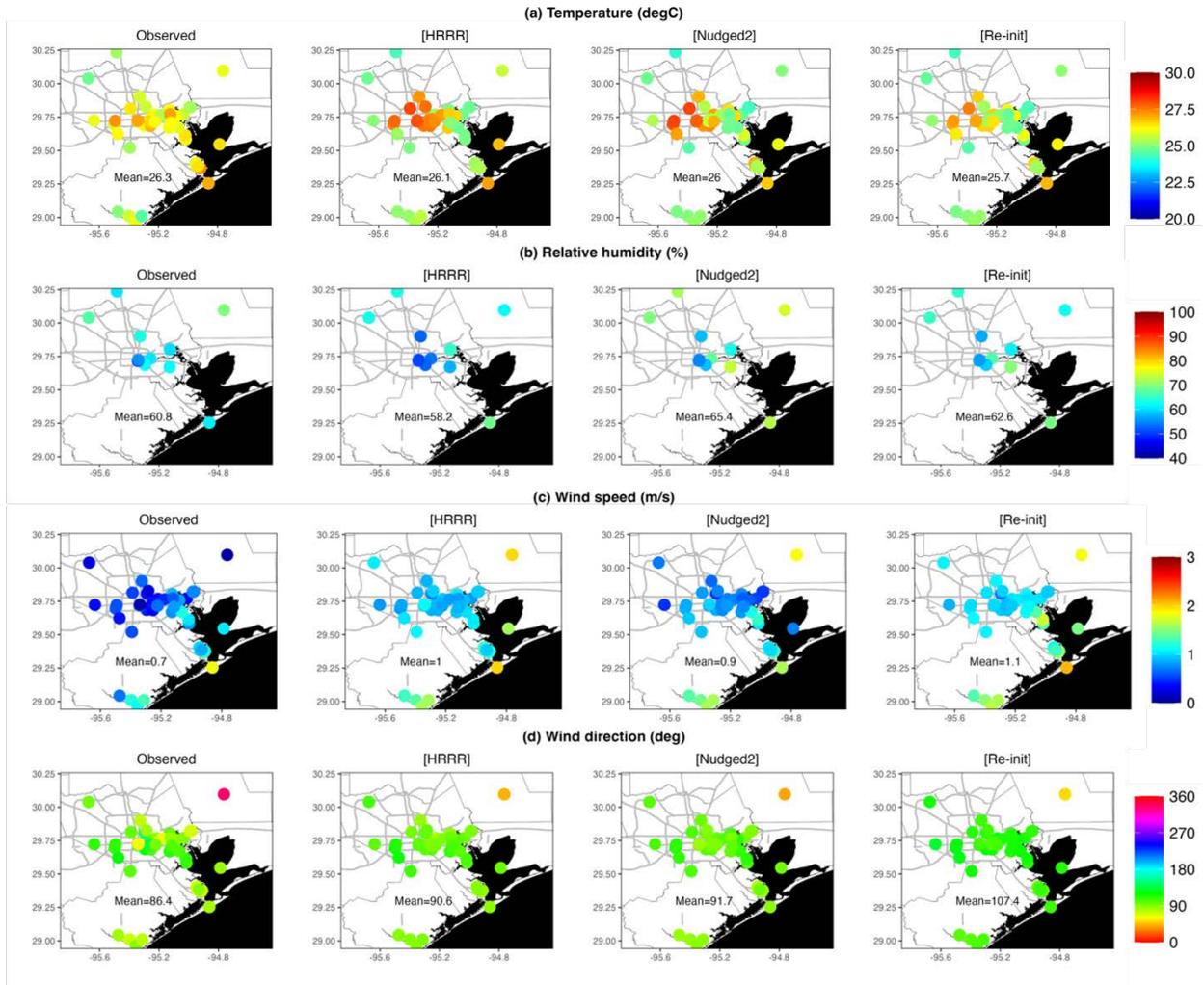


Figure 7. Spatial distribution of CAMS-observed and modeled mean meteorology of (a) air temperature, (b) relative humidity, (c) wind speed, and (d) wind direction during ozone episodes.

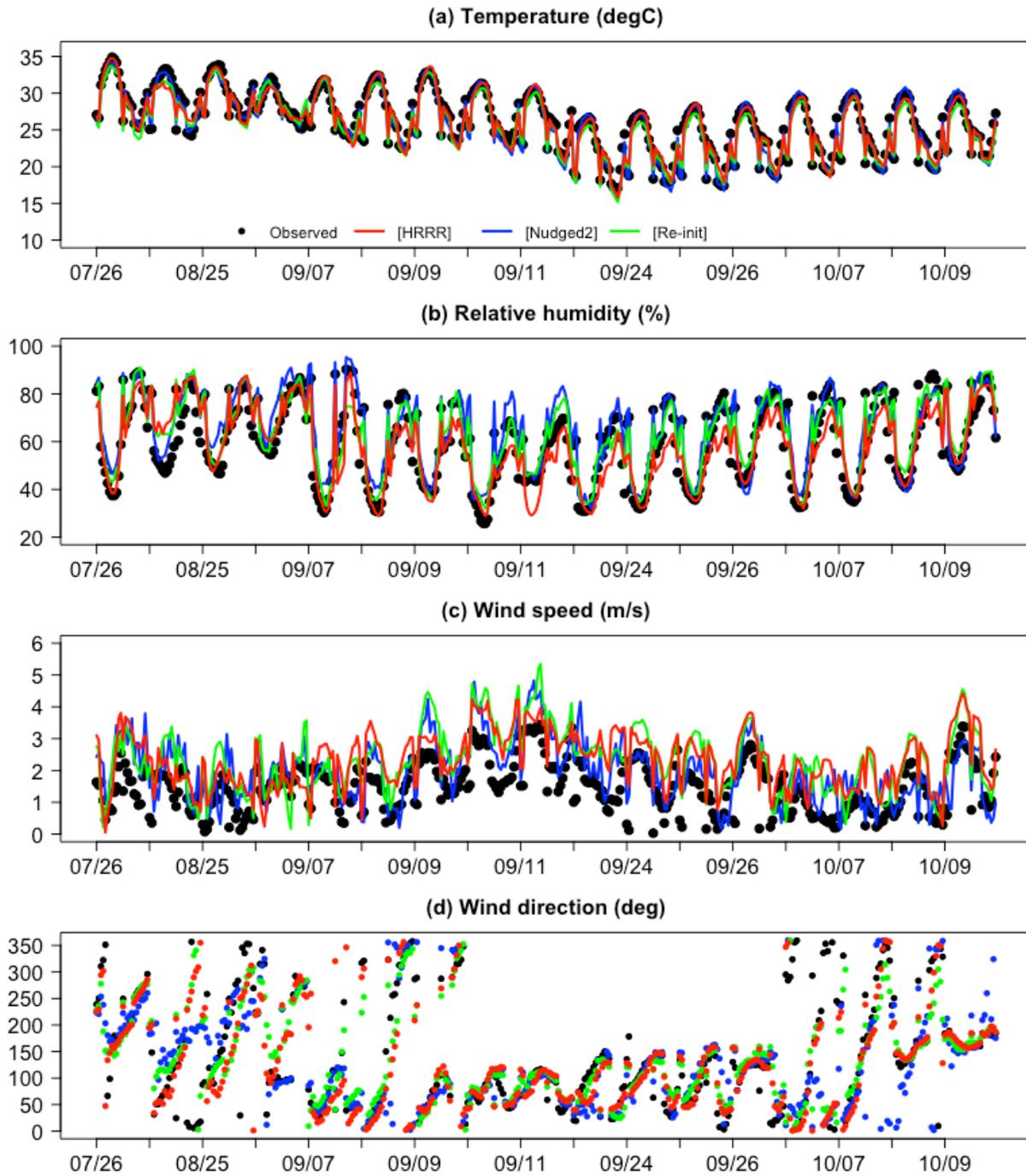


Figure 8. Hourly time series of (a) air temperature, (b) relative humidity, (c) wind speed, and (d) wind direction between CAMS observations and WRF model simulations for five ozone episodes, that is Jul 28, Aug 25, Sep 6-11, Sep 23-26, and Oct 6-9.

Table 6. Performance metrics of spatiotemporal variability between CAMS-observed and WRF-modeled meteorology for five ozone episodes, that is Jul 28, Aug 25, Sep 6-11, Sep 23-26, and Oct 6-9. Hourly temperature ($^{\circ}\text{C}$), relative humidity (%), wind speed (m s^{-1}), and wind direction (deg) at all stations is used for the calculation of the 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 stand for observation and model, respectively.

Variables	Simulation	OBS	MOD	R	NMB	MB	MAE	RMSE
Temperature	[Base]	26.18	25.82	0.88	-0.01	-0.36	1.69	2.15
	[WSM6]	26.18	25.84	0.89	-0.01	-0.35	1.57	1.99
	[YSU]	26.18	26.29	0.89	0.00	0.11	1.65	2.11
	[ACM2]	26.18	25.95	0.86	-0.01	-0.23	1.76	2.23
	[ERA5]	26.18	24.91	0.85	-0.05	-1.28	2.17	2.71
	[HRRR]	26.18	26.12	0.89	0.00	-0.06	1.59	2.05
	[Nudged]	26.18	25.82	0.92	-0.01	-0.36	1.47	1.89
	[Nudged2]	26.18	25.92	0.92	-0.01	-0.26	1.43	1.84
	[Reinit]	26.18	25.69	0.92	-0.02	-0.49	1.41	1.77
	[SIP]	26.18	25.25	0.85	-0.04	-0.93	2.03	2.54
Relative humidity	[Base]	60.12	60.94	0.76	0.01	0.82	10.25	13.04
	[WSM6]	60.12	62.21	0.78	0.03	2.09	9.85	12.28
	[YSU]	60.12	58.45	0.80	-0.03	-1.68	9.54	12.31
	[ACM2]	60.12	62.73	0.71	0.04	2.60	11.40	14.71
	[ERA5]	60.12	64.21	0.77	0.07	4.08	10.55	12.76
	[HRRR]	60.12	57.82	0.79	-0.04	-2.30	9.13	12.13
	[Nudged]	60.12	62.22	0.79	0.03	2.10	10.46	13.38
	[Nudged2]	60.12	64.63	0.82	0.08	4.51	9.54	12.05
	[Reinit]	60.12	62.57	0.84	0.04	2.45	8.37	10.66
	[SIP]	60.12	62.42	0.67	0.04	2.30	12.07	15.37
Wind speed	[Base]	0.67	1.29	0.35	0.59	1.01	1.40	1.70
	[WSM6]	0.67	1.67	0.37	0.61	1.04	1.39	1.72
	[YSU]	0.67	0.80	0.39	0.75	1.29	1.55	1.87

	[ACM2]	0.67	1.16	0.38	0.66	1.12	1.44	1.77
	[ERA5]	0.67	1.76	0.43	0.64	1.09	1.38	1.66
	[HRRR]	0.67	1.00	0.54	0.49	0.83	1.12	1.36
	[Nudged]	0.67	0.86	0.50	0.31	0.53	1.00	1.28
	[Nudged2]	0.67	0.89	0.55	0.30	0.51	0.96	1.20
	[Reinit]	0.67	1.14	0.61	0.48	0.82	1.07	1.31
	[SIP]	0.67	1.30	0.33	0.69	1.19	1.51	1.82
Wind direction	[Base]	87.76	72.32	0.43	-0.05	-7.67	56.5	73.36
	[WSM6]	87.76	72.56	0.38	-0.04	-5.51	56.41	72.93
	[YSU]	87.76	53.26	0.41	-0.08	-12.14	60.30	77.29
	[ACM2]	87.76	54.87	0.37	-0.07	-10.64	64.15	81.29
	[ERA5]	87.76	47.32	0.43	-0.07	-10.92	58.05	74.83
	[HRRR]	87.76	92.51	0.61	-0.02	-3.43	40.16	57.55
	[Nudged]	87.76	79.81	0.54	-0.03	-4.82	48.41	65.44
	[Nudged2]	87.76	93.29	0.48	0.02	3.00	46.05	64.70
	[Reinit]	87.76	109.03	0.47	0.00	-0.32	39.99	57.67
	[SIP]	87.76	64.15	0.37	-0.05	-7.30	60.44	77.55

3.3.2 Boat observations

Figure 9 shows the evaluation of spatial variabilities between boat-observed and WRF-modeled meteorology. The mean air temperature of 27°C averaged over five ozone episodes is slightly underestimated by the modeled value of 26–27°C. The temperature over water is ~1°C higher than that on land during the same period. The mean relative humidity of 61% is overestimated by the modeled value of 69–73%. The mean humidity over water and on land is the same. For wind estimates, the modeled mean wind speed of 3–4 m s⁻¹ is significantly higher than the observed value of 2 m s⁻¹, but the wind direction is captured to different extents (R=0.45–0.62). Offshore winds (2 m s⁻¹) are observed to be a lot stronger than winds on land (0.7 m s⁻¹). Meanwhile, **Figure 10** shows temporal variabilities, and the corresponding statistics are shown in **Table 7**. Compared to observations, the modeled mean temperatures are 0.1–1.7 °C lower with correlation coefficients in the ranges of 0.75–0.88. Modeled mean relative humidity is 7–12% higher than observations with correlation coefficients in the ranges of 0.56–0.68. Modeled wind speed is 0.9–1.9 m s⁻¹ higher than observations with correlation coefficients in the ranges of 0.13–0.30. Observed and modeled offshore winds are both southeasterlies, in comparison with northeasterlies on land.

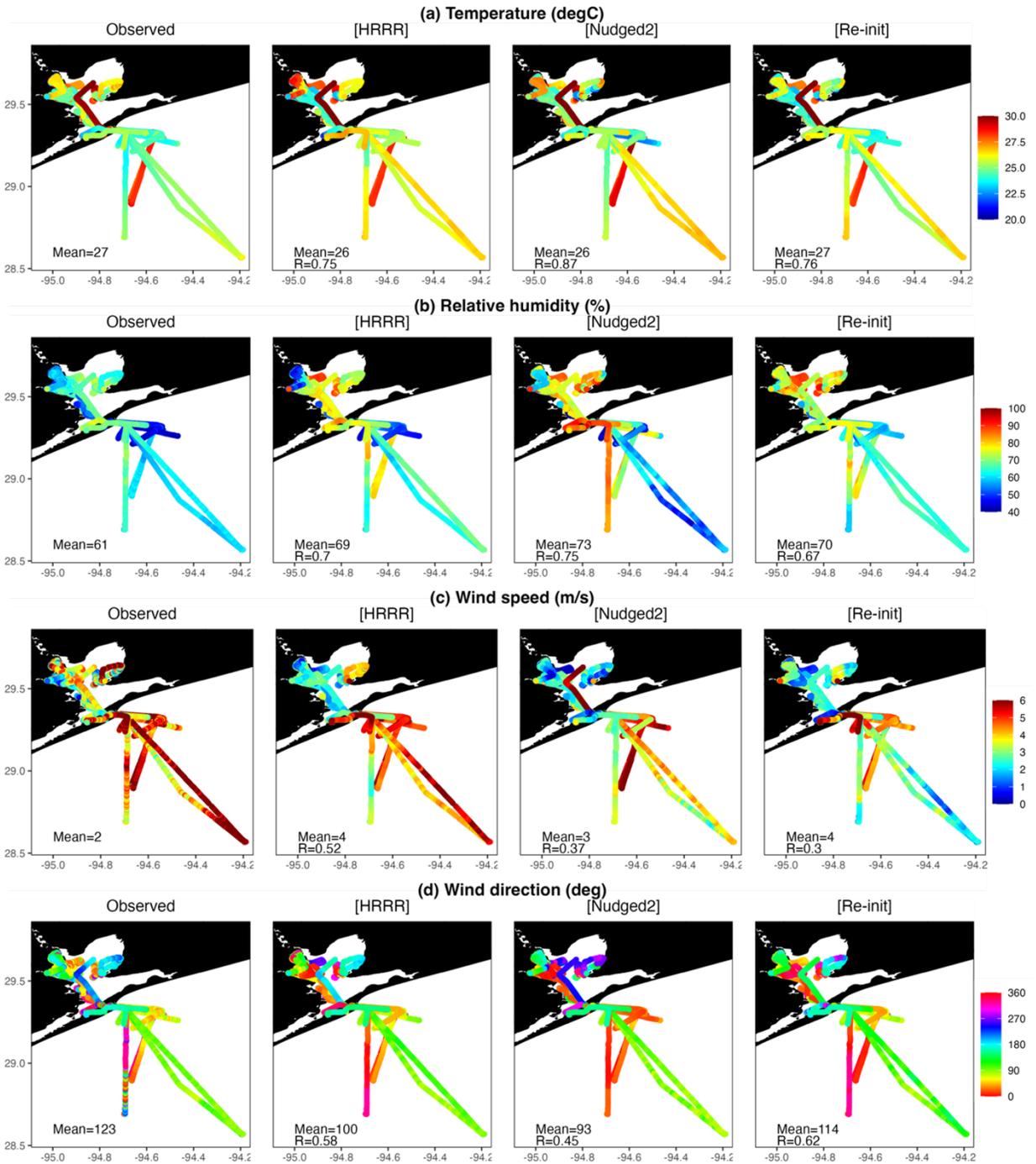


Figure 9. Spatial distribution of boat-observed and modeled meteorology during ozone episodes.

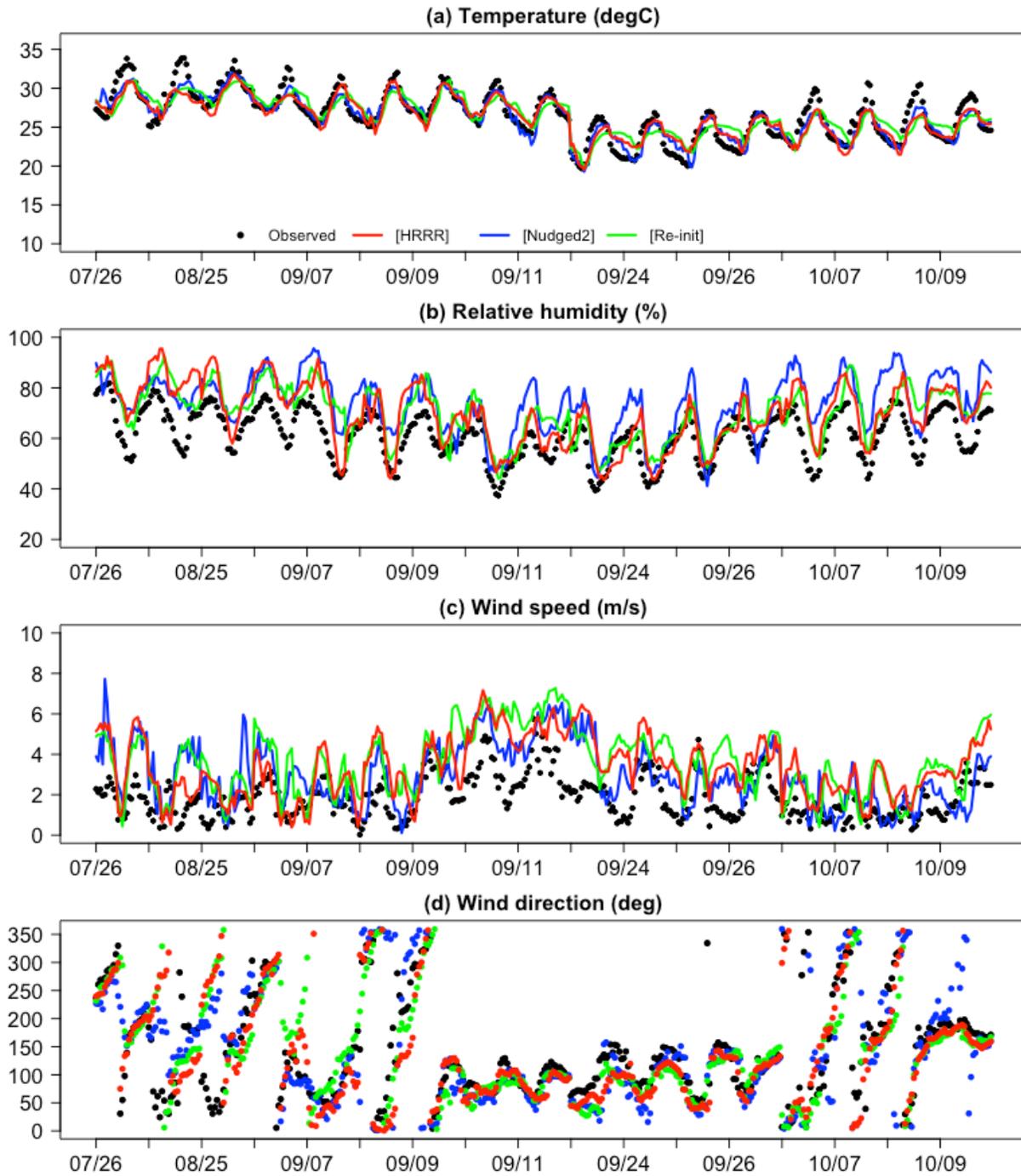


Figure 10. Hourly time series of (a) air temperature, (b) relative humidity, (c) wind speed, and (d) wind direction between boat observations and WRF model simulations for five ozone episodes, that is Jul 28, Aug 25, Sep 6-11, Sep 23-26, and Oct 6-9 in 2021.

Table 7. Performance metrics of spatiotemporal variability between boat-observed and WRF-modeled temperature ($^{\circ}\text{C}$), relative humidity (%), wind speed (m s^{-1}), and wind direction (deg) for five ozone episodes. 1-minute meteorology used for 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 stand for observation and model, respectively.

Variables	Simulation	OBS	MOD	R	NMB	MB	MAE	RMSE
Temperature	[Base]	26.55	26.45	0.77	0.00	-0.11	1.71	2.14
	[WSM6]	26.55	26.50	0.75	0.00	-0.05	1.77	2.20
	[YSU]	26.55	26.78	0.78	0.01	0.22	1.71	2.10
	[ACM2]	26.55	26.51	0.75	0.00	-0.04	1.78	2.21
	[ERA5]	26.55	24.85	0.75	-0.06	-1.70	2.21	3.00
	[HRRR]	26.55	26.30	0.75	-0.01	-0.25	1.89	2.29
	[Nudged]	26.55	26.19	0.88	-0.01	-0.36	1.27	1.64
	[Nudged2]	26.55	26.30	0.87	-0.01	-0.25	1.26	1.65
	[Reinit]	26.55	26.53	0.76	0.00	-0.02	1.71	2.15
	[SIP]	26.55	24.88	0.77	-0.06	-1.67	2.28	2.89
Relative humidity	[Base]	60.96	70.24	0.64	0.15	9.28	11.95	14.59
	[WSM6]	60.96	71.09	0.61	0.17	10.14	11.76	14.38
	[YSU]	60.96	68.20	0.65	0.12	7.24	10.96	13.29
	[ACM2]	60.96	69.35	0.56	0.14	8.40	12.75	15.33
	[ERA5]	60.96	74.38	0.60	0.22	13.42	14.66	17.23
	[HRRR]	60.96	69.20	0.70	0.14	8.24	10.38	12.68
	[Nudged]	60.96	71.60	0.68	0.17	10.65	12.54	15.78
	[Nudged2]	60.96	73.35	0.75	0.20	12.39	12.87	14.92
	[Reinit]	60.96	69.68	0.67	0.14	8.72	10.25	12.42
	[SIP]	60.96	72.52	0.63	0.19	11.57	14.51	17.3
Wind speed	[Base]	0.73	2.47	0.16	0.74	1.67	2.20	2.78
	[WSM6]	0.73	2.62	0.14	0.82	1.85	2.33	2.92
	[YSU]	0.73	2.17	0.13	0.99	2.22	2.63	3.19
	[ACM2]	0.73	1.99	0.15	0.92	2.07	2.49	3.09
	[ERA5]	0.73	1.89	0.22	0.78	1.74	2.21	2.72
	[HRRR]	0.73	1.68	0.52	0.59	1.32	1.69	2.05
	[Nudged]	0.73	1.66	0.30	0.41	0.92	1.63	2.11
	[Nudged2]	0.73	1.75	0.37	0.41	0.92	1.57	1.96
	[Reinit]	0.73	2.02	0.30	0.69	1.55	2.00	2.41
	[SIP]	0.73	2.01	0.30	0.65	1.46	2.01	2.43
Wind direction	[Base]	144.15	118.78	0.32	-0.08	-11.45	57.74	75.38
	[WSM6]	144.15	113.5	0.26	-0.13	-19.10	60.40	77.29
	[YSU]	144.15	135.77	0.26	-0.11	-16.44	63.52	81.13

	[ACM2]	144.15	125.25	0.27	-0.11	-17.20	68.93	85.92
	[ERA5]	144.15	96.69	0.18	-0.17	-25.20	69.00	85.30
	[HRRR]	144.15	137.93	0.58	-0.08	-12.53	41.54	58.16
	[Nudged]	144.15	124.48	0.44	-0.10	-14.86	52.59	69.23
	[Nudged2]	144.15	146.95	0.45	-0.05	-7.68	47.87	65.51
	[Reinit]	144.15	146.96	0.62	-0.10	-14.98	42.98	59.66
	[SIP]	144.15	103.91	0.28	-0.14	-20.87	64.49	81.48

3.3.3 La Porte surface measurements

We show temporal variability between surface measurements and WRF-modeled meteorology at the La Porte site throughout the whole September 2021 in **Figure 11** and **Table 8**. No spatial variability is shown because the measurements are stationary and do not have spatial variability. Other periods are not shown due to data availability. Observations at 30-second resolution are averaged into hourly. Compared to observations, the modeled mean temperatures are 0.5-1.8 °C lower with correlation coefficients in the ranges of 0.81-0.91. Modeled mean relative humidity is 1-5% higher than observations with correlation coefficients in the ranges of 0.78-0.85. Modeled wind speed is 0.1-0.5 m s⁻¹ lower than observations with correlation coefficients in the ranges of 0.41-0.62. Observed and modeled winds are both southeasterlies but have 26-50 degree differences.

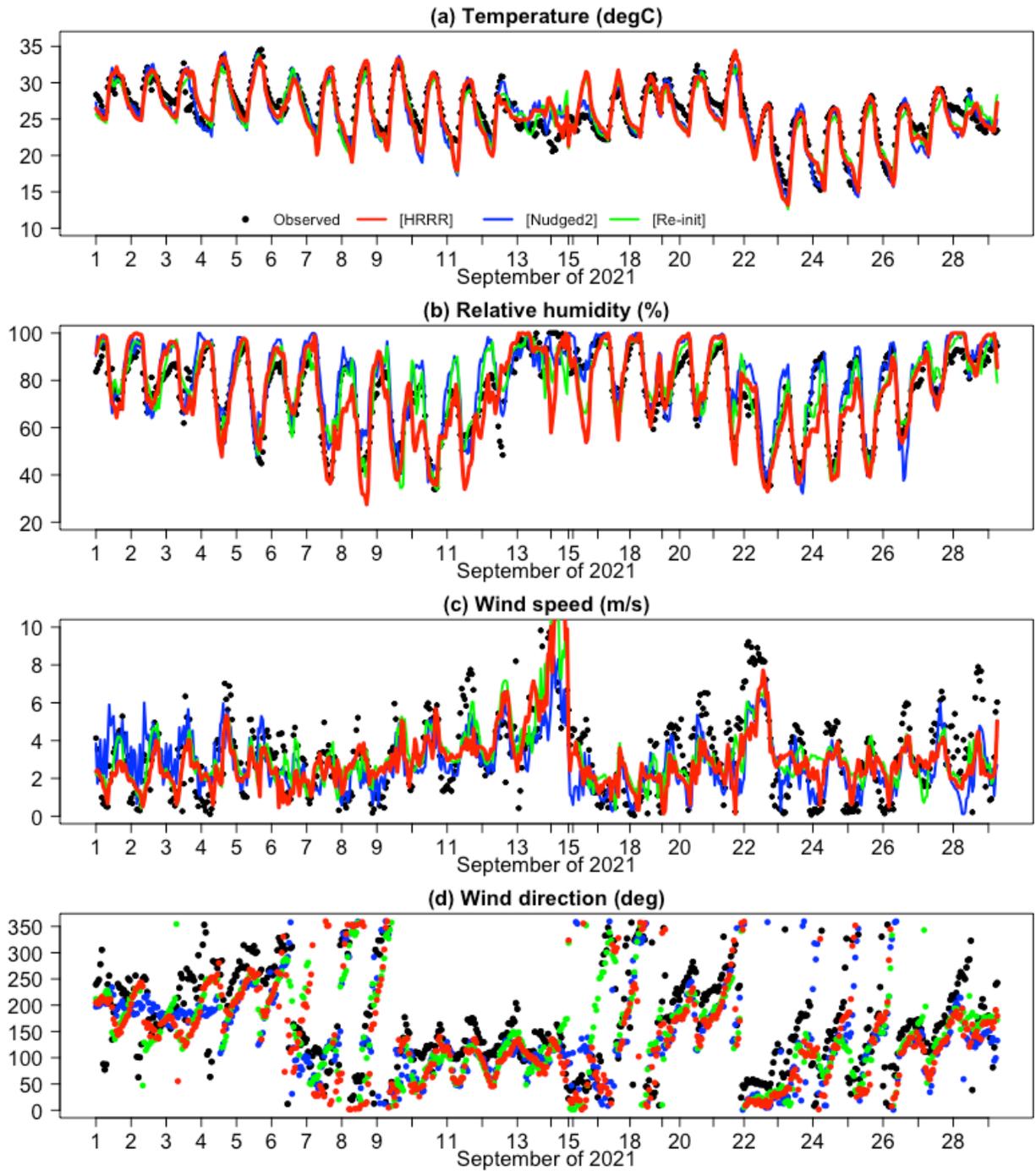


Figure 11. Hourly time series of (a) air temperature, (b) relative humidity, (c) wind speed, and (d) wind direction between surface observations and WRF model simulations at the La Porte site during the whole September in 2021.

Table 8. Performance metrics of temporal variability between La Porte-observed and WRF-modeled temperature ($^{\circ}\text{C}$), relative humidity (%), wind speed (m s^{-1}), and wind direction (deg) for the whole September 2021. Meteorology at the 30-second interval at the La Porte site is averaged hourly and then 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 stand for observation and model, respectively.

Variables	Simulation	OBS	MOD	R	NMB	MB	MAE	RMSE
Temperature	[Base]	26.10	25.26	0.86	-0.03	-0.84	1.75	2.17
	[WSM6]	26.10	25.34	0.88	-0.03	-0.75	1.65	1.96
	[YSU]	26.10	25.59	0.84	-0.02	-0.50	1.74	2.22
	[ACM2]	26.10	25.40	0.82	-0.03	-0.69	1.85	2.29
	[ERA5]	26.10	24.32	0.81	-0.07	-1.77	2.47	2.94
	[HRRR]	26.10	25.71	0.89	-0.01	-0.39	1.45	1.85
	[Nudged]	26.10	25.36	0.91	-0.03	-0.74	1.44	1.85
	[Nudged2]	26.10	25.60	0.92	-0.02	-0.49	1.26	1.62
	[Reinit]	26.10	25.61	0.90	-0.02	-0.49	1.43	1.75
	[SIP]	26.10	24.93	0.85	-0.04	-1.16	1.92	2.36
Relative humidity	[Base]	75.75	77.93	0.82	0.03	2.18	7.93	10.21
	[WSM6]	75.75	79.78	0.82	0.05	4.03	8.22	10.59
	[YSU]	75.75	75.92	0.84	0.00	0.17	8.15	10.31
	[ACM2]	75.75	77.06	0.78	0.02	1.32	9.14	11.42
	[ERA5]	75.75	81.13	0.79	0.07	5.38	8.94	11.26
	[HRRR]	75.75	74.73	0.86	-0.01	-1.02	7.16	9.39
	[Nudged]	75.75	76.53	0.84	0.01	0.78	8.05	10.42
	[Nudged2]	75.75	79.75	0.90	0.05	4.00	6.83	8.38
	[Reinit]	75.75	76.85	0.90	0.01	1.11	5.60	7.38
	[SIP]	75.75	76.81	0.85	0.01	1.06	6.92	8.83
Wind speed	[Base]	3.26	3.24	0.46	-0.01	-0.03	1.49	1.97
	[WSM6]	3.26	3.06	0.45	-0.06	-0.21	1.48	2.00
	[YSU]	3.26	3.27	0.41	0.00	0.00	1.57	2.05
	[ACM2]	3.26	3.16	0.47	-0.03	-0.11	1.50	1.97
	[ERA5]	3.26	3.11	0.50	-0.05	-0.15	1.48	1.93
	[HRRR]	3.26	2.98	0.63	-0.09	-0.28	1.36	1.76
	[Nudged]	3.26	2.74	0.62	-0.16	-0.52	1.35	1.80
	[Nudged2]	3.26	2.81	0.67	-0.14	-0.46	1.31	1.69
	[Reinit]	3.26	3.19	0.71	-0.02	-0.07	1.25	1.55
	[SIP]	3.26	3.38	0.43	0.04	0.12	1.60	2.02
Wind direction	[Base]	168.12	132.39	0.42	-0.17	-29.31	67.48	82.07
	[WSM6]	168.12	129.83	0.44	-0.19	-32.45	64.48	77.37

	[YSU]	168.12	141.99	0.47	-0.21	-36.05	69.08	82.81
	[ACM2]	168.12	134.66	0.40	-0.16	-27.63	72.85	87.07
	[ERA5]	168.12	117.89	0.37	-0.20	-33.31	71.19	83.86
	[HRRR]	168.12	140.05	0.65	-0.18	-30.41	54.89	67.08
	[Nudged]	168.12	137.40	0.59	-0.21	-35.39	59.41	72.99
	[Nudged2]	168.12	144.84	0.64	-0.18	-29.70	53.43	65.66
	[Reinit]	168.12	147.45	0.59	-0.15	-25.92	52.55	64.92
	[SIP]	168.12	123.52	0.42	-0.18	-30.59	71.83	86.58

3.4. Summary

We evaluated ten WRF configurations of different meteorological inputs, physics options, and data assimilation options against both onshore and offshore observations. 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], [Nudged2] and [Reinit] are found to be the three best simulations based on campaign-wide statistics. Considering that [Nudged2] requires additional efforts to prepare observational datasets and [Reinit] needs to automate the model running process, [HRRR] is the easiest but the most effective option to reproduce meteorology during the TRACER-AQ 2021 campaign.

4. Photochemical Model Evaluation and Model Intercomparison

A comparative evaluation of multiple photochemical models against the TRACER-AQ observations will provide insights into the validity of photochemical model configurations and help identify the strengths and shortcomings of the many components in the complex meteorological-photochemical modeling system used for the State Implementation Plan (SIP). The project team evaluated three photochemical models: CAMx, WRF-GC, and WRF-Chem, all using WRF meteorology to drive photochemistry and transport.

The CAMx simulation was performed for two periods, July 20 – 30 and August 20 – October 13, 2021, to cover all six high-O₃ episodes. A 10-day spin-up before each period was applied. The WRF-Chem simulation was performed from September 1–30, covering three high-O₃ episodes in September 2021, with a 1-day spin-up applied. The WRF-GC simulation was performed from September 1– October 1, covering three high-O₃ episodes in September 2021.

4.1 Evaluation of Photochemical Models with TRACER-AQ Observations

In this section, we evaluated all three models against in situ observations of O₃ collected at CAMS sites, by boats, and by mobile laboratories during the TRACER-AQ field campaign period. Model comparisons with remote-sensing-based observations (e.g., aircraft and ground-based ozone lidar) are presented in the next Section to aid model intercomparison. The CAMx model was evaluated for all six episodes, while WRF-GC and WRF-Chem were evaluated for the three episodes in September 2021 because their simulations were conducted for this month

only. The evaluation statistics are separately shown for all days with spatiotemporally coincident model results and observations, ozone episode days as defined above, and clean days which exclude episode days from all days. Among the three episodes in September 2021, the second one (Sep 17-19) was caused by the passing of Hurricane Nicolas. Because this was a unique meteorological occurrence, we only provide general statistics on this episode without further analysis.

4.1.1. CAMx Evaluation

(1) Evaluation against CAMS sites: We first evaluated CAMx against land-based O₃ measurements from the CAMS sites. The time series of the daytime mean O₃ from simulations and observations from CAMS sites are displayed in **Figure 12**, and the evaluation statistics are summarized in **Table 9**. The model captures the onshore O₃ variability ($R=0.79$) with an overall overestimation of 7.89 ppb (20%), mainly due to the high positive bias of 10.93 ppb (34%) on clean days. This result is comparable with the model performance from previous studies focusing on the same area (e.g., Xiao et al., 2010; Pan et al., 2015; Kommalapati et al., 2016), which further verifies the reliability of our model settings.

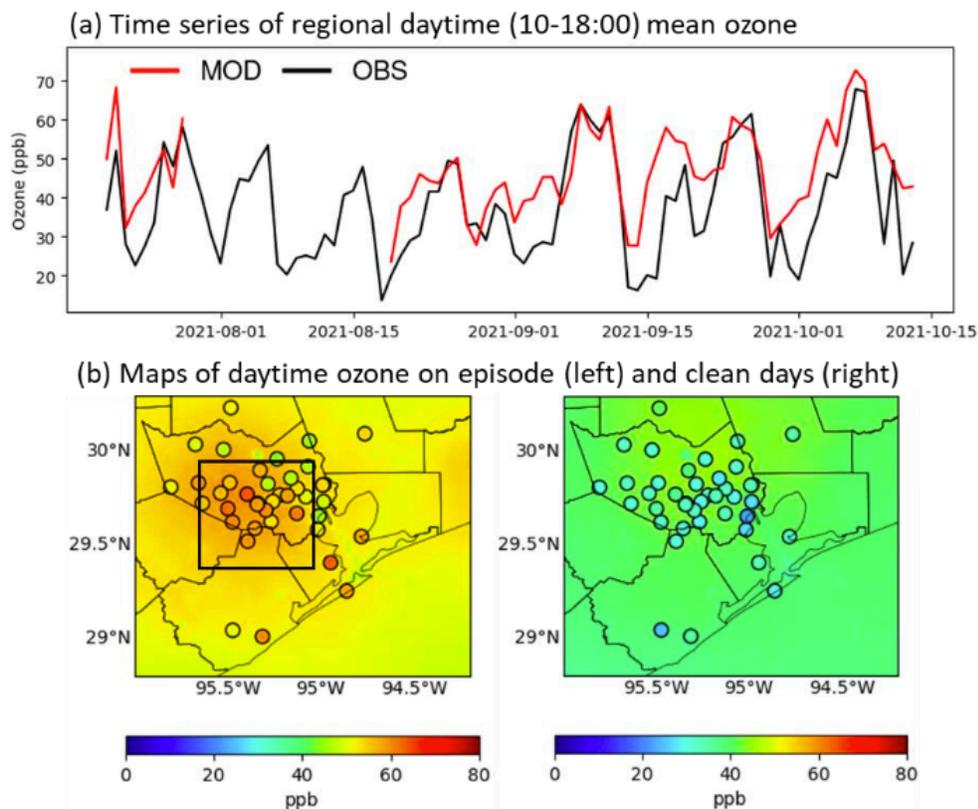


Figure 12. (a) Time series of daytime (10:00 – 18:00) mean ozone for observations at CAMS sites (OBS; black line) and simulations (MOD; red line). (b) Maps of observed (points) and simulated (background) daytime ozone during ozone episodes (left) and clean days (right).

Table 9. Daytime (10:00 – 18:00) ozone evaluation metrics for CAMx model at CAMS sites, including the observed and simulated mean values, correlation coefficient (R), mean bias (MB), normalized mean bias (NMB), mean absolute error (MAE), and root mean square error (RMSE).

Sites	Period	Observed mean (ppb)	Simulated mean (ppb)	R	MB (ppb)	NMB (%)	MAE (ppb)	RMSE (ppb)
CAMS	all days	38.87	46.76	0.79	7.89	20.32	9.41	11.72
	ozone episode	54.63	56.17	0.64	1.54	2.81	5.31	7.15
	clean days	31.34	42.28	0.64	10.93	34.88	11.35	13.37

(2) Evaluation against offshore measurements: The time series of the daytime (10:00 – 18:00) mean O₃ at the three boats are shown in **Figure 13a**, and the evaluation statistics are listed in **Table 10**. The evaluation does not include nighttime data to reduce the effects from land as the boats stayed at the dock at night. We also tested hourly time series evaluations (not shown) and it shows a larger bias between modeled ozone and boat observations overnight than during the daytime. The spatiotemporal variability of daytime O₃ at the three boats is well captured by the model with a correlation coefficient (R) value greater than 0.70. Overall, the model overestimates daytime O₃ by 4.57 ppb (11%), 7.82 ppb (22%), and 4.35 ppb (9%) for the pontoon boat, red eagle, and shrimp boat, respectively. On episode days, high O₃ mixing ratios can be found over Galveston Bay and the Gulf of Mexico (Figure 2b). The model captures some of the variability (R=0.42 – 0.51), with negative mean bias (MB) values of ~4.5 ppb (8%) for the pontoon and shrimp boats and a nearly unbiased simulation (MB=0.05 ppb) for the Red Eagle boat. While the O₃ variability is better predicted on clean days (R=0.69 – 0.76), the model shows higher values of MB than those on high-O₃ days ranging from 9.15 ppb (29%) to 11.28 ppb (41%), which drives the overall model overestimation.

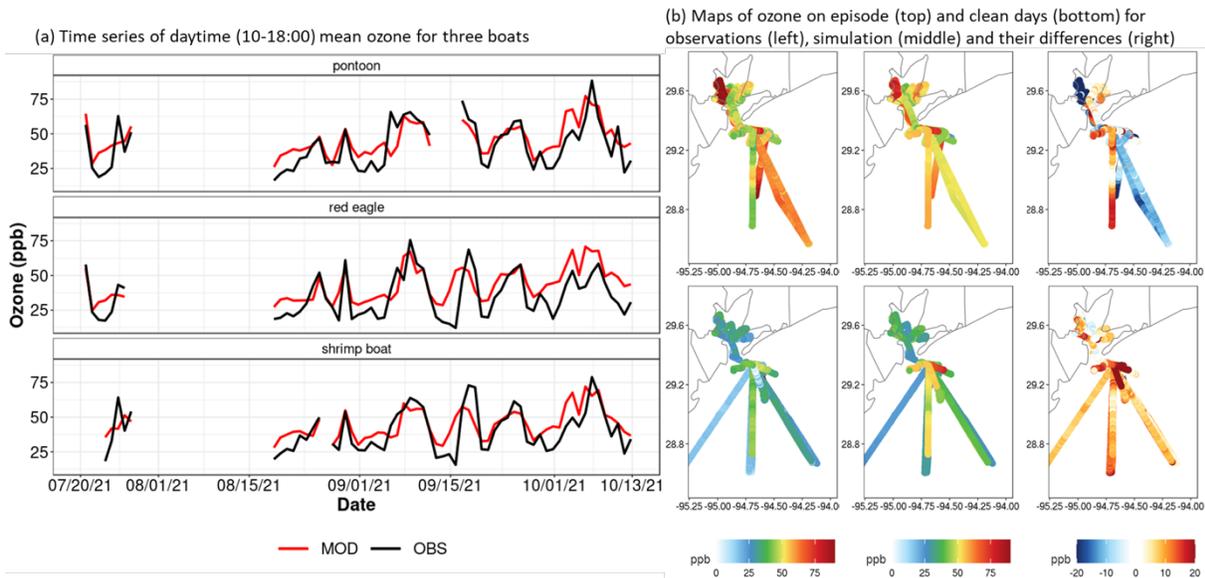


Figure 13. (a) Time series of daytime (10:00 – 18:00) mean ozone for observations at three boats (black) and simulations (red). (b) Maps of observed (left column), simulated (middle column), and their difference (right column) of ozone during ozone episodes (top row) and clean days (bottom row).

Table 10. Daytime (10:00 – 18:00) ozone evaluation metrics for CAMx model against offshore observations collected by three boats.

Boat	Period	Observed mean (ppb)	Simulated mean (ppb)	R	MB (ppb)	NMB (%)	MAE (ppb)	RMSE (ppb)
pontoon	all days	41.18	45.76	0.77	4.57	11.12	9.75	11.57
	ozone episode	58.57	54.21	0.51	-4.36	-7.44	8.34	11.31
	clean days	32.06	41.33	0.76	9.27	28.93	10.50	11.71
red eagle	all days	34.86	42.69	0.71	7.82	22.45	11.15	13.42
	ozone episode	51.20	51.25	0.42	0.05	0.08	9.71	11.92
	clean days	27.60	38.88	0.69	11.28	40.89	11.80	14.03
shrimp boat	all days	39.99	44.35	0.73	4.35	10.89	9.15	11.47
	ozone episode	57.22	52.22	0.43	-5.00	-8.74	8.88	11.65
	clean days	31.17	40.32	0.69	9.15	29.36	9.28	11.38

(3) Evaluation against mobile observations: Figure 14 shows the spatial distributions of O_x ($O_3 + NO_2$) observed by the mobile lab and modeled by CAMx over the six high-ozone episodes and on clean days. Here we use O_x , instead of O_3 , to account for the fast photochemical reactions of O_3 and NO_x taking place in fresh emissions along the roads. The mobile lab measurements during the six episodes and on clean days were directly shown without averaging, as the latitude and longitude of the mobile lab were fast changing and thus not averageable over a period of time (e.g., one hour). On clean days, the model shows an overestimation of daytime O_x at most of the locations with a mean bias value of 9.50 ppb. Over the high-ozone periods, there are low to middle levels of O_x near the urban center, which is overestimated by the model. The model underestimates high levels of O_x in the western part of the region. Overall, the model overestimates O_x by 3.35 ppb on ozone episode days.

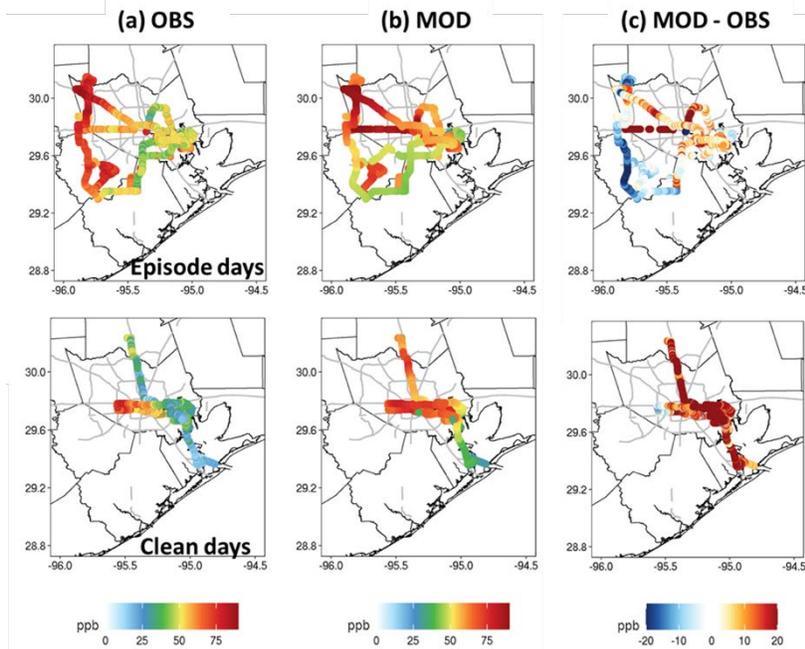


Figure 14. The mixing ratio of O_x ($O_3 + NO_2$) observed by the mobile lab (OBS; a), simulated by CAMx (MOD; b), and O_x differences between the simulation and observations (c) on days with high ozone in the six episodes (first row) and clean days (second row).

4.1.2. WRF-GC Evaluation

The following figures and tables present the evaluation of the WRF-GC model with the same in situ O_3 observations as in **Section 4.1.1** but for September 2021 only. The WRF-GC model has similar correlation coefficients as CAMx, but a significantly higher positive bias for all platforms. Throughout September, WRF-GC has the largest positive biases during several clean days of September 1-5 and the ozone episode of September 17-19 (associated with the passing of Hurricane Nicolas), which dominates the overall ozone overestimation in WRF-GC. The model failure in simulating this unique ozone episode indicates a misrepresentation of meteorology or transport patterns associated with the passing of Hurricane Nicolas.

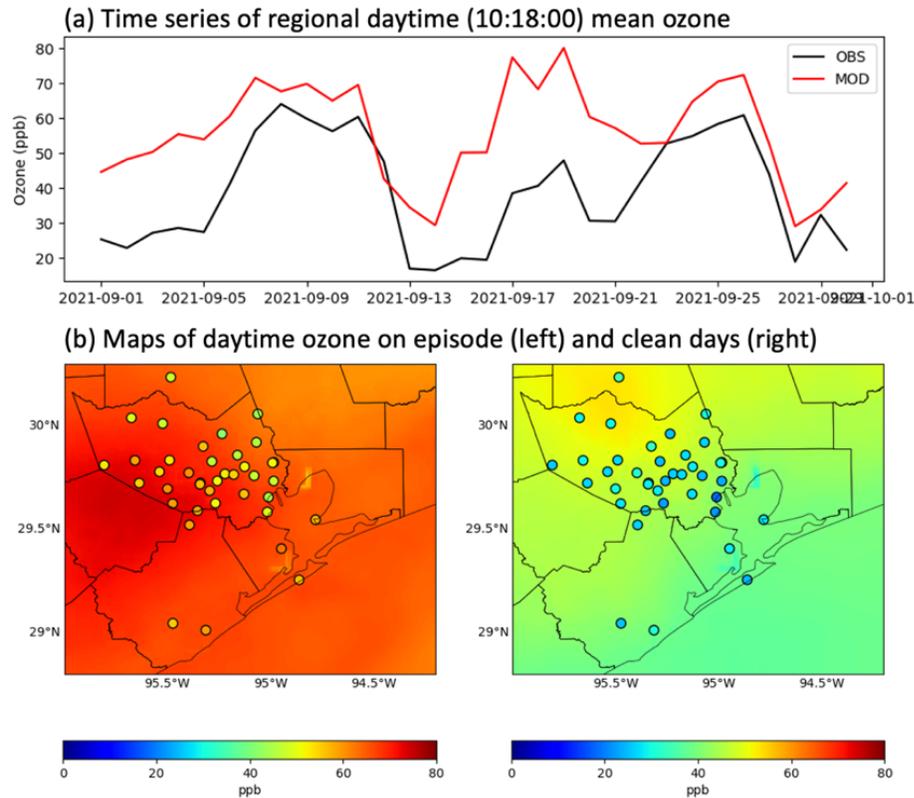


Figure 15. (a) Time series of daytime (10:00 – 18:00) mean ozone for observations at CAMS sites (OBS; black line) and WRF-GC simulations (MOD; red line). (b) Maps of observed (points) and simulated (background) daytime ozone during ozone episodes (left) and clean days (right).

Table 11. Daytime (10:00 – 18:00) ozone evaluation metrics for WRF-GC model at CAMS sites. For the second row, metrics during September 6-11 and September 23-26 (excluding September 17-19) are shown outside of parenthesis while metrics during all three September episodes are shown in parenthesis.

Sites	Period	Observed mean (ppb)	Simulated mean (ppb)	R	MB (ppb)	NMB (%)	MAE (ppb)	RMSE (ppb)
CAMS	all days	38.83	55.93	0.74	17.10	44	17.43	20.15
	ozone episode	56.54 (53.27)	66.51 (68.54)	0.62 (-0.07)	9.97 (15.27)	18 (29)	9.97 (15.27)	11.20 (18.76)
	clean days	27.79	46.29	0.39	18.50	67	19.09	21.15

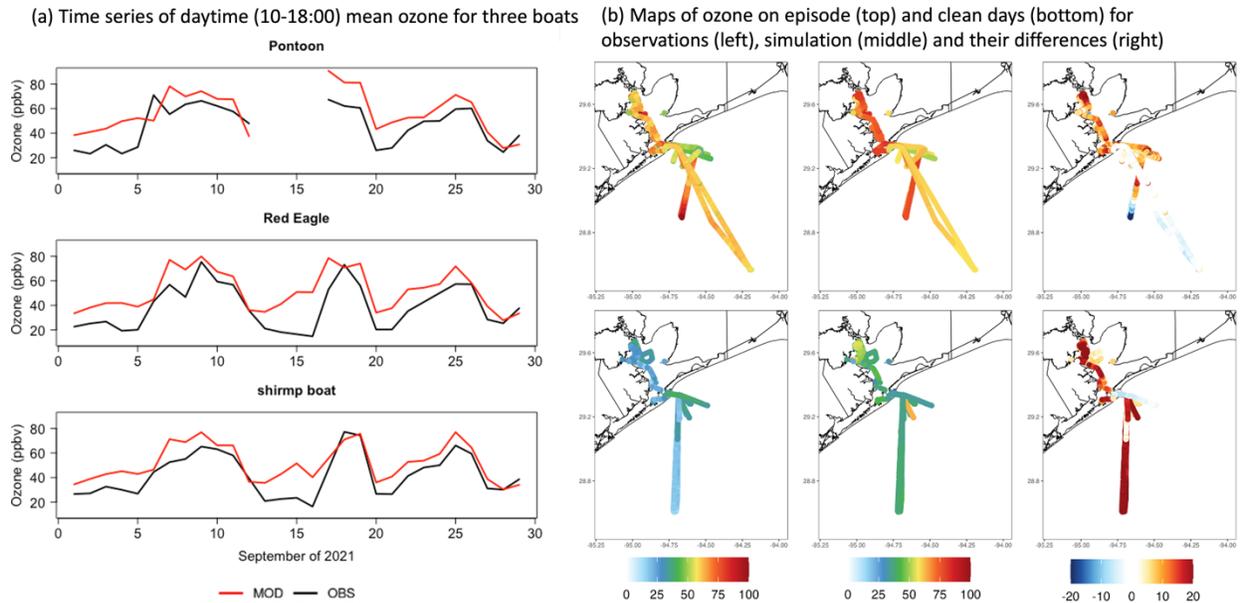


Figure 16. (a) Time series of daytime (10:00 – 18:00) mean ozone for observations at three boats (black) and WRF-GC simulations (red). (b) Maps of observed (left column), simulated (middle column), and their difference (right column) of ozone during ozone episodes (top row) and clean days (bottom row).

Table 12. Daytime (10:00 – 18:00) ozone evaluation metrics for WRF-GC model against offshore observations collected by three boats.

Boat	Period	Observed mean (ppb)	Simulated mean (ppb)	R	MB (ppb)	NMB (%)	MAE (ppb)	RMSE (ppb)
pontoon	all days	46.31	56.76	0.78	10.45	23	13.53	15.23
	ozone episode	60.43	70.18	0.25	9.75	16	12.95	14.78
	clean days	31.01	42.22	0.05	11.21	36	14.16	15.71
red eagle	all days	38.46	51.73	0.83	13.27	34	13.72	16.55
	ozone episode	55.99	66.72	0.62	10.73	19	11.1	13.69
	clean days	24.22	39.55	0.19	15.33	63	15.86	18.55
shrimp boat	all days	41.91	51.45	0.89	9.54	23	10.51	12.46
	ozone episode	59.49	66.44	0.78	6.95	12	7.99	9.40
	clean days	28.73	40.21	0.06	11.48	40	12.39	14.34

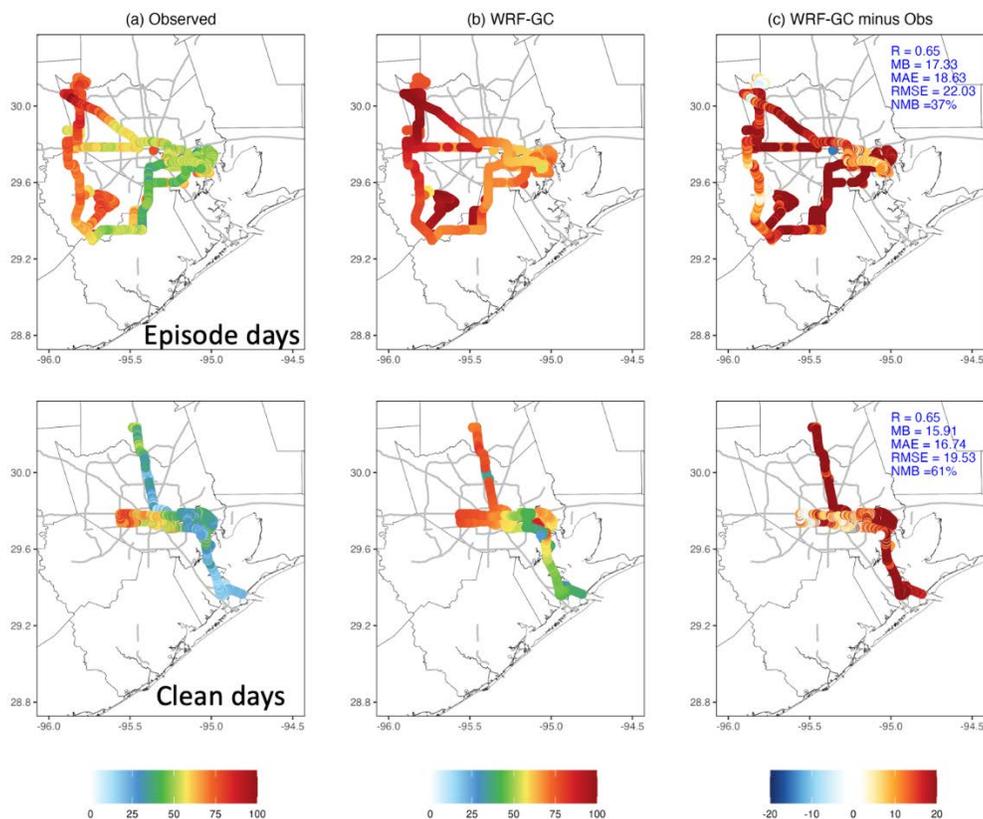


Figure 17. The mixing ratio of O_x ($O_3 + NO_2$) observed by the mobile lab (a) and simulated by WRF-GC (b), and O_x differences between the simulation and observations (c) on days with high ozone in the three episodes of September (first row) and clean days (second row).

4.1.3. WRF-Chem Evaluation

The following figures and tables present the evaluation of the WRF-Chem model with the same in situ ozone observations as in **Section 4.1.1** but for September 2021 only. Compared to the CAMS sites, WRF-Chem tends to underestimate surface ozone (**Figure 18**). The lower ozone concentration simulated by WRF-Chem during episode 4, Sep 17-19, slightly dominates the overall low ozone bias. Note this ozone episode is caused by the passing of Hurricane Nicolas. After removing this unique ozone episode from the evaluation statistics, the WRF-Chem model has a higher correlation coefficient than the other two models and higher average ozone levels than when episode 4 is included.

This same pattern is more profound in the boat comparisons (**Figure 19**). The WRF-Chem model is fairly able to reproduce higher ozone over the Gulf and Bay but notably underestimates ozone during Sep. 17-19 while the other two models are better able to capture this episode. Since the initial meteorology conditions used for the WRF-Chem simulation are the same as used by the other two models (HRRR), the underestimation of ozone during this episode could be due to a misrepresentation of the transport during this event or the meteorological model dynamics specific to the WRF-Chem simulation. Removing this unique ozone episode from the evaluation statistics, the WRF-Chem model has similar correlation coefficients as the other two models and much better results than when the episode is included.

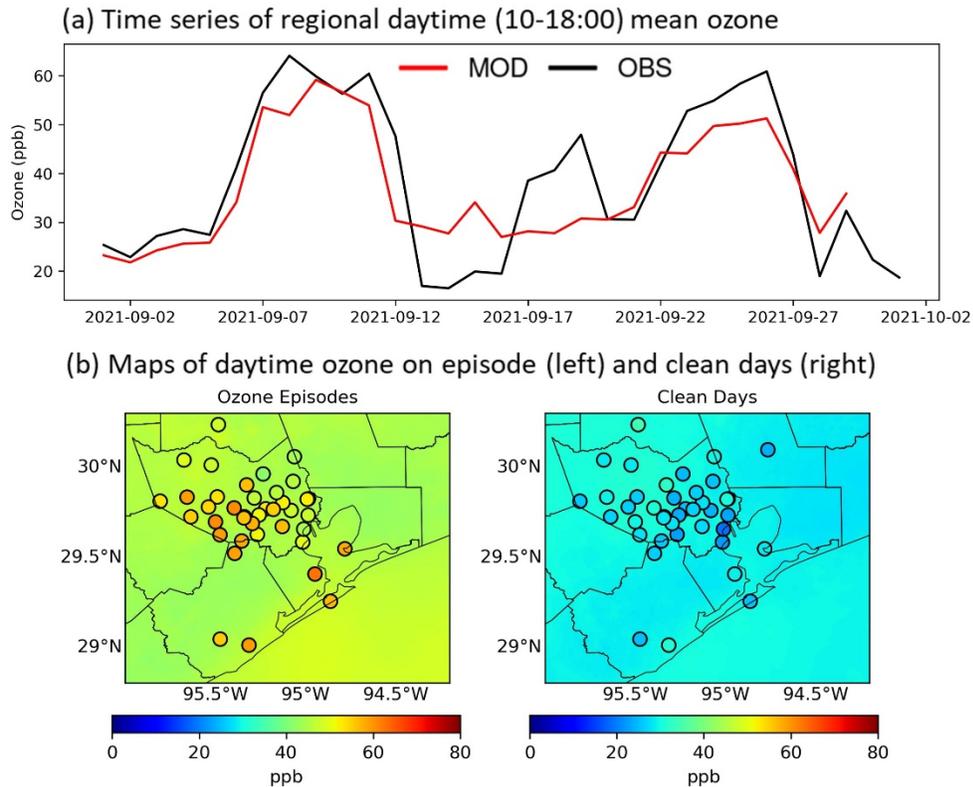


Figure 18. (a) Time series of daytime (10:00 – 18:00) mean ozone for observations at CAMS sites (OBS; black line) and WRF-Chem simulation (MOD; red line). (b) Maps of observed (points) and simulated (background) daytime ozone during ozone episodes (left) and clean days.

Table 13. Daytime (10:00 – 18:00) ozone evaluation metrics for WRF-Chem model at CAMS sites. For the second row, metrics during September 6-11 and September 23-26 (excluding September 17-19) are shown outside of parenthesis while metrics during all three September episodes are shown in parenthesis.

Sites	Period	Observed mean (ppb)	Simulated mean (ppb)	R	MB (ppb)	NMB (%)	MAE (ppb)	RMSE (ppb)
CAMS	all days	39.40	36.99	0.86	-2.41	-6.11	6.74	8.41
	ozone episode	53.27 (56.54)	45.49 (50.47)	0.91 (0.83)	-7.78 (-6.07)	-14.61 (-10.74)	7.83 (6.14)	9.11 (7.15)
	clean days	28.13	30.09	0.59	1.96	6.97	5.85	7.79

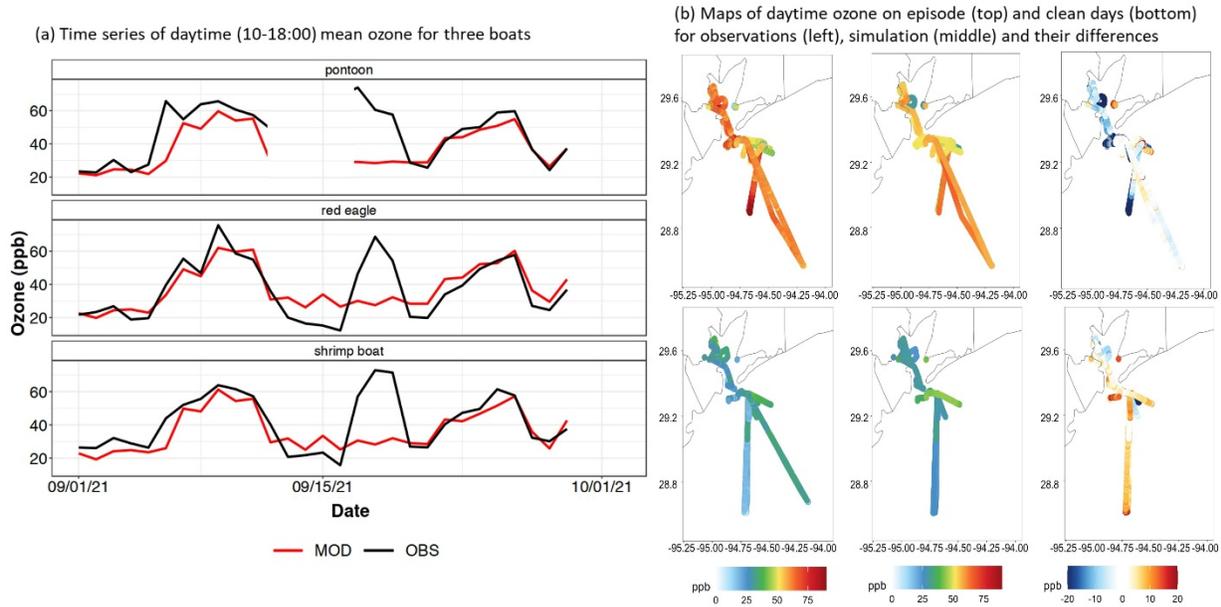


Figure 19. (a) Time series of daytime (10:00 – 18:00) mean ozone for observations at three boats (black) and WRF-Chem simulation (red). (b) Maps of observed (left column), simulated (middle column), and their difference (right column) of ozone during ozone episodes (top row) and clean days (bottom row).

Table 14. Daytime (10:00 – 18:00) ozone evaluation metrics for WRF-Chem model against offshore observations collected by three boats. The metrics are the same as in Table 2. Correlations derived in parenthesis exclude results from September 17-19.

Boat	Period	Observed mean (ppb)	Simulated mean (ppb)	R	MB (ppb)	NMB (%)	MAE (ppb)	RMSE (ppb)
pontoon	all days	45.90	37.19	0.62	-8.71	-18.97	9.40	15.53
	ozone episode	59.80	45.02	-0.30 (0.71)	-14.78	-24.72	14.78	20.67
	clean days	30.84	28.71	0.70	-2.13	-6.89	3.57	6.31
red eagle	all days	37.01	37.30	0.73	0.29	0.78	8.56	11.82
	ozone episode	53.91	46.85	0.33 (0.79)	-7.06	-13.09	9.69	14.65
	clean days	23.29	29.55	0.58	6.26	26.88	7.64	8.87
shrimp boat	all days	41.61	36.13	0.64	5.48	-13.17	8.90	13.79
	ozone episode	57.83	44.89	-0.02 (0.85)	-12.94	-22.38	12.94	19.32
	clean days	28.43	29.01	0.54	0.58	2.05	5.61	6.43

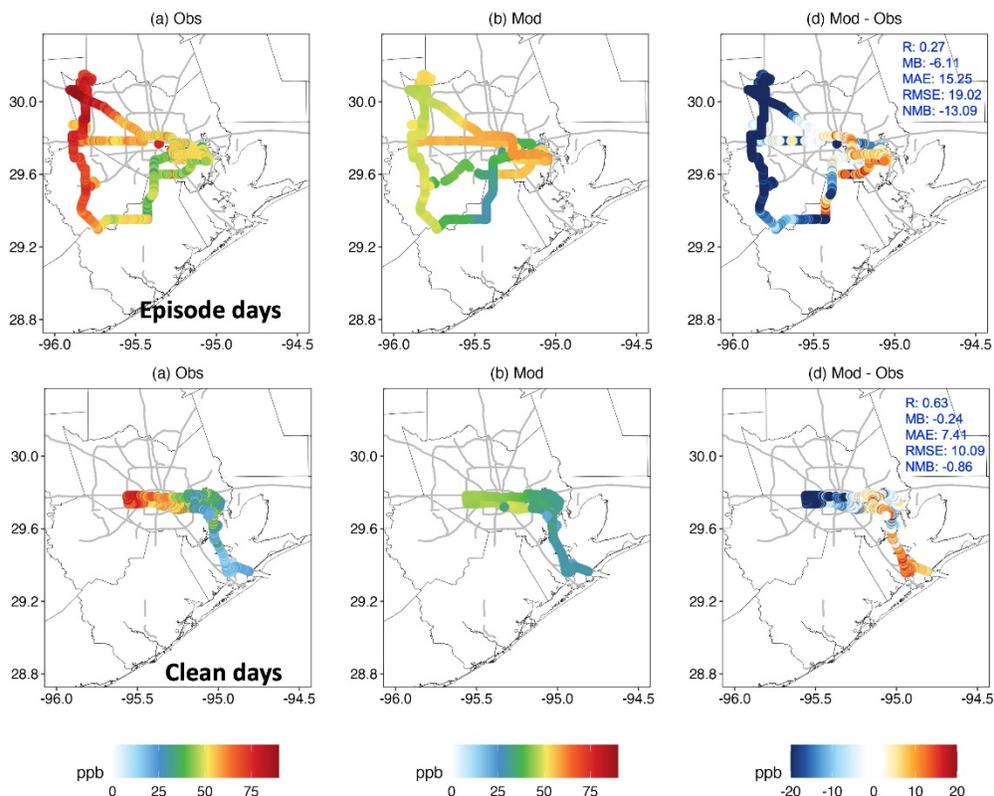


Figure 20. The mixing ratio of O_x ($O_3 + NO_2$) observed by the mobile lab (a) and simulated by WRF-Chem (b), and O_x differences between the simulation and observations (c) on days with high ozone in the three episodes of September (first row) and clean days (second row).

4.2 Model Intercomparison

4.2.1. Mean state of ozone and precursors

The previous section evaluated individual models against TRACER-AQ observations. Since observations are only snapshots, it is necessary to conduct model intercomparison beyond the observation location and time. We first compared the mean state of ozone, nitrogen dioxide, carbon monoxide, and formaldehyde over the surface and the vertical profile of the inner-most model domain (d03) during September 2021. **Figure 21** contains the mean surface concentrations of ozone, nitrogen dioxide, carbon monoxide, and formaldehyde simulated from 3 different models, WRF-GC, WRF-Chem, and CAMx. The values for the model were calculated by taking the daily average at each grid point, and then averaging that value over the whole September 2021. Observation values were calculated using the same method. Some notable features were the anomalies for ozone and nitrogen dioxide seen near Clear Lake for WRF-GC, issues with formaldehyde over the gulf for WRF-GC, issues with gridding for nitrogen dioxide over the gulf for WRF-Chem, and a possible issue with the initiation of formaldehyde for WRF-Chem. All these issues will be investigated and corrected when necessary in subsequent work of the project. WRF-GC had the highest ozone bias of the three and WRF-Chem had the lowest bias.

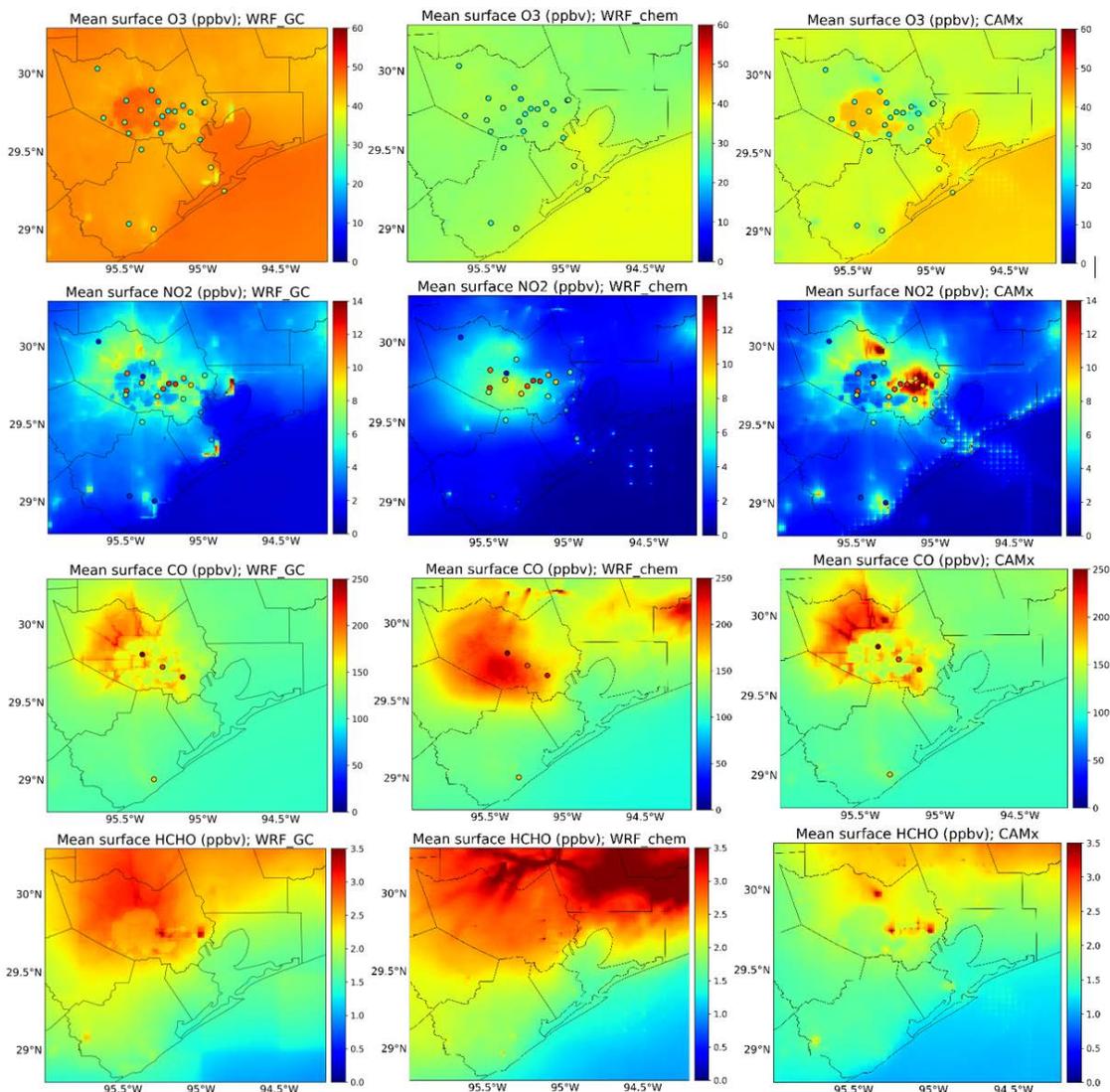


Figure 21. Domain (d03)-mean surface concentrations of (a) ozone, (b) NO₂, (c) CO, and (d) formaldehyde (HCHO) simulated by WRF-GC, WRF-chem, and CAMx over September 2021. CAMS site observations of O₃, NO₂, and CO (points) are overlaid in (a), (b), (c), respectively.

Figure 22 contains time series of site-averaged surface ozone as well as domain-averaged surface ozone. Areas highlighted in yellow indicate high ozone periods in September 2021. In the whole d03 domain, the temporal variability of surface ozone, such as the overall ozone diurnal cycle and ozone increase in the ozone episodes, is consistently simulated across three models (**Figure 22b**). However, there are some differences between models in detail. For example, WRF-Chem does not simulate ozone increase in the second ozone episode (Sep 17-19) as described in **Section 4.1.3**. The detailed model intercomparison on the ozone diurnal cycle can be found in the next section. The mean ozone concentration for the whole September is also different in each model with the highest value of 46.08 ppbv in WRF-GC, 37.65 ppbv in CAMx, and 33.97 ppbv in WRF-Chem. The standard deviation is also different in each model with 14.55 ppbv in WRF-GC, 10.34 ppbv in CAMx, and 10.87 ppbv in WRF-Chem.

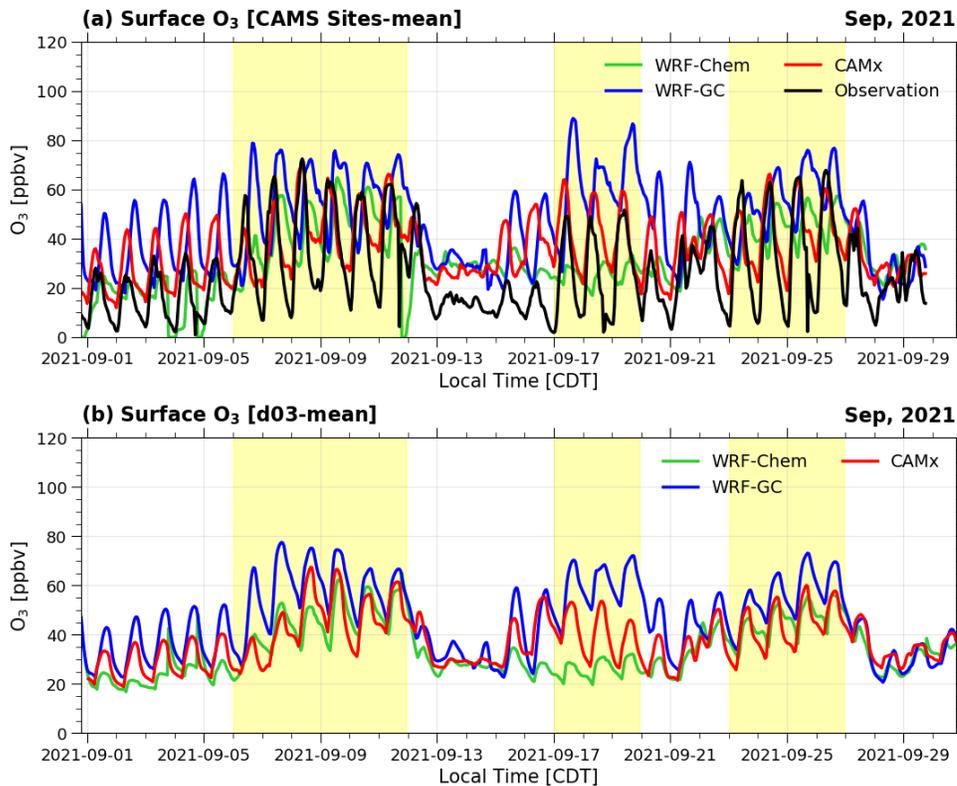


Figure 22. (a) Observed (black) and simulated surface ozone time series averaged over all CAMS sites in the d03 domain. (b) Domain (d03)-mean surface ozone time series simulated by each model. The yellow highlights indicate the three episodes in September 2021.

Figure 23 contains domain mean vertical profiles of ozone, nitrogen dioxide, carbon monoxide, and formaldehyde. The models are more consistent with each other above 850 hPa for O_3 , but not for other species shown. As long-range transported fire emissions and lightning can significantly impact the primary species above the boundary layer, the differences in CO, NO_x , and HCHO above 850 hPa indicate the effects of different fire emissions and lightning NO_x emissions used in each model, as shown in **Table 1**. One thing to note is the reduction of vertical resolution for CAMx.

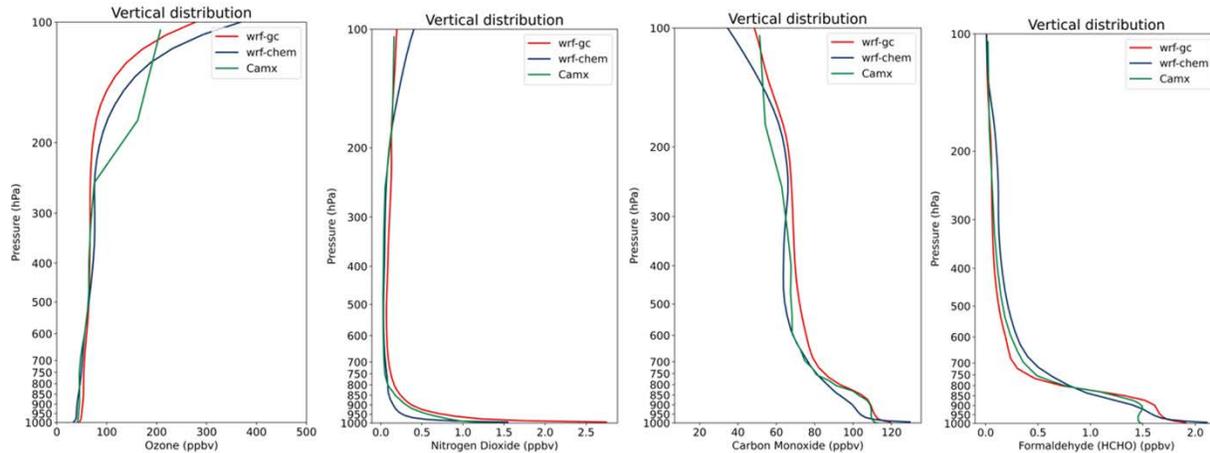


Figure 23. Domain (d03)-mean vertical profiles of O₃, NO₂, CO, and formaldehyde (HCHO) simulated by three models.

Figure 24 shows the September-mean spatial ozone anomalies in each model. The ozone anomalies in each grid are calculated by removing each model’s domain-mean ozone. All models show positive anomalies offshore and negative anomalies on land, which indicates that higher ozone and lower ozone than the domain-mean ozone are simulated offshore and on land, respectively. CAMx and WRF-GC show positive anomalies in the city of Houston (**Figures 24a** and **24b**), but this feature is not characterized in WRF-Chem (**Figure 24c**). CAMx also shows strong negative anomalies in the east and north of the city of Houston (**Figure 24a**), which might be associated with the high mean NO₂ concentration in these regions in CAMx (**Figure 21**).

Figure 25 shows the same spatial ozone anomalies but for the first and third ozone episodes. In the ozone episodes, positive anomalies are characterized over most of the d03 domain including land and offshore in all three models. In the first episode, all models show stronger positive anomalies offshore than land (**Figures 25a-c**), and strong positive anomalies are also captured in the city of Houston in CAMx (**Figure 25a**) and WRF-GC (**Figure 25b**). In WRF-GC, the strong positive anomalies offshore are characterized along the coast (**Figure 25b**), but the magnitude of positive anomalies increases from offshore to the Gulf of Mexico in WRF-Chem (**Figure 25c**). In the third episode (**Figures 25d-f**), the overall characteristics are similar, but the magnitude of positive anomalies offshore tends to be less than that in the first episodes in all three models. In WRF-Chem, the center of positive anomalies offshore is captured near Brazoria and Matagorda counties (**Figure 25f**). Note that negative anomalies in CAMx are still captured in the east and north of the city of Houston in these two ozone episodes (**Figures 25a** and **25d**).

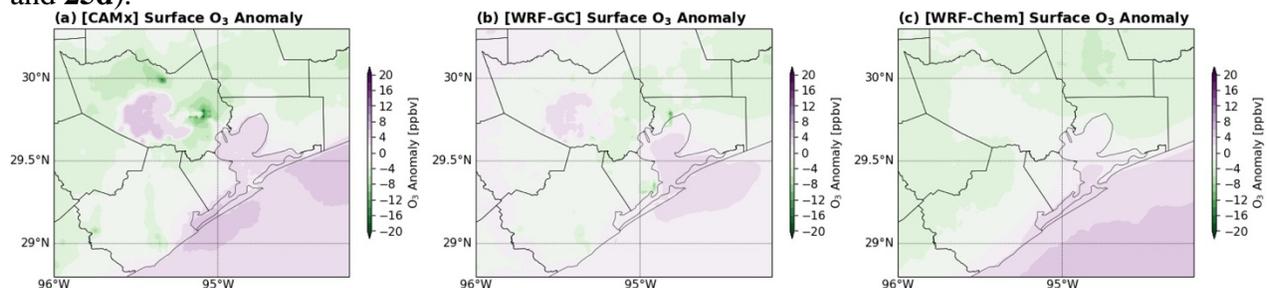


Figure 24. September-mean spatial ozone anomalies simulated by each model. The anomalies are calculated by removing each model’s domain-mean ozone from individual grids.

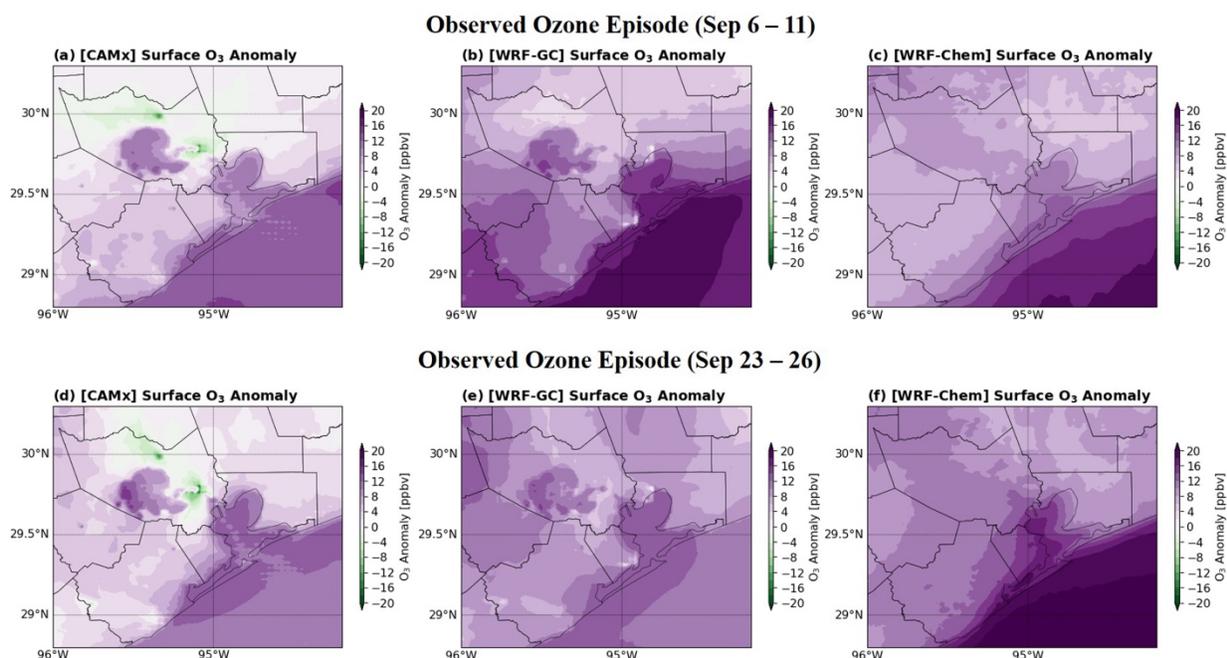


Figure 25. Spatial ozone anomalies in the first ozone episode (upper panel) and the third ozone episode (lower panel) by each model. The anomalies are calculated by removing each model’s September-mean and domain-mean ozone from each grid.

4.2.2. O₃ diurnal cycle

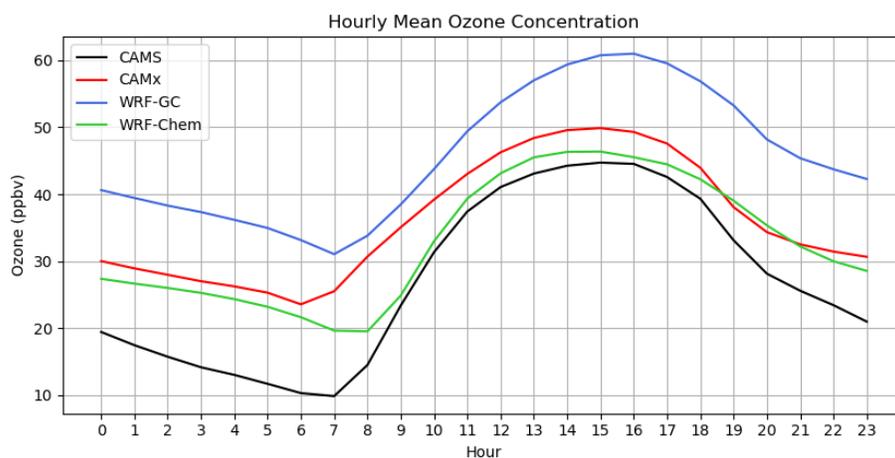


Figure 26. Diurnal variations in site-averaged surface ozone for observations (black), CAMx (red), WRF-GC (blue), and WRF-Chem (green) averaged over the September 2021 period.

Our model intercomparison analysis so far explored the spatial and temporal average patterns of chemical species over the study region. We further explore the average diurnal cycle of surface ozone to identify useful information such as the daily pattern of ozone accumulation, and how ozone measurements in the morning can predict modeled ozone behavior in the afternoon. The clearest diurnal cycle pattern can be visualized through a site-averaged average diurnal cycle plot for the September 2021 period. Each site-averaged diurnal cycle of CAMx,

WRF-GC, WRF-Chem, and CAMS observations sites, most of which are coastal or urban sites, are compared. Only surface ozone data from the latitude and longitude points nearest to CAMS observation sites are included in the diurnal time series (**Figure 26**). The first feature of notice is the overnight ozone pattern during the early morning hours (00:00-05:00). During this period is typically when ozone levels over the region are decreasing. The rate of ozone decrease is underdone by all of the models. Additionally, this is the time of day when model bias (modeled value - observed value) is greatest. Modeling finds it most difficult to predict the cleaner ozone environment observed at night. The next period between 06:00 and 09:00 is where modeled and observed ozone concentration minimums occur. Modeling is inconsistent here: CAMx predicts ozone production beginning between 06:00 and 07:00, WRF-GC between 07:00 and 08:00, and WRF-Chem between 08:00 and 09:00, with observations predicting ozone production after 07:00. As a result, each model has a rate of ozone production during the daytime that is different, and this rate increases the later in the morning the model indicates the ozone minimum occurring. The afternoon period (12:00-17:00) displays the leveling off of ozone production and ozone decreases across models and observations. Modeling continuously overpredicts ozone concentrations during the afternoon, consistent with morning ozone forecasts. The model biases are smaller though, and some of the modeling captures the cycling pattern seen in observations. CAMx and WRF-Chem follow the generally observed afternoon pattern, roughly capturing the ozone maximum and the periods of small positive and small negative rates. WRF-GC ozone maximum is translated later in the afternoon. While each model showed overprediction, the models' overprediction relative to each other was consistent throughout the diurnal cycle. In a general sense, the morning ozone forecast was predictive of the afternoon ozone forecast. Model performance generally increased throughout the day to about the time peak ozone levels were forecast, where CAMx and WRF-Chem both showed model biases of about 5 ppbv. WRF-GC did not have periods of observed ozone pattern emulation that the other models displayed.

The surface ozone data were further separated into an average of the morning hours (7:00-11:00 CDT) and afternoon hours (12:00-17:00 CDT) on the high ozone episode days identified earlier in the report, and clean ozone episode days. Each of the CAMx, WRF-GC, and WRF-Chem surface ozone outputs were compared against observed ground sites. The surface ozone model biases (modeled values – observed values) during the morning and afternoon hours on high and clean ozone episode days were calculated to capture the model performance of the diurnal ozone cycle. **Figure 27a** shows the scatter plot of morning versus afternoon model biases on the high ozone episode days, and **Figure 27b** shows the scatter plot of morning versus afternoon model biases on the clean ozone episode days. The relationship between morning and afternoon model bias during both ozone periods is generally positive and linear. WRF-GC displays the weakest of those relationships, with a slope of just 0.25 during high ozone days and 0.50 during clean ozone days. Its correlations during high ozone days are the lowest of all modeled periods, showing weak predictability of surface ozone from morning to afternoon. CAMx and WRF-Chem are well correlated with strong positive slopes on high ozone days, but WRF-Chem correlation weakens compared to CAMx on clean ozone days. CAMx maintains a strong positive slope on clean ozone days and a linear correlation above 0.85. Model bias tends to be positive for each model during the mornings, positive during clean ozone day afternoons, and positive or negative during high ozone day afternoons. Model performance of ozone bias varied greatly from model to model, but CAMx displayed the greatest consistency. We expect a high morning ozone bias to lead to higher afternoon ozone biases in each model, but CAMx displayed this characteristic clearly during both periods.

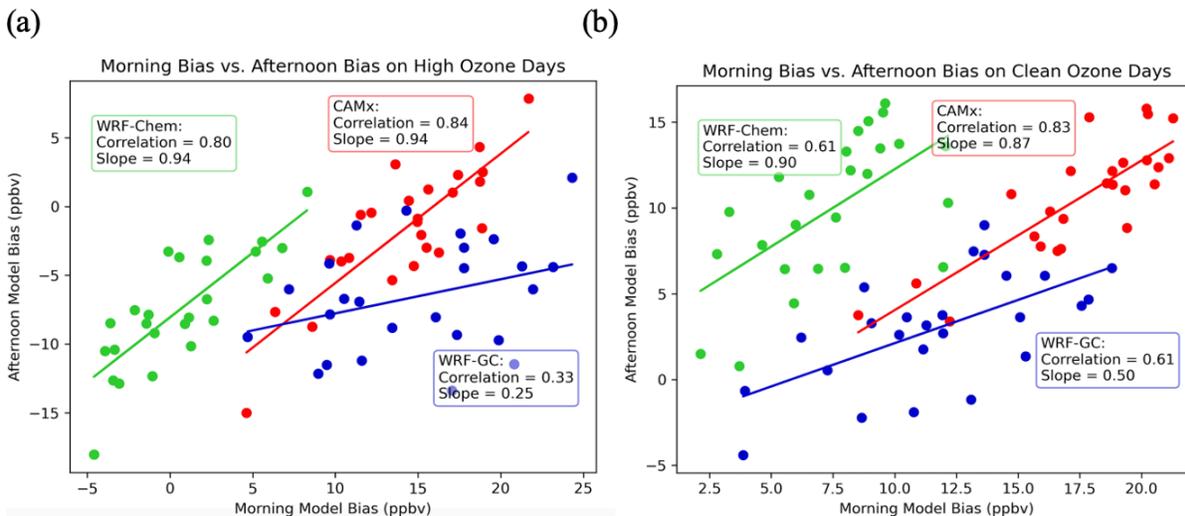


Figure 27. Scatter plot of the morning (7:00-11:00 CDT) vs. afternoon (12:00-17:00 CDT) surface ozone model biases, simulated by CAMx, WRF-GC, and WRF-Chem over September 2021 high ozone episode days (a) and clean days (b).

4.2.3. Column distribution of O₃ and precursors

This section first presents inter-model comparisons against column concentrations of HSRL-2 ozone and GCAS NO₂ and HCHO, and then compares the column-mean HCHO-to-NO₂ ratio (FNR) between observation and models. An ozone-episode day on September 9, 2021 is taken as an example of such comparisons. **Figure 28** shows the spatial distributions of the ozone column (0-1 km) from HSRL-2 and models in the morning (8-11.5 CDT), at noon (11.5-14.3 CDT), and in the afternoon (14.3-17.5 CDT). Land-water differences and daytime changes are the two aspects used for comparisons below. For land-water differences, the land has higher ozone than the water in the morning and at noon, while the water has higher ozone than the land in the afternoon. Such land-water differences are captured by WRF-Chem in the morning and are captured by CAMx and WRF-GC in the afternoon. At noon, land-water differences are captured by all three models. For daytime changes, in-land ozone increases from morning to noon and decreases from noon to afternoon; such an observed pattern is captured by all three models. Offshore ozone increases throughout all three periods; such an observed pattern is captured by WRF-GC and CAMx. Among the three models, WRF-GC shows the smallest bias under all conditions (land, water, morning, noon, and afternoon).

GCAS vertical column densities (0-8 km) of NO₂ (**Figure 29**) and HCHO (**Figure 30**) are shown for morning, noon, and afternoon. Three models underestimate NO₂ and HCHO to different extents. NO₂ peaks in the morning due to large traffic emissions and low boundary layer heights; such an observed pattern is captured by CAMx and WRF-GC. In contrast, daytime changes of HCHO are minor because secondary HCHO can be formed throughout the atmosphere via oxidation of VOCs in addition to primary emissions at the surface that are affected by diurnal changes of boundary layer heights. All three models show such a pattern with the smallest negative biases in WRF-Chem.

The FNR is widely used as an indicator of the ozone formation regime because the relative ambient concentrations of HCHO and NO₂ reflect the reactivity-weighted concentrations of VOCs and NO_x, respectively (Valin et al., 2016). A tropospheric column FNR less than 2.6 indicates a VOC-sensitive regime, a ratio between 2.6 to 3.5 indicates a transitional regime and a

ratio greater than 3.5 indicates a NO_x -sensitive regime; these values are specially derived for the Houston region according to relationships between multi satellite observations of ozone precursors and ground-based ozone observations (Jin et al., 2020). We use the above values to indicate different ozone formation regimes while simultaneously noting that these values should not be static in all scenarios. According to GCAS-derived FNR, southeastern Houston and the western coast of Galveston Bay are VOC-sensitive, other regions at the urban core are in a transitional regime and surrounding regions of Houston and the Bay are NO_x -sensitive. Among the three models, CAMx best capture such spatial patterns while the other two models both simulate most of the Houston region and adjacent Bay to be NO_x -sensitive (**Figure 31**). For daytime changes, the in-land region changes from the transitional regime in the morning to NO_x -sensitive in the afternoon, while the waterbody remains NO_x -sensitive throughout the day. Such observed trends are best captured by WRF-GC (**Figure 32**).

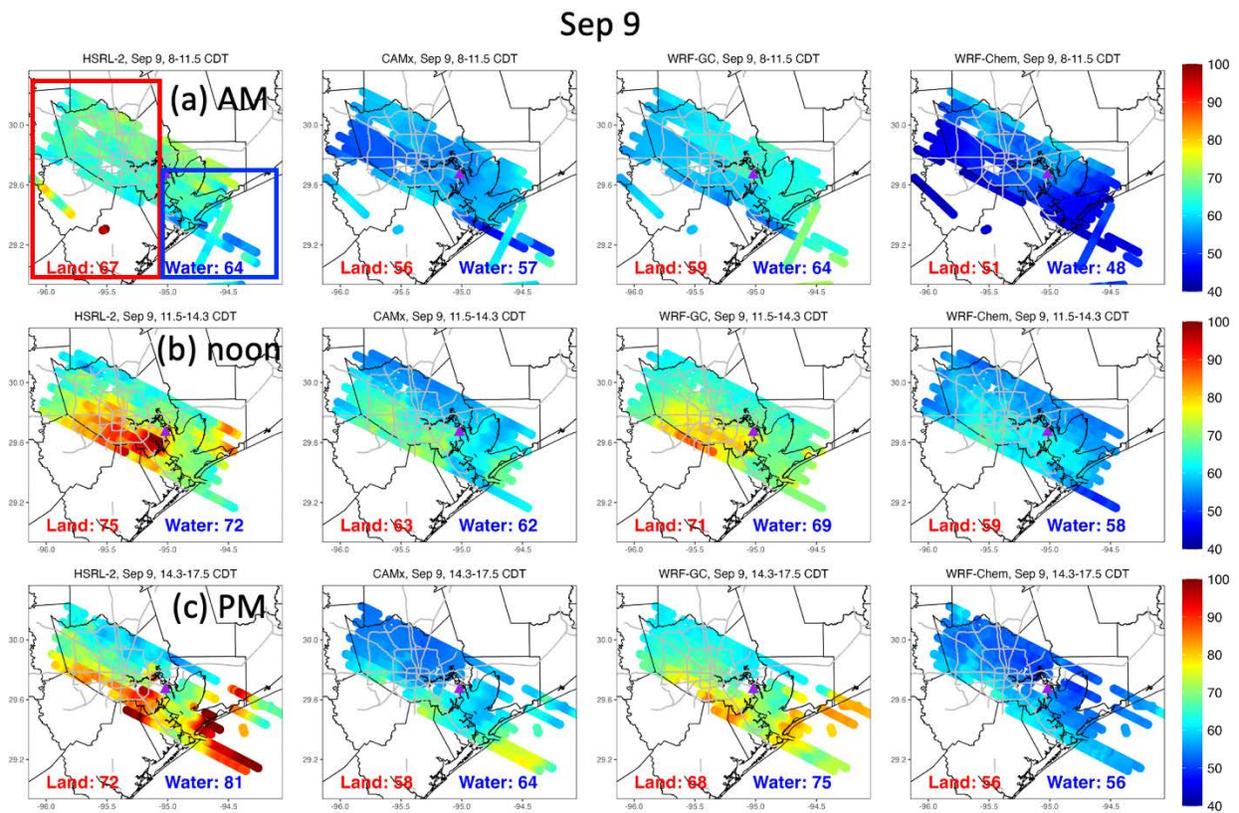


Figure 28. Spatial distributions of column (0-1 km) mean ozone from HSRL-2 measurements in comparison with CAMx, WRF-GC, and WRF-Chem models in the morning (8-11.5 CDT), at noon (11.5-14.3 CDT) and in the afternoon (14.3-17.5 CDT) on September 9, 2021. Red and blue boxes indicate land and water, respectively. Mean concentrations (ppbv) are inserted.

Sep 9

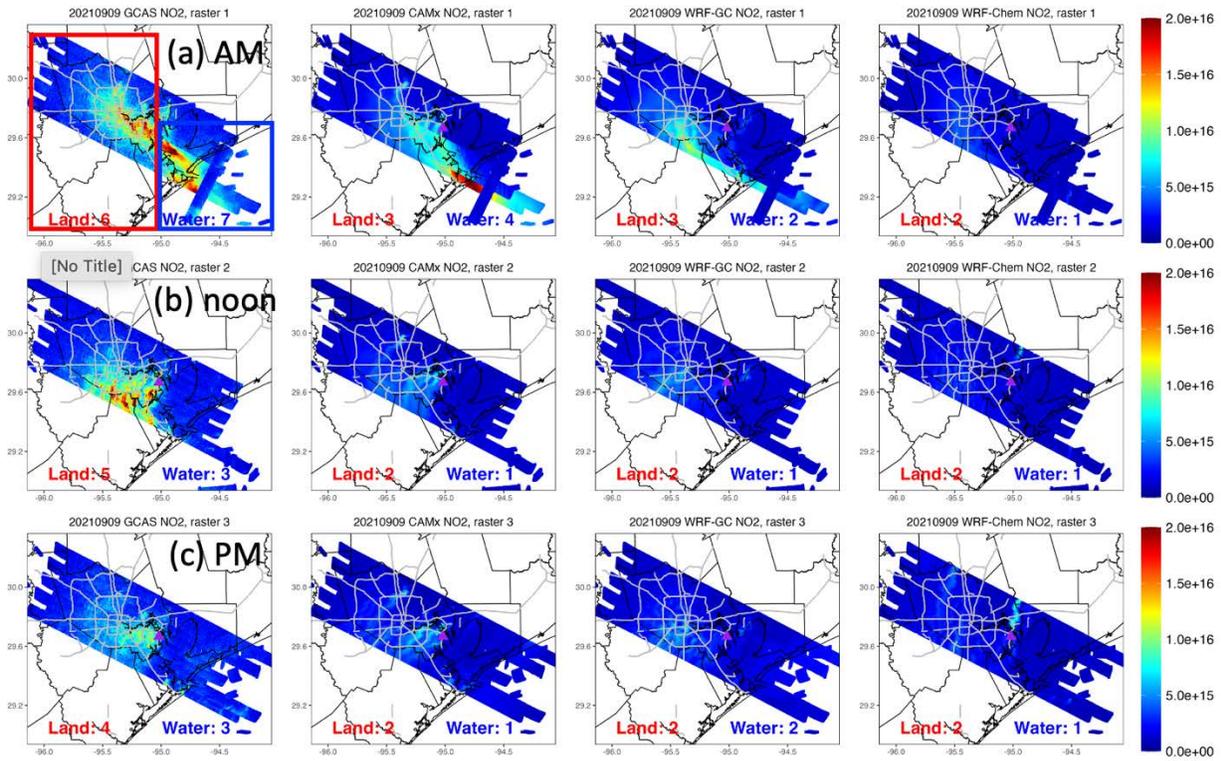


Figure 29. Spatial distributions of column (0-8 km) mean NO₂ from GCAS measurements in comparison with CAMx, WRF-GC, and WRF-Chem models in the morning (8-11.5 CDT), at noon (11.5-14.3 CDT) and in the afternoon (14.3-17.5 CDT) on September 9, 2021. Red and blue boxes indicate land and water, respectively. Mean concentrations ($\times 10^{15} \text{ molec cm}^{-2}$) are inserted.

Sep 9

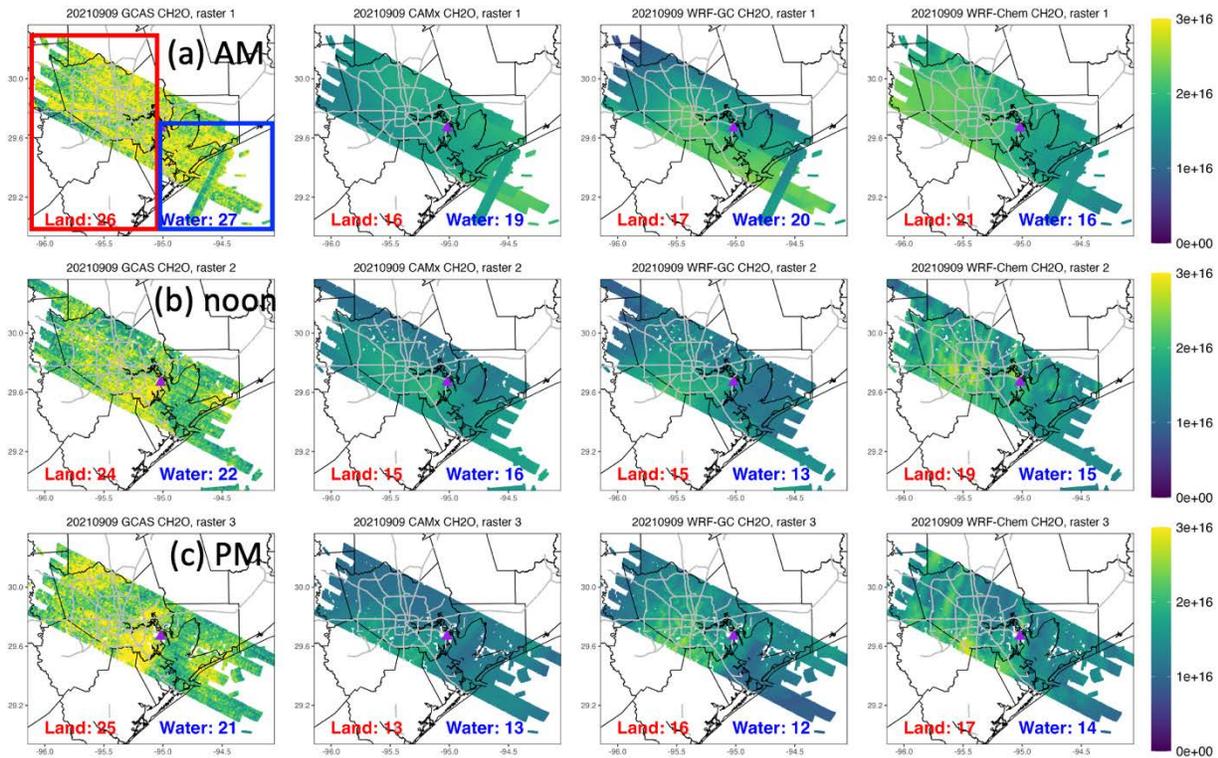


Figure 30. Spatial distributions of column (0-8 km) mean HCHO from GCAS measurements in comparison with CAMx, WRF-GC, and WRF-Chem models in the morning (8-11.5 CDT), at noon (11.5-14.3 CDT) and in the afternoon (14.3-17.5 CDT) on September 9, 2021. Red and blue boxes indicate land and water, respectively. Mean concentrations ($\times 10^{15} \text{ molec cm}^{-2}$) are inserted.

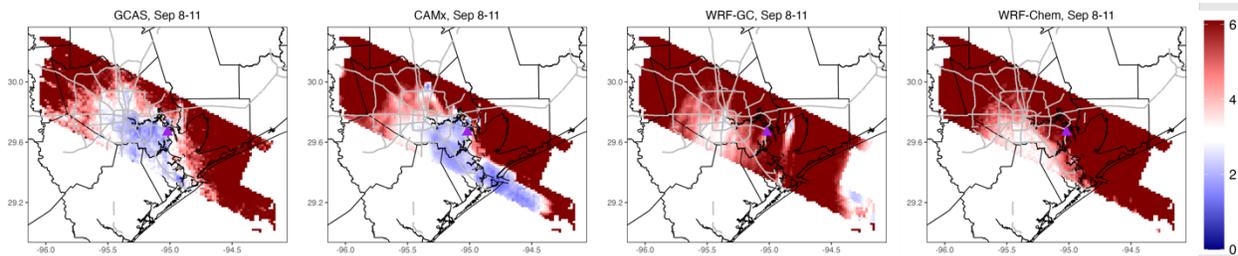


Figure 31. Mean column (0-8 km) formaldehyde-to-NO₂ ratio (FNR) observed by GCAS and simulated by CAMx, WRF-GC, and WRF-Chem over the ozone episode of September 8-11, 2021. The FNR thresholds defining ozone regimes, i.e., VOC-sensitive (FNR < 2.6; blue), transitional (2.6 ≤ FNR ≤ 3.5), and NO_x-sensitive (FNR > 3.5) regimes, are taken from Jin et al. (2020).

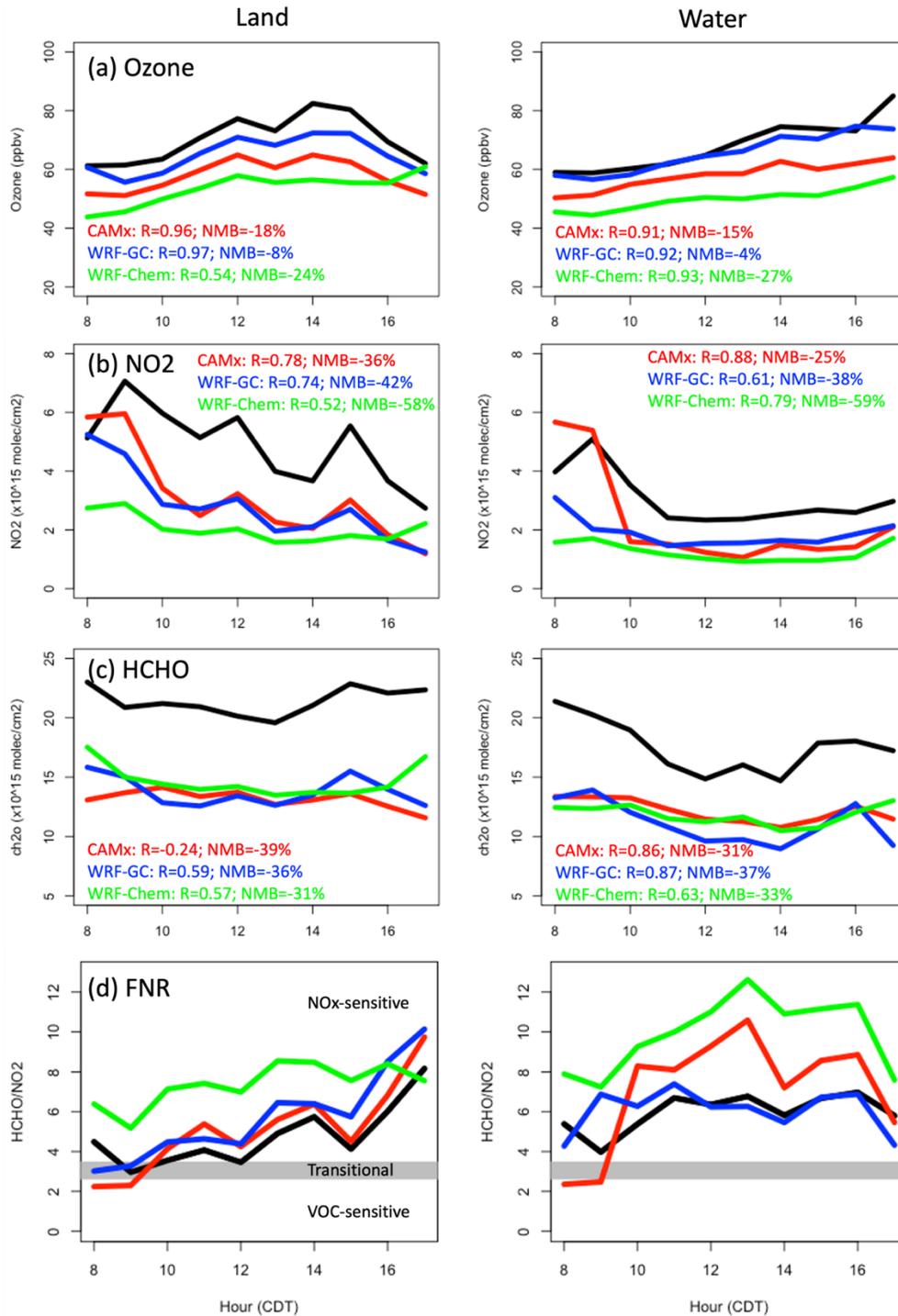


Figure 32. Diurnal variations for (a) ozone, (b) NO₂, (3) formaldehyde (HCHO) and (d) formaldehyde-to-NO₂ ratio (FNR) from observations (black), CAMx (red), WRF-GC (blue), and WRF-Chem (green) averaged over land and water during the ozone episode of September 8-11, 2021. In (a), ozone observations are from HSRL-2 column concentrations (0-1 km). In (b) and (c), NO₂ and HCHO observations are from GCAS column concentrations (0-8 km). In (d), ozone transitional regime is defined with values specially derived for Houston ($2.6 \leq \text{FNR} \leq 3.5$) in Jin et al. (2020).

4.2.4. Vertical distribution of ozone

In this section, the modeled ozone curtain profiles for September 6-11 and September 23-26 were evaluated against the lidar-derived ozone curtain profiles from the LMOL and GSFC ozone lidars. During the study period, the two lidars were stationed at two different sites around the region. The LMOL lidar was stationed at the University of Houston (UH) and the GSFC lidar at La Porte. The full curtain profiles for September 6-11 and September 23-26 are shown in **Figure 33** and **Figure 34**. During the September 6-11 episode, WRF-GC does best to capture the consistently high ozone concentrations in the lower troposphere (0-2000 m). WRF-Chem consistently underestimates the higher ozone near the surface during the 3rd episode. CAMx does best capturing the highest measured ozone event day (September 8) during episode 3. Both CAMx and WRF-GC underestimate the high ozone captured aloft (> 2500 m) later in the episode (after September 9) but WRF-Chem does well to capture this event, most notably at the La Porte site. For the September 23-26 episode, the results vary depending on which station, UH or La Porte. CAMx and WRF-GC do well again to capture the higher ozone near the surface at the UH site. Again, CAMx does better than both models to capture the highest surface ozone day (September 24) at the UH site. WRF-GC also simulates the higher surface ozone at the La Porte site well but CAMx underestimates ozone at this site. WRF-Chem captures some of the higher ozone near the surface towards the end of the episode at both sites but with underestimations. WRF-Chem also does better during this episode to capture the higher ozone aloft while WRF-GC and CAMx underestimate ozone aloft.

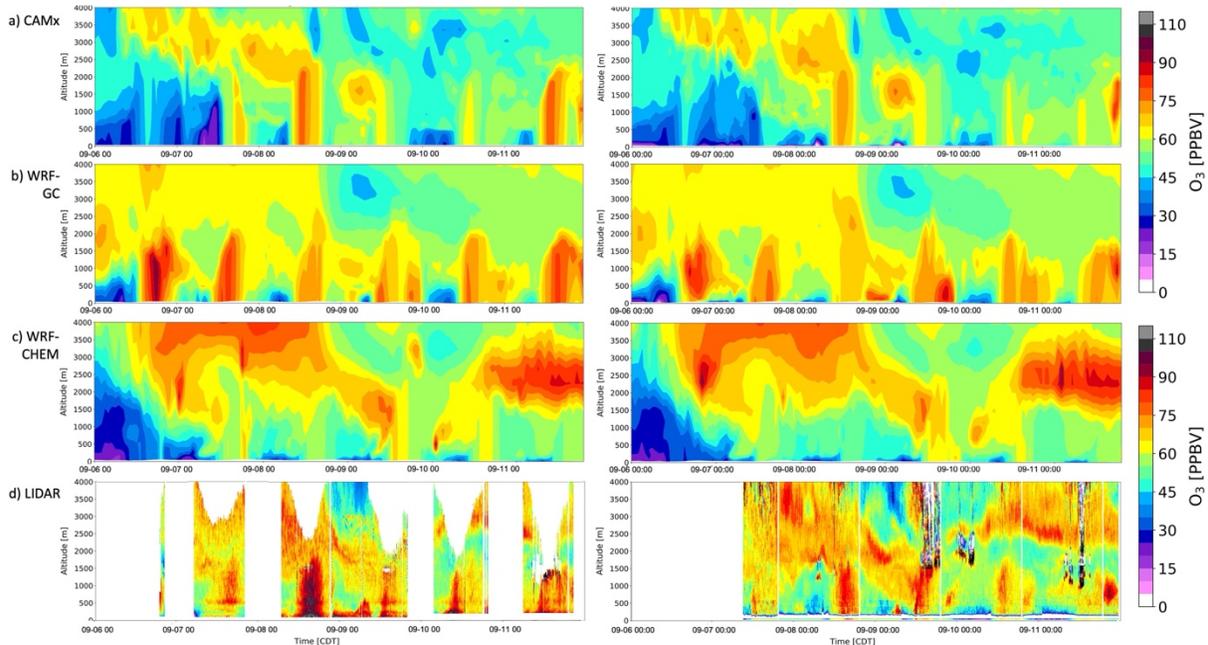


Figure 33. Ozone curtain profiles from September 6-11. Three model intercomparisons vs. lidar curtain at UH - LMOL lidar (left panel) and La Porte - GSFC lidar (right panel). White space denotes missing/unavailable lidar data.

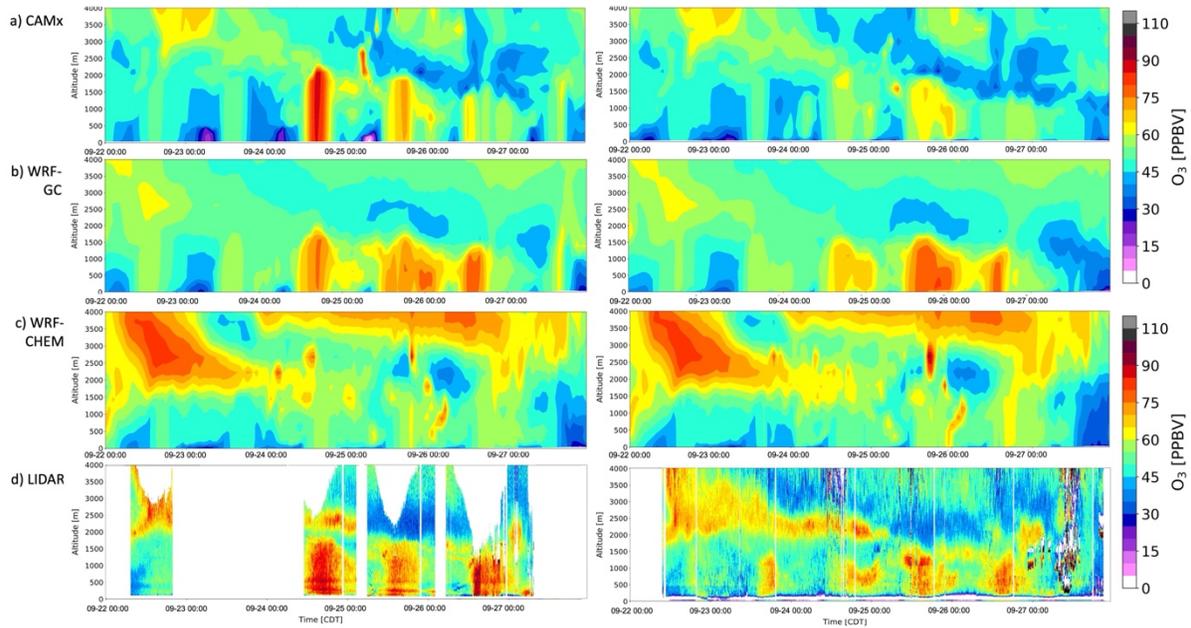


Figure 34. Ozone curtain profiles from September 23-26. Three model intercomparisons vs. lidar curtain at UH - LMOL lidar (left panel) and La Porte - GSFC lidar (right panel). White space denotes missing/unavailable lidar data.

The modeled vertical O_3 profiles were also intercompared against the afternoon ozonesondes launched over Galveston Bay and the Gulf of Mexico. During the study period, there were nine afternoon launches over Galveston Bay while the Gulf of Mexico only had five afternoon launches during high- O_3 events. The average O_3 profiles from these launches are shown in **Figure 35**. Over Galveston Bay, CAMx best captures ozone concentrations near the surface (0-0.5 km) while WRF-GC shows overestimation and WRF-Chem shows underestimation. At 0.5-4 km aloft, all three models show different extents of ozone underestimation, indicating the long-range transported O_3 is underrepresented by all models. At 4-6 km aloft, WRF-GC best captures ozone concentrations while CAMx has a negative bias and WRF-Chem has a positive bias. Over the Gulf, WRF-GC best captures surface ozone below 1 km, WRF-Chem best captures ozone at 2-4 km aloft, and CAMx and WRF-GC best capture ozone at 4-6 km aloft.

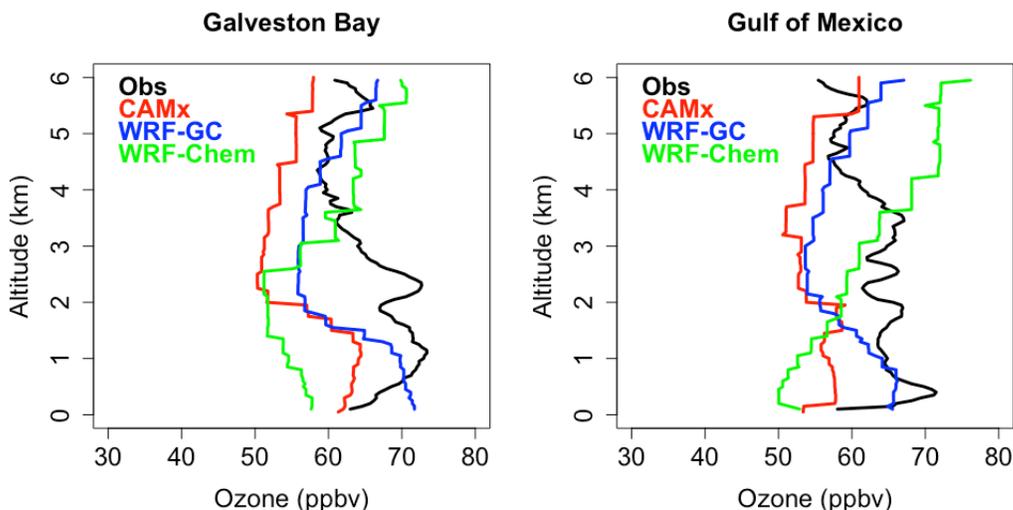


Figure 35. Ozone vertical distribution from afternoon (12:00-18:00) ozonesonde launches at Galveston Bay and the Gulf of Mexico averaged over ozone episodes.

4.2.5. VOC and NO_x sensitivity

This section compares model-predicted ozone and its precursor concentrations to examine ozone formation regimes, using the first ozone episode of September 8-11, 2021 as an example. **Figure 36a** shows domain-mean ozone column concentrations of 57 ± 2 ppbv, 63 ± 3 ppbv, and 59 ± 1 ppbv for CAMx, WRF-GC, and WRF-Chem, respectively. Despite different concentrations, the three models share a common feature of higher ozone over water than over land. **Figure 36b** shows domain-mean NO₂ column concentrations of 0.48 ± 0.26 ppbv, 0.43 ± 0.18 ppbv, and 0.32 ± 0.13 ppbv for CAMx, WRF-GC, and WRF-Chem, respectively. CAMx has high NO₂ near Houston Ship Channel. **Figure 36c** shows domain-mean HCHO column concentrations of 1.62 ± 0.23 ppbv, 1.68 ± 0.35 ppbv, and 1.75 ± 0.29 ppbv for CAMx, WRF-GC, and WRF-Chem, respectively.

Figure 36d shows FNR simulated by three models. CAMx simulates Harris County to be mostly VOC-sensitive because of high NO_x and low HCHO concentrations, while WRF-GC simulates a smaller region of the urban core to be VOC-sensitive than CAMx. WRF-Chem shows the region in a transitional regime. Over Galveston Bay, CAMx simulates the western Ship Channel to be VOC-sensitive, while the other two models simulate the whole Galveston Bay to be NO_x-sensitive. The inter-model differences over Galveston Bay have mainly resulted from offshore NO_x emissions, that is CAMx shows more offshore NO_x than the other two models. The Gulf is simulated to be mostly NO_x-sensitive by all three models.

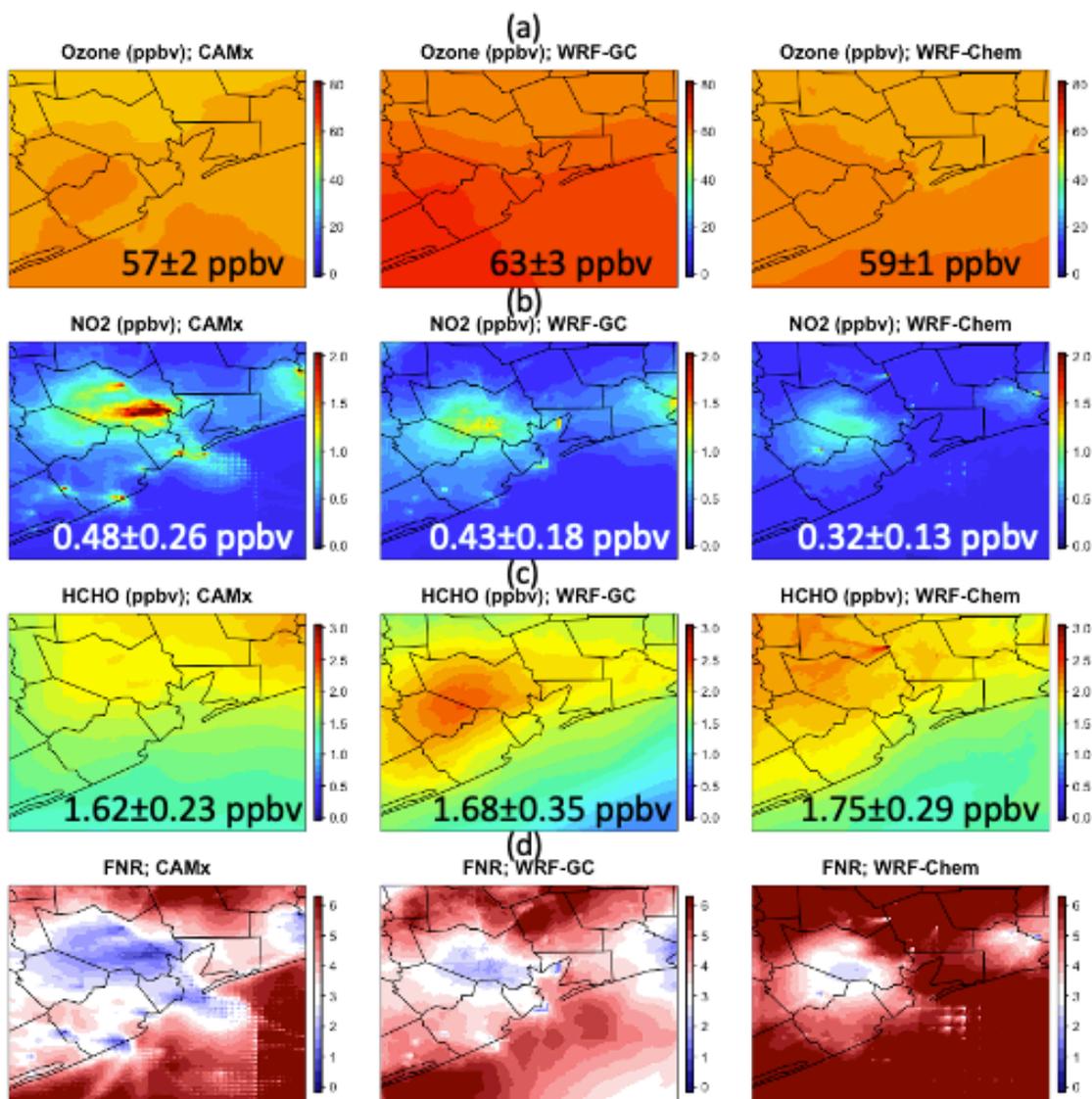


Figure 36. Domain mean column (0-8 km) concentrations of (a) ozone, (b) NO₂, (c) formaldehyde (HCHO), and (d) formaldehyde-to-NO₂ ratio (FNR) simulated by CAMx, WRF-GC, and WRF-Chem over the ozone episode of September 8-11, 2021. In (d), ozone transitional regime is defined by values specially derived for Houston ($2.6 \leq \text{FNR} \leq 3.5$) in Jin et al. (2020).

4.3. Summary

This section compares inter-model differences among three photochemical models, i.e., CAMx, WRF-GC, and WRF-Chem, used to simulate ozone and precursor gases during the 2021 TRACER-AQ campaign in Houston and adjacent waters. Observations from multi-platforms are used; surface measurements include (1) CAMS sites, (2) mobile boats operating in Galveston Bay and the Gulf of Mexico, and (3) mobile driving in urban Houston, while remote sensing measurements include airborne (1) GCAS and (2) HSRL-2, ground-based (3) LMOL and (4) TROPOZ, and (5) ozonesondes. In comparison with surface measurements, CAMx best captures ozone concentrations, while WRF-GC and WRF-Chem show significantly higher positive and negative biases, respectively. One of the major reasons for such differences originates from the

ozone episode of September 17-19, which is caused by the passing of Hurricane Nicolas. Models have different extents of difficulties to represent meteorology and chemistry after the hurricane. The second major reason is associated with clean days. All three models show substantially larger biases and less correlation on clean days than on ozone episode days. In comparison with column concentrations measured by remote sensing instruments, all models underestimate ozone at 0-1 km aloft and NO_x and HCHO at 0-8 km aloft, but WRF-GC shows the smallest bias compared with the other two models. Ozone sensitivity derived from column concentrations of NO_x and HCHO is best captured by WRF-GC (diurnal variations) and CAMx (spatial patterns). WRF-Chem has the best performance in simulating long-range transported ozone plumes above the boundary layer as observed by the two ozone lidars, which may be attributed to the fact that only WRF-Chem implemented 2021 fire emissions.

5. Investigation of Elevated Offshore Ozone's Sources

The 2021 offshore monitoring data confirmed that high O₃ does occur over water and can be associated with weak synoptic forcing and local recirculation events as well as after frontal passages under northerly flow. Under such conditions, on-land ozone monitors also registered high concentrations. It is not clear, based on observations alone, to what extent high ozone over land and the water are interrelated and what emission sources are responsible for high ozone over the water. In addition to *in situ* ozone production, the fact that northerly flow brings higher ozone over the waters suggests regional background ozone should be an important factor for observed ozone variability over the waters.

To address these questions, the project team carried out sensitivity experiments in two photochemical models, CAMx and WRF-GC. In **Section 5.1**, the project team investigated the background concentrations of air pollutants by turning off local anthropogenic emissions, a standard modeling approach to estimate regional background ozone. In addition, the project team carried out *soft* emission perturbation experiments, in which anthropogenic emissions over the land and the water are decreased separately by 10% to preserve the ozone chemical regimes. **Section 5.2** investigated the propagating effects of these perturbed anthropogenic emissions. Such effects were used to identify how precursor emissions from the land affect *in situ* ozone formation over the waters. The inter-model differences in the simulated ozone responses to the same emission perturbation experiments were also examined. It is worth noting that the majority of the emissions over Houston and adjacent Galveston Bay are land-based emissions. Therefore, this report focuses on the effects of emission reductions on land, while emissions reductions over water will be analyzed in the following monthly and final reports.

Table 15 describes the emission perturbation experiments implemented in the two models. The simulations were performed from September 5 to September 11, 2021, with the first day of September 5 as a spin-up. All the anthropogenic emission perturbations were conducted in the inner-most nested domain at 1.33 km resolution; full emissions (incl. anthropogenic, natural) were adopted in the parent domains. Natural emissions remain unchanged among simulations. Biogenic emissions were the MEGAN model (v2.1) in WRF-GC and the BEIS (v5) in CAMx. The MEGAN in WRF-GC inherits the same implementation as in GEOS-Chem where high-resolution emission factors and land functional types from the original MEGAN were regridded into a coarse resolution of 2x2.5 to fit the global model. That is the reason why WRF-GC shows coarse resolution for isoprene in the following figures.

Table 15. Emission perturbation experiments for anthropogenic emissions. Anthropogenic emissions of NO_x and non-methane VOCs are perturbed separately on land.

	Land NO _x	Land VOC	Water NO _x	Water VOC
[CTR]	Full	Full	Full	Full
[BGD]	Zero	Zero	Zero	Zero
[Land_NO _x]	10% reduction	Full	Full	Full
[Land_VOC]	Full	10% reduction	Full	Full
[Water_NO _x]	Full	Full	10% reduction	Full
[Water_VOC]	Full	Full	Full	10% reduction

5.1 Regional Background Ozone

Background ozone is generally taken to represent the concentration of ozone in the absence of local or regional anthropogenic influences. In this study, we define the ‘Houston background’ as the surface pollutant concentration in the absence of anthropogenic emissions in the Houston area. It represents a floor below which air quality cannot be improved by eliminating local emissions in the Houston area. This ‘Houston background’ concept is different from the ‘North American background’ defined by the US Environmental Protection Agency (EPA) which is the surface ozone concentration that would be present over the US in the absence of North American anthropogenic emissions (e.g., the ensemble of the US, Canada, and Mexico) (US EPA, 2006). We note that Houston background ozone is a model concept, similar to the North American background, because it is derived from model experiments.

Figure 37 shows the Houston background concentrations ([BGD]) derived from WRF-GC in comparison to surface concentrations in the control simulation ([CTR]) for the ozone episode during two September episodes of Sep 6-11 and Sep 23-26, 2021. The contribution of Houston background to surface NO₂ concentrations is ~22% during Sep 6-11 and ~30% during Sep 23-26. The contribution of Houston background to surface HCHO concentrations is ~73% during Sep 6-11 and ~65% during Sep 23-26. This suggests that the majority of HCHO in Houston is either biogenic or originates from outside of Houston and is transported into this region while NO₂ is primarily from local contributions. Surface isoprene concentrations are higher in [BGD] compared to [CTR], suggesting that the atmosphere in [CTR] has a larger oxidation capacity resulting from local anthropogenic emissions and therefore promotes increased isoprene oxidation. It is worth noting that the background isoprene concentrations in Houston are 6-12 ppbv simulated by the MEGAN model in WRF-GC (**Figure 37**), much higher than the 1-2 ppbv simulated by the BEIS model in CAMx (**Figure 38**). These high background concentrations are likely to cause a high bias of surface ozone as observed in Section 4.

The abundance and reactivity of these ozone precursors determine ozone production regimes, which can be determined by the loss of HO_x radicals (HO_x=OH+HO₂) as the termination of ozone chain reactions. In remote and clean troposphere with low NO_x concentrations, peroxy radicals may react with themselves instead of with NO to produce peroxides and other oxygenated compounds. The most important process, the self-reaction of hydroperoxyl radical (HO₂), produces hydrogen peroxide (H₂O₂) which is therefore used to represent NO_x-limited ozone production. In urban areas with very high NO_x concentrations, the dominant sink for HO_x radicals is the oxidation of NO₂ by OH, resulting in the production of HNO₃. Therefore, HNO₃ is used to represent ozone production under a VOC-limited regime. **Figure 37** shows higher HNO₃ in [CTR] than [BGD], indicating that local anthropogenic emissions lead to much higher VOC-limited ozone production. In contrast, [CTR] and [BGD] have similar levels of H₂O₂ in **Figure 37**, suggesting that NO_x-limited ozone production remains

relatively consistent with and without local anthropogenic emissions. The combined effects of NO_x - and VOC-limited ozone production led to Houston background ozone of 53 ppbv during Sep 6-11 and 51 ppbv during Sep 23-26, which accounts for the majority of surface ozone of 61 ppbv and 58 ppbv, respectively during said periods. The differences between background and control concentrations are the local contributions of ~ 7-8 ppbv both on land and over water. In Section 4, we found WRF-GC has a high bias of surface ozone during Sep 2021. This high bias may be caused by relatively high background concentration in the model, suggesting either natural emissions or regional anthropogenic emissions are overestimated in the WRF-GC model.

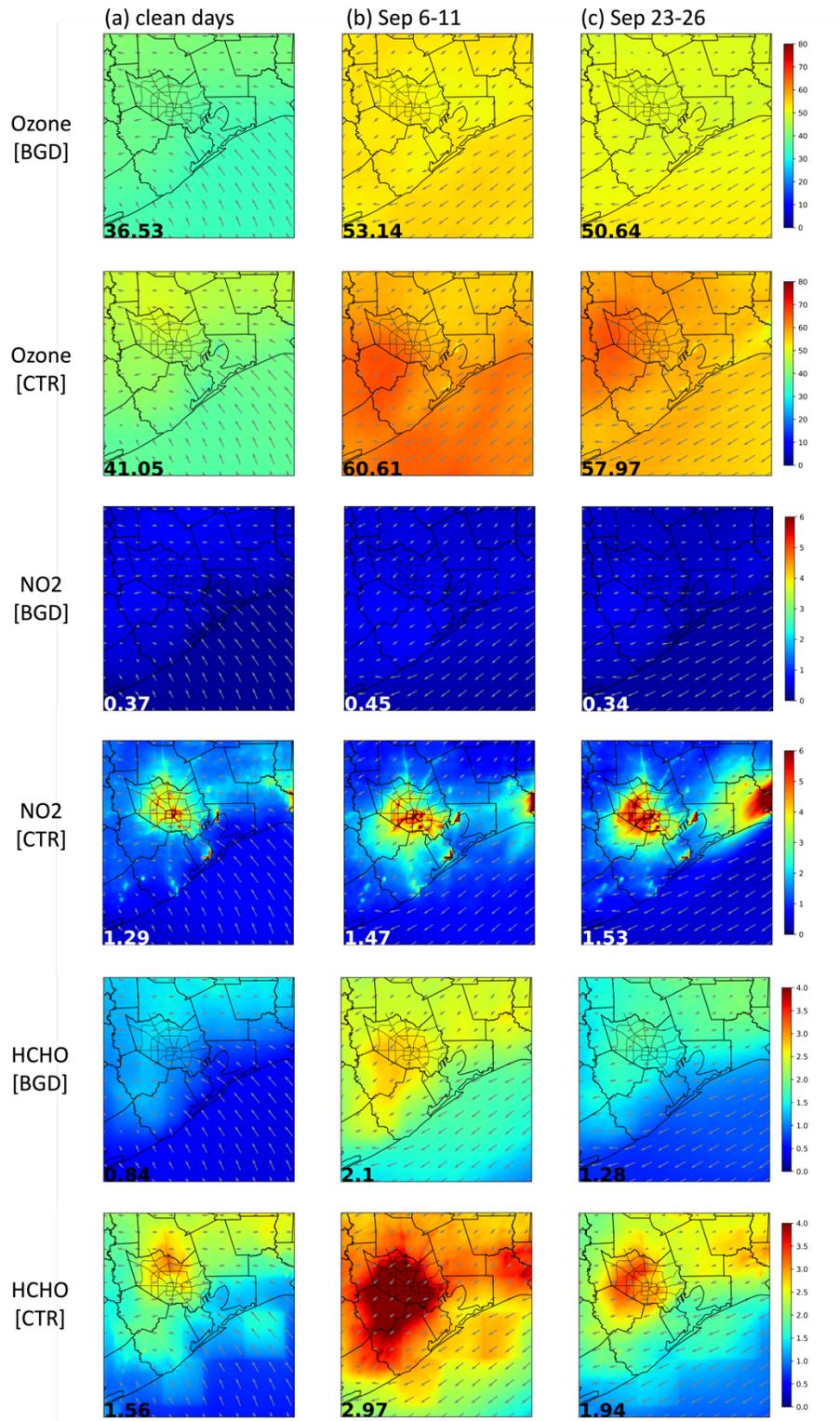


Figure 37. WRF-GC simulated surface concentrations (in ppbv) of O_3 , NO_2 , and HCHO from the background simulation [BGD] and the control simulation [CTR] during the daytime (8-18 CDT) on (a) clean days, (b) September 6-11 and (c) September 23-26, 2021. Domain averages are inserted.

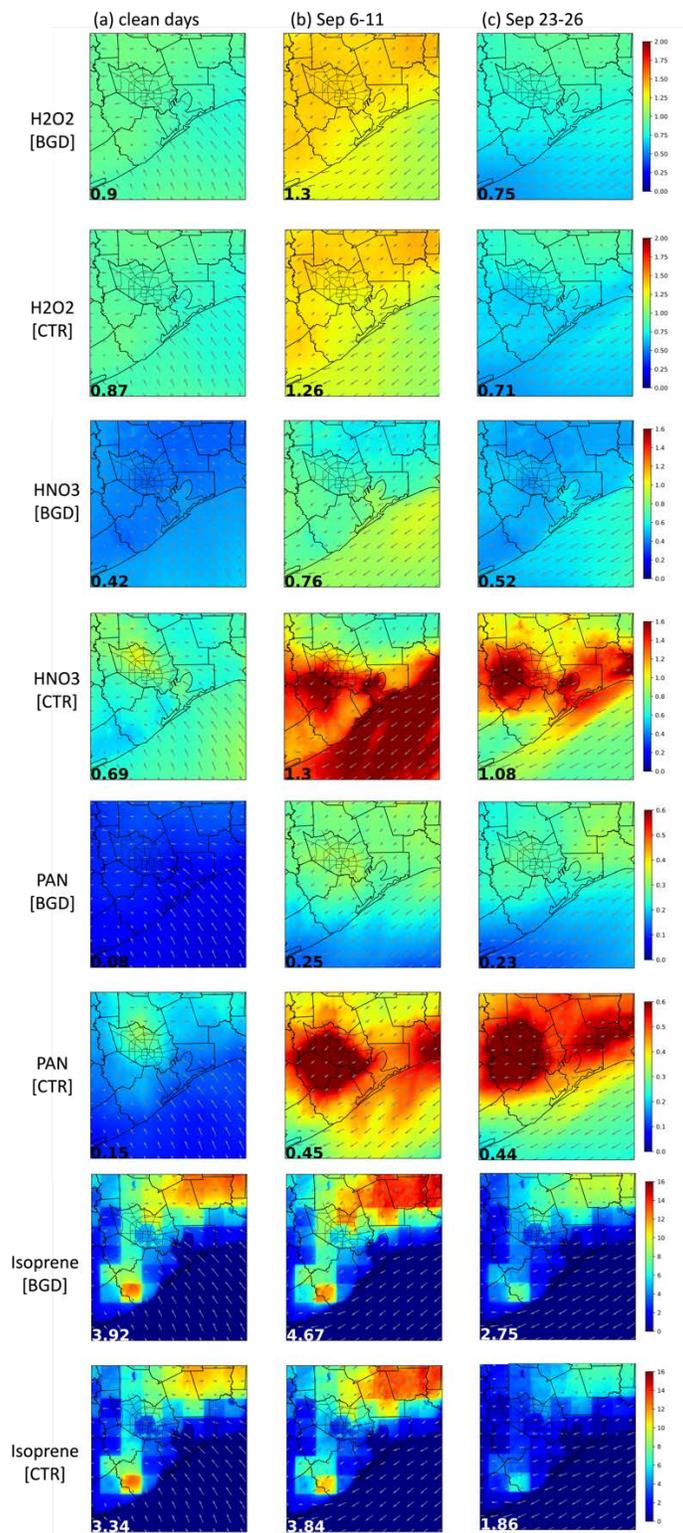


Figure 37 (continued). WRF-GC simulated surface concentrations (in ppbv) of H₂O₂, HNO₃, PAN, and isoprene from the background simulation [BGD] and the control simulation [CTR] during the daytime (8-18 CDT) on (a) clean days, (b) September 6-11 and (c) September 23-26, 2021. Domain averages are inserted.

Figure 38 shows CAMx-derived Houston background surface concentrations of various pollutants (NO_2 , HCHO, O_3 , H_2O_2 , HNO_3 , PAN, and isoprene) compared to the corresponding control simulations for ozone episode days during September 6-11 and September 23-36. The background concentrations of NO_2 only account for ~16% of NO_2 during Sep 6-11 and ~17% during Sep 23-26 in the control simulation. This suggests that most of the NO_2 concentrations are contributed by local anthropogenic emissions, similar to the findings from the WRF-GC simulation where background NO_2 accounts for 22-30% of surface NO_2 concentrations. The difference in background contributions indicated from the two models (16%-17% from CAMx vs 22-30% from WRF-GC) is likely due to the difference in how the model accounts for natural NO_x emissions. For example, CAMx does not have lightning NO_x emissions but WRF-GC does. In terms of HCHO, the CAMx model suggests that background HCHO accounts for ~75% of the surface concentrations of HCHO during Sep 6-11 and 77% during Sep 23-26, consistent with the WRF-GC simulation results. On the other hand, the H_2O_2 concentrations in the background simulation are ~85% of the control simulation during Sep 6-11 and ~92% of the control simulation during Sep 23-26, suggesting that NO_x -limited ozone production was slightly increased by local anthropogenic emissions. HNO_3 is much higher in the control simulation compared to the background for both episodes, suggesting that local anthropogenic emissions lead to increased VOC-limited ozone production.

While both the WRF-GC and CAMx models find similar contributions from background NO_2 and HCHO concentrations to the overall surface concentrations, there are differences in the absolute values of the estimated concentrations. The CAMx model consistently shows lower NO_2 and HCHO concentrations than WRF-GC, resulting in lower ozone estimations as well. Previous reports have indicated that CAMx underestimates HCHO and NO_2 in the Texas and Houston region by up to 20% (Goldberg et al., 2022). During Sep 6-11 episode, Houston background ozone derived from CAMx is ~38 ppbv compared to ~47 ppbv from the control simulation, while during Sep 23-26, Houston background ozone is ~40 ppbv compared to ~47 ppbv from the control simulation. Although both values are lower than those from WRF-GC, the differences in surface ozone between [BGD] and [CTR] as estimated by CAMx are ~9 and ~7 ppbv over land and water during these two episodes, respectively. These values are similar to those of 7-8 ppbv derived from WRF-GC. However, CAMx predicted a slightly higher local contribution to the total ozone (19% during Sep 6-11 and 15% during Sep 23-26) than WRF-GC (~13% for both episodes). Given the large contributions of non-local emissions to surface ozone in Houston, there is a need to better model regional transport dynamics for improving ozone simulation.

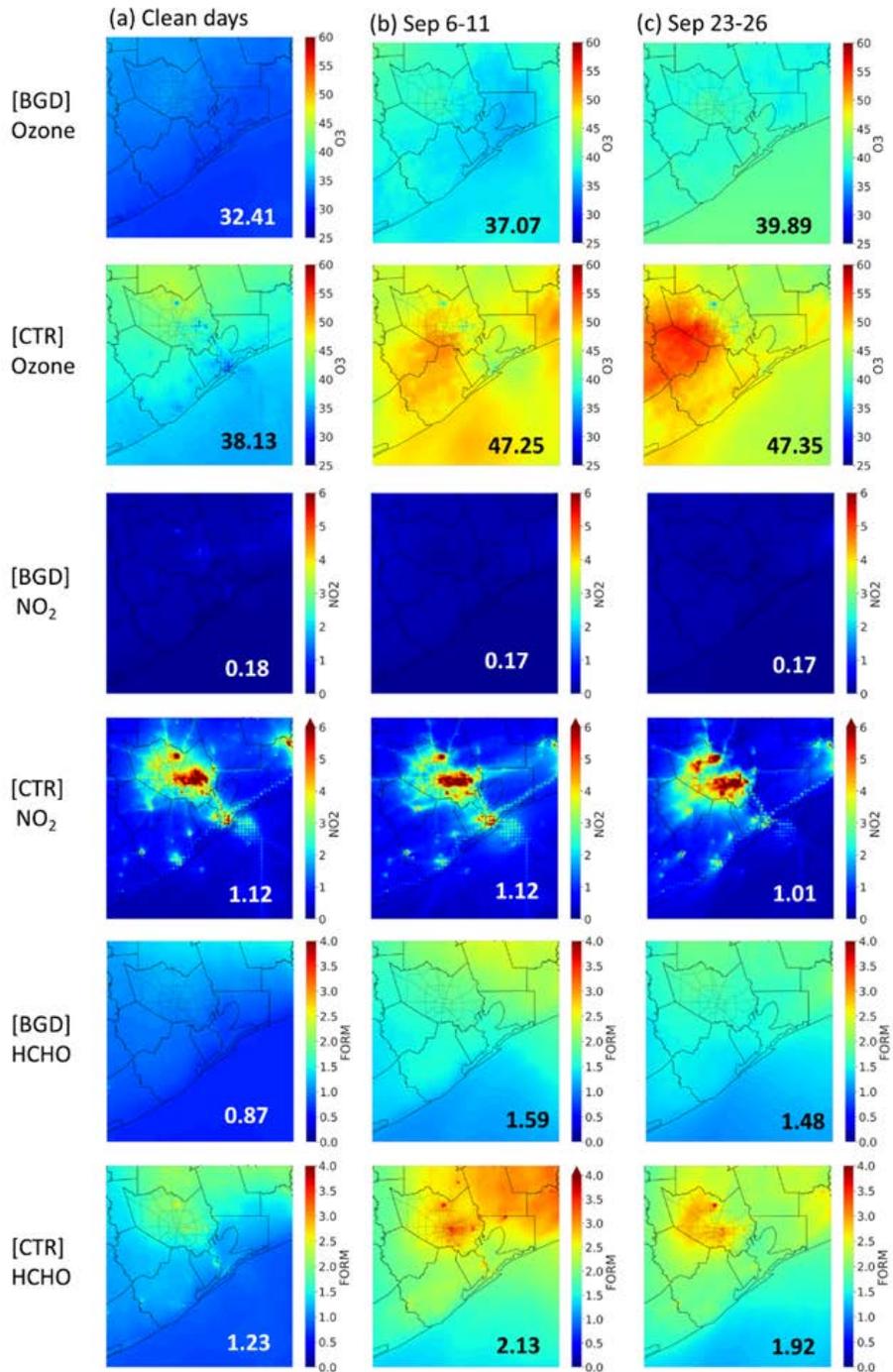


Figure 38. CAMx simulated surface concentrations (in ppbv) of O₃, NO₂, and HCHO from the background simulation [BGD] and the control simulation [CTR] during the daytime (8-18 CDT) on (a) clean days, (b) September 6-11 and (c) September 23-26, 2021. Domain averages are inserted.

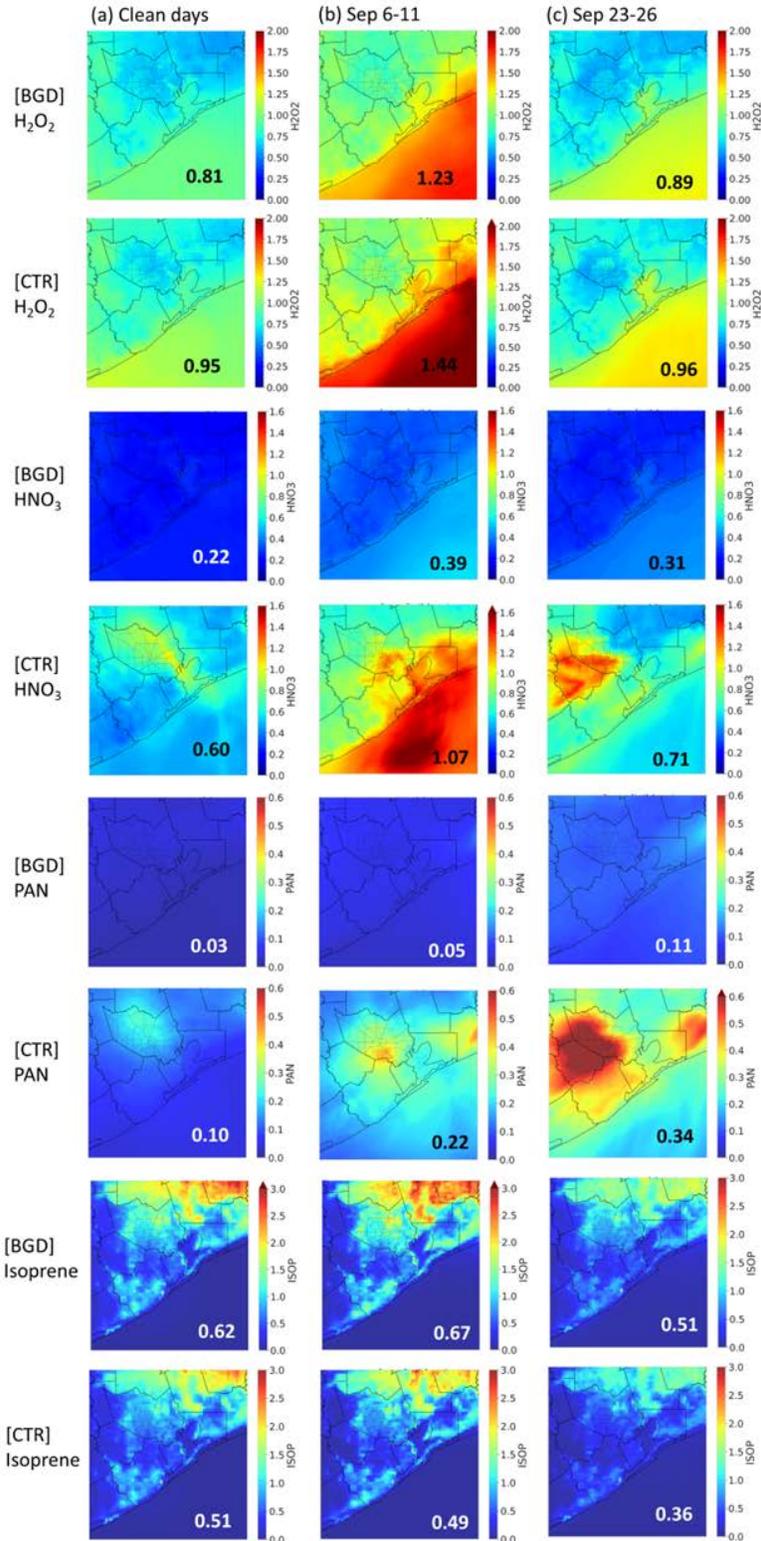


Figure 38 (continued). CAMx simulated surface concentrations (in ppbv) of H₂O₂, HNO₃, PAN, and isoprene from the background simulation [BGD] and the control simulation [CTR] during the daytime (8-18 CDT) on (a) clean days, (b) September 6-11 and (c) September 23-26, 2021. Domain averages are inserted.

Figure 39 and **Figure 40** display the vertical profiles of different species over land and over water from the [BGD] and [CTR] simulations of WRF-GC and CAMx, respectively. Common findings from both models are background concentrations dominate above 2 km for most species, while local emission contributions are confined below 2 km. The main differences between the two models are: (1) WRF-GC predicts ~ 3 ppbv higher Houston background ozone over water than over land, whereas CAMx predicts the same Houston background ozone over both locations; and (2) CAMx predicts lower H₂O₂ in [BGD] at both locations, whereas WRF-GC predicts no change, if not slightly higher, in H₂O₂ in [BGD]. The second difference can be explained by much higher background NO₂ in WRF-GC (0.34-0.44 ppbv) than CAMx (0.17-0.18 ppbv), not only in absolute concentrations but also in relative contributions to total NO_x in Houston. The high background NO_x in WRF-GC over land locations results in less sensitivity in this model to local NO_x emissions changes and hence no change in H₂O₂ between [BGD] and [CTR]. Since both models have minimal natural emissions over water, the first difference indicates regional transport has a larger contribution to Houston background in WRF-GC than CAMx leading to higher background ozone over the water.

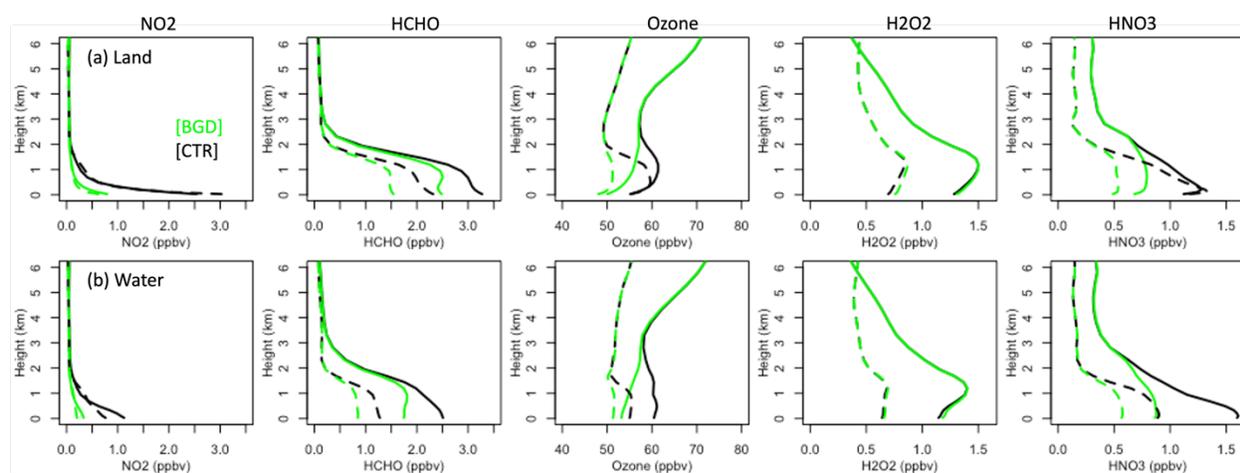


Figure 39. Vertical profiles of NO₂, HCHO, ozone, H₂O₂, and HNO₃ (a) on land and (b) over water obtained from WRF-GC [BGD] and [CTR] simulations during the daytime (8-18 CDT) over September 6-11 (solid lines) and September 23-26 (dotted lines), 2021.

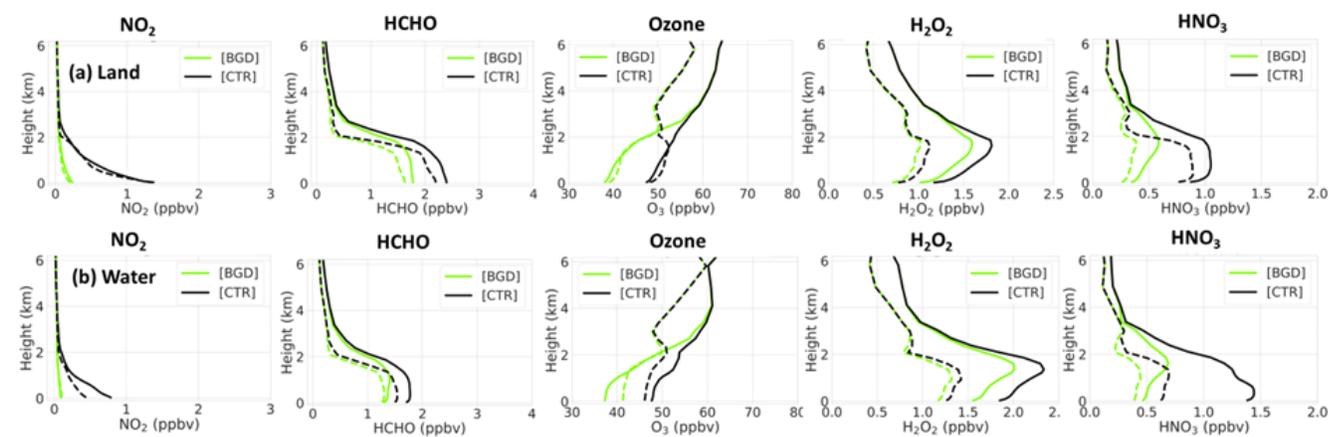


Figure 40. Same as Figure 39 but for CAMx.

5.2 Emission Perturbation Experiments

The above section shows that the surface ozone in Houston has large background contributions of 38-53 ppbv from outside of Houston in comparison with local contributions of within 10 ppbv. These local contributions are expected to respond to anthropogenic emission reductions in Houston. Therefore, this section investigates the corresponding changes of surface ozone and related chemical species in response to 10% reductions in on-land and over-water anthropogenic NO_x and non-methane VOC emissions, respectively. The choice of 10% perturbation is made in order to preserve ozone production regions in both models.

Such changes are shown in **Figures 41-44** for WRF-GC and in **Figures 45-48** for CAMx. In general, WRF-GC shows more spatial heterogeneity to local emission perturbation than CAMx, possibly due to complex aerosol-radiation and aerosol-cloud interactions. Conversely, CAMx shows more spatial homogeneity because no aerosol components interfere with ozone formation in the model. It is noteworthy that WRF-GC has much higher isoprene concentrations in [CTR] than CAMx, because the two models use different biogenic isoprene emission algorithms. This difference further explains why WRF-GC has higher background ozone than CAMx. To compare different responses on land and over water, we select three specific regions of interest: urban Houston, Galveston Bay, and the Gulf of Mexico. The regional differences are summarized in **Figure 42** (September 6-11) and **Figure 44** (September 23-26) for WRF-GC and **Figure 46** (September 6-11) and **Figure 48** (September 23-26) for CAMx.

Figure 42 shows that on-land emission reductions can affect ozone concentrations both on land and over water, while over-water emission reduction exhibits more influences over the waters than on land. A 10% reduction of on-land NO_x emissions in WRF-GC leads to a 0.21 ppbv (0.37%) ozone decrease in urban Houston, no ozone changes in the Bay, and a 0.12 ppbv (0.19%) ozone increase in the Gulf during September 6-11. A 10% reduction of on-land non-methane VOC emissions in WRF-GC leads to a 0.01 ppbv (0.01%) ozone decrease in urban Houston, a 0.10 ppbv (0.18%) ozone decrease in the Bay, but a 0.15 ppbv (0.25%) ozone increase in the Gulf during September 6-11. Ozone changes do not exhibit large responses because of the high background in this model, making it less sensitive to small local emission perturbations. Despite the small response in ozone (less than 0.4% in all locations) because of the small perturbation in emissions, we can infer urban Houston is in NO_x limited regime in WRF-GC because surface ozone responds more readily to NO_x emission reductions. It is noteworthy that the ozone change over the waters (Bay and the Gulf) is opposite in sign to that over the land. NO_y species (NO₂, PAN, and HNO₃) over the Gulf show an increase following a 10% reduction of land emissions. This is likely due to an increase in the lifetime of NO_y species over land that allows more transport of those land-based emissions toward the waters, which then leads to increased ozone production over the water. This effect demonstrates the resiliency of high ozone over water to small changes in land emissions, making it difficult to control.

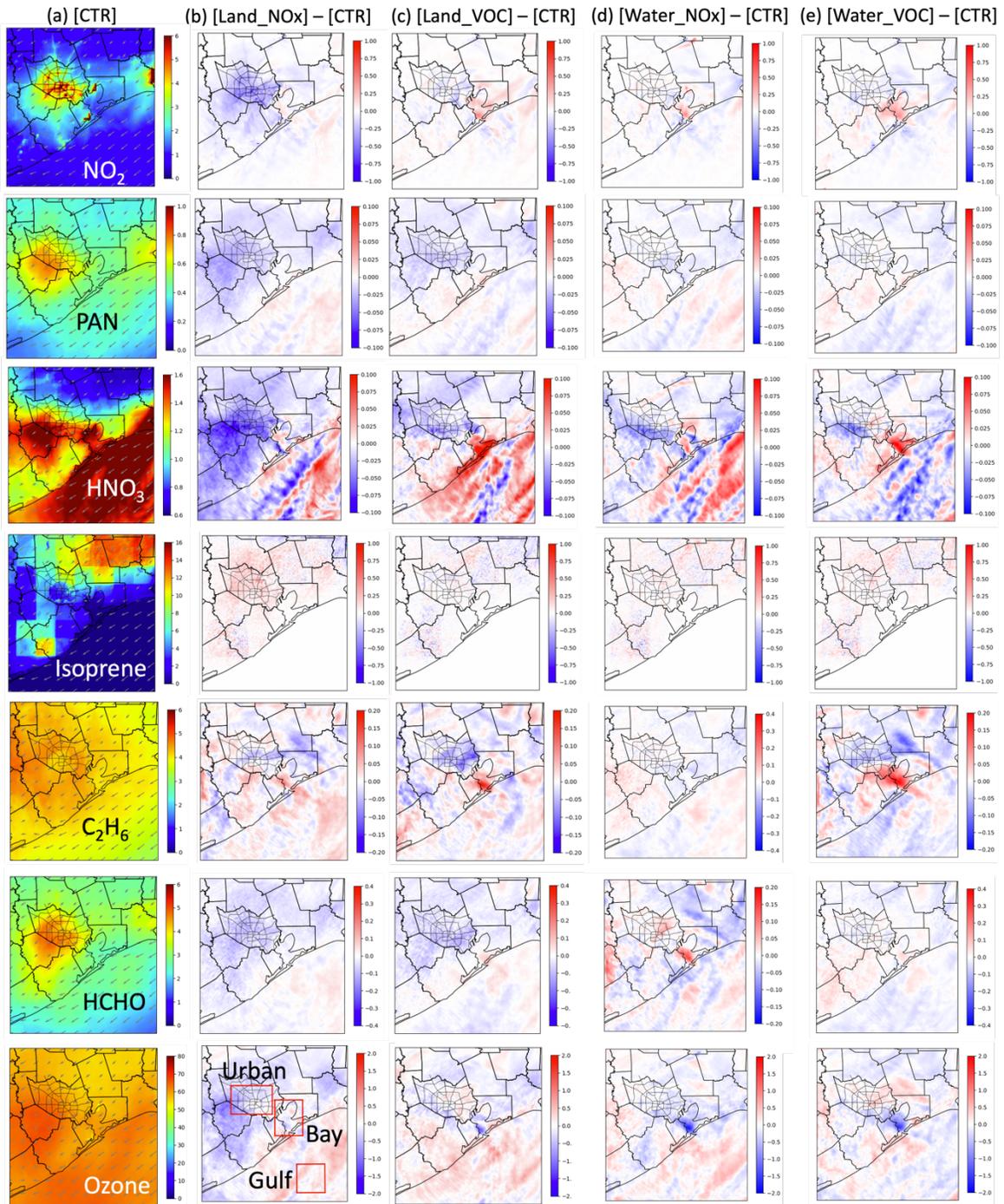
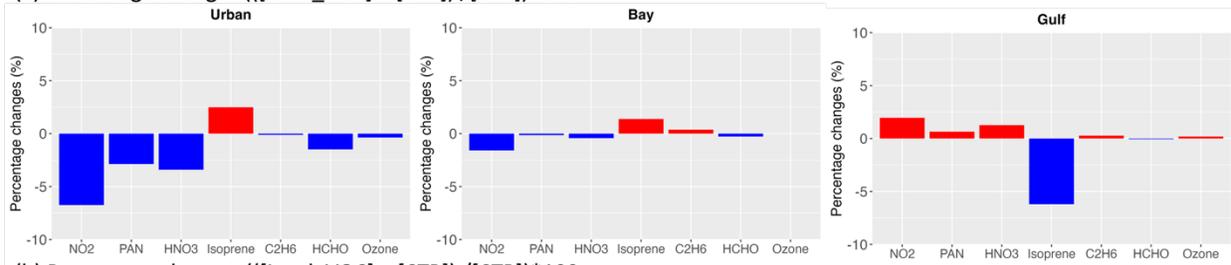
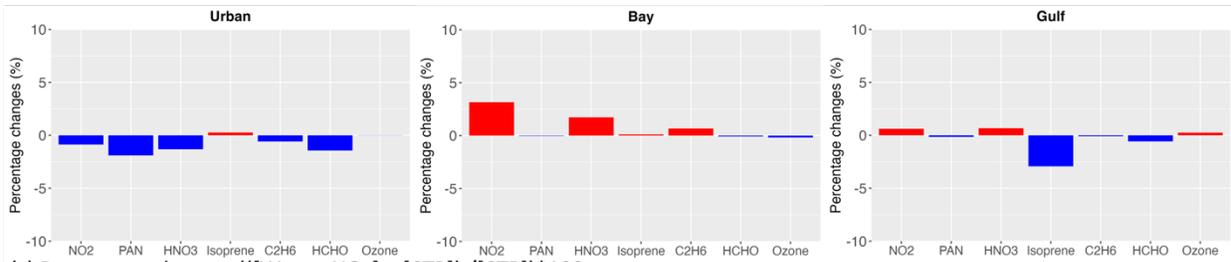


Figure 41. (a) WRF-GC surface concentrations in the control simulation [CTR]. Concentration changes of (b) the onshore NO_x emission reduced simulation [Land_ NO_x] minus [CTR], (c) the onshore VOC emission reduced simulation [Land_VOC] minus [CTR], (d) the offshore NO_x emission reduced simulation [Water_ NO_x] minus [CTR], and (e) the offshore VOC emission reduced simulation [Water_VOC] minus [CTR] during the daytime (8-18 CDT) over September 6-11, 2021.

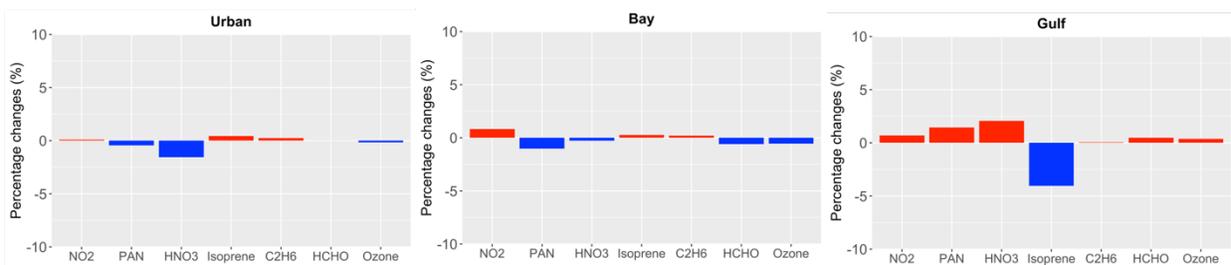
(a) Percentage changes $(([\text{Land_NO}_x] - [\text{CTR}]) / [\text{CTR}]) * 100$



(b) Percentage changes $(([\text{Land_VOC}] - [\text{CTR}]) / [\text{CTR}]) * 100$



(c) Percentage changes $(([\text{Water_NO}_x] - [\text{CTR}]) / [\text{CTR}]) * 100$



(d) Percentage changes $(([\text{Water_VOC}] - [\text{CTR}]) / [\text{CTR}]) * 100$

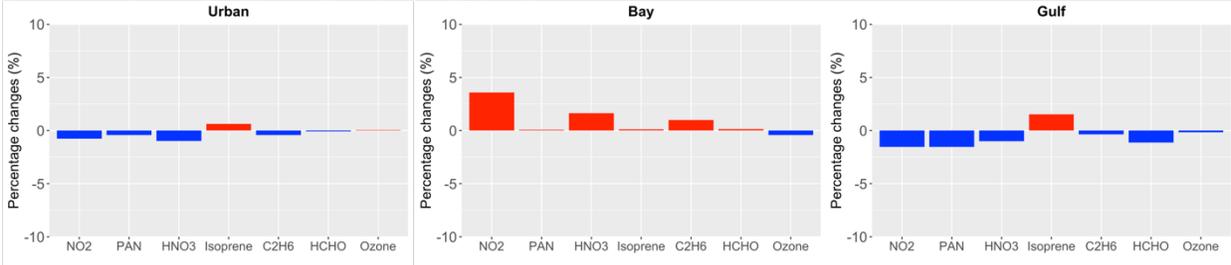


Figure 42. Percentage changes of WRF-GC surface concentrations from the difference between (a) [Land_NO_x] and [CTR], (b) [Land_VOC] and [CTR], (c) [Water_NO_x] and [CTR], and (d) [Water_VOC] and [CTR] relative to [CTR] in urban Houston, the Galveston Bay, and the Gulf of Mexico during the daytime (8-18 CDT) over September 6-11, 2021. The regions are given in Figure 41.

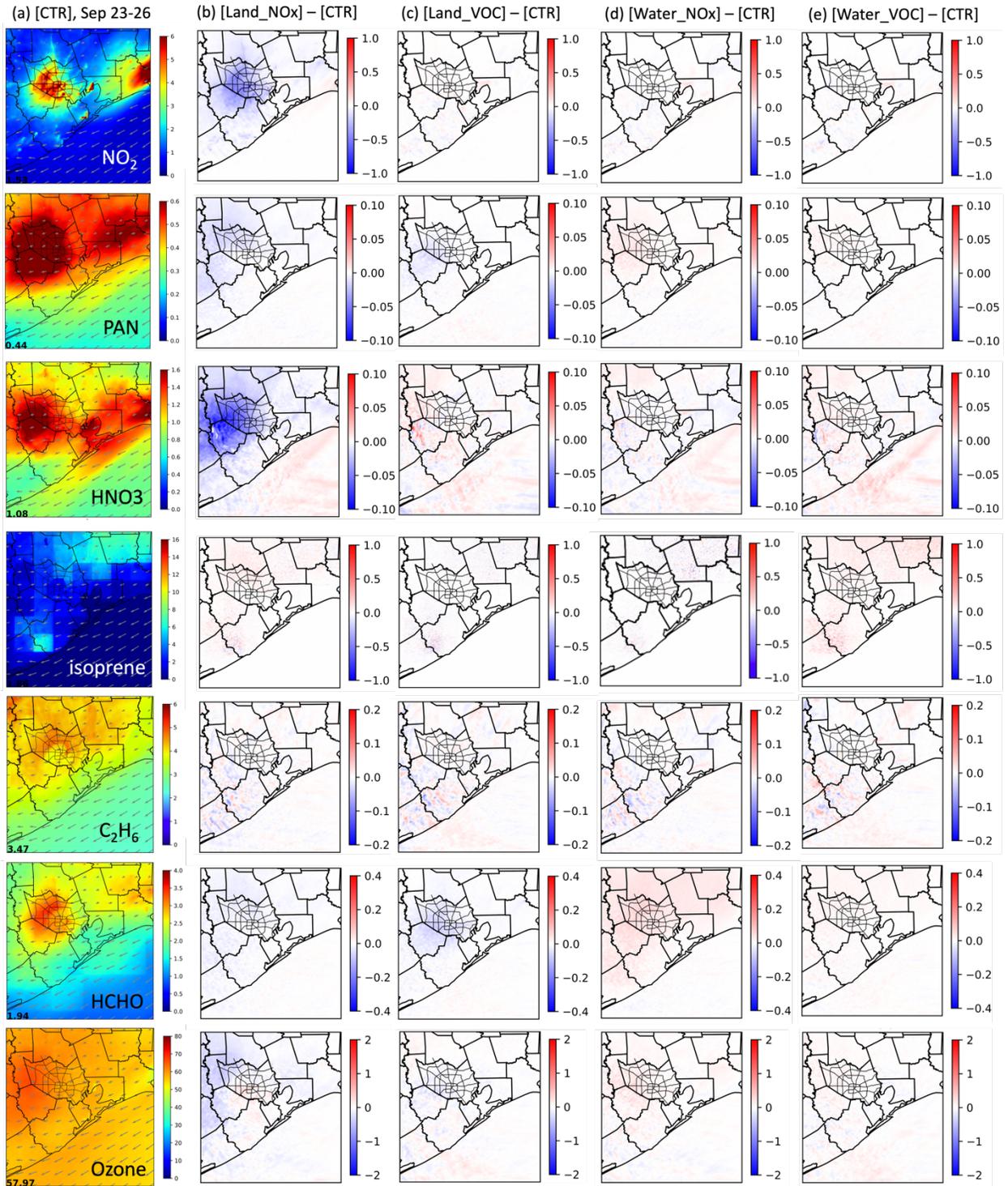


Figure 43. Same as Figure 41 but for September 23-26, 2021.

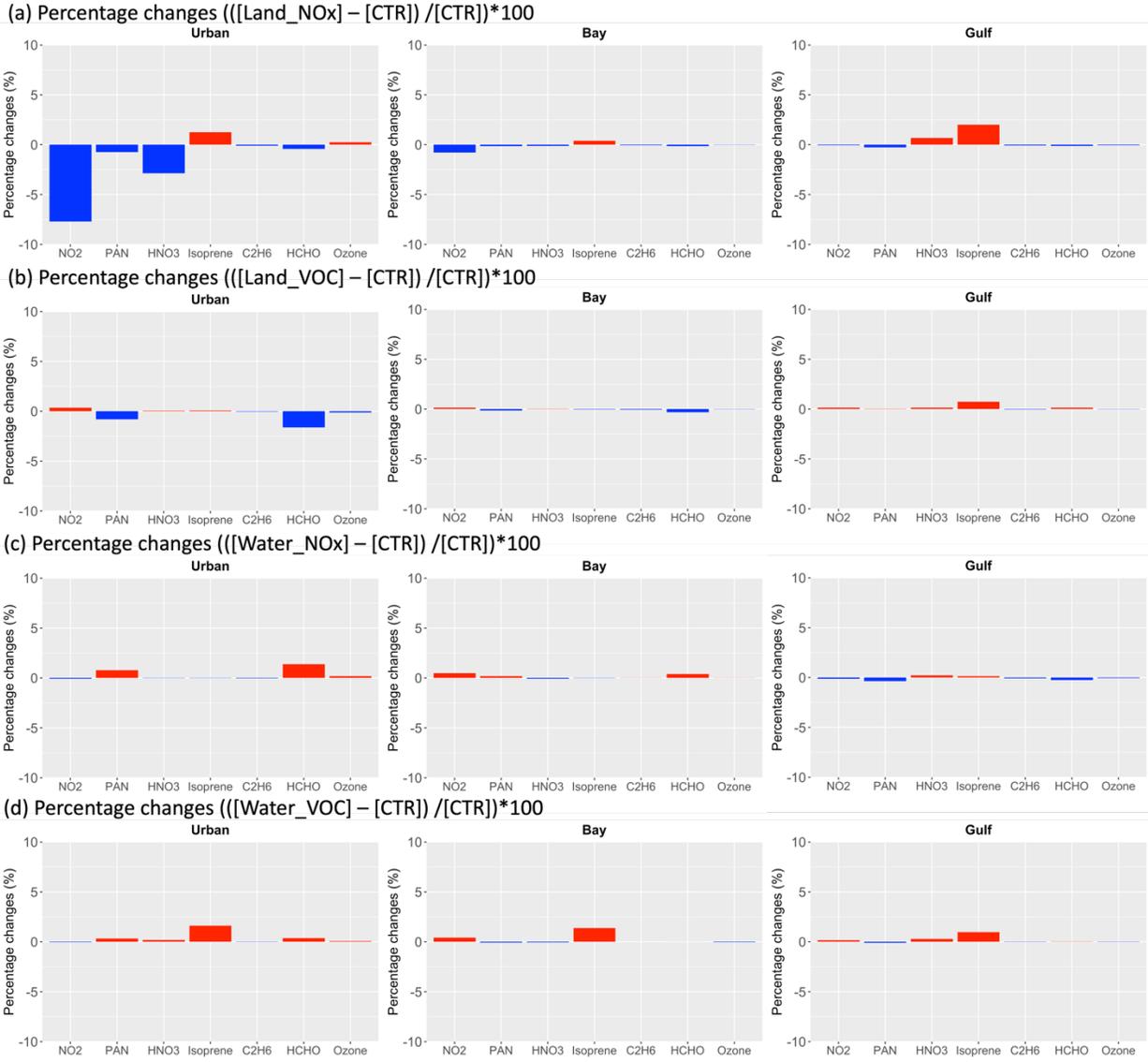


Figure 44. Same as Figure 42 but for September 23-26, 2021.

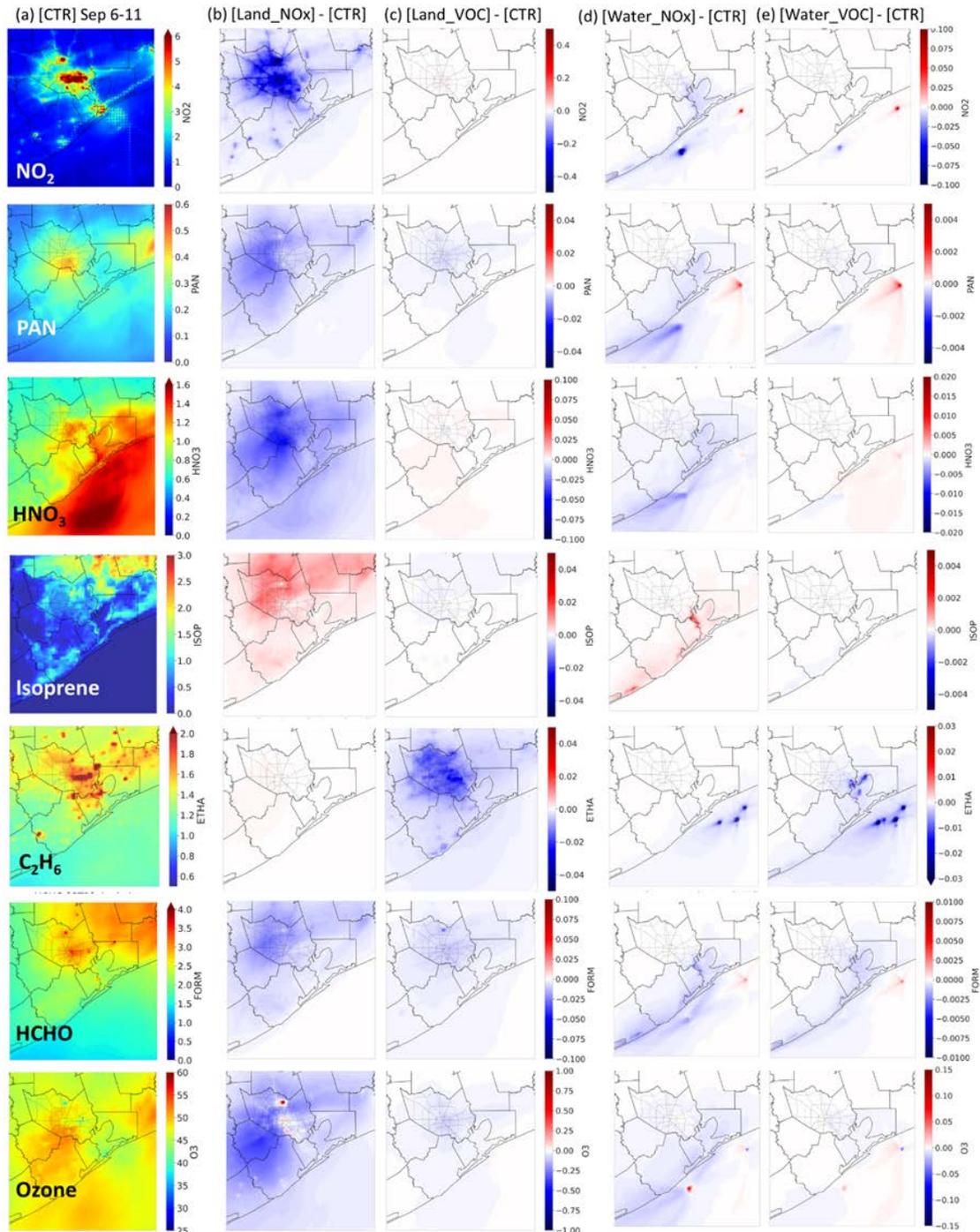


Figure 45. Same as Figure 41 but for CAMx.

Figures 46 and 48 show the percentage changes in ozone and related chemical species resulting from a 10% reduction in anthropogenic NO_x and non-methane VOC emissions over land and water, as obtained from the CAMx model. Similar to the WRF-GC model, CAMx also predicts ozone decreases both on land and over the Bay. During Sep 6-11, a 10% reduction in on-land NO_x emissions leads to ozone decreases of 0.16 ppbv (0.33%) in urban Houston, 0.09

ppbv (0.17%) in the Bay, and 0.03 ppbv (0.07%) in the Gulf. A 10% reduction in on-land non-methane VOC emissions results in an ozone decrease of 0.04 ppbv (0.08%) in urban Houston, 0.01 ppbv (0.04%) in the Bay, and 0.002 ppbv (0.005%) in the Gulf. Consistent with the WRF-GC findings, CAMx simulations indicate that ozone responds more to reductions in NO_x emissions than to reductions in VOC emissions.

In contrast to the responses from on-land emission perturbation, the impact of a 10% reduction in over-water NO_x and non-methane VOC emissions on ozone is relatively small. CAMx predicts ozone decreases of 0.002 ppbv (0.003%) in urban Houston and 0.006 ppbv (0.014%) in the Bay, along with a slight increase of 0.001 ppbv (0.001%) in the Gulf in response to the reduction in over-water NO_x emissions. Similarly, for a 10% reduction in over-water non-methane VOC emissions, CAMx predicts an ozone decrease of 0.001 ppbv (0.002%) in urban Houston, 0.002 ppbv (0.005%) in the Bay, and a slight increase of 0.001 ppbv (0.004%) in the Gulf. These results show that in CAMx, over-water surface ozone is more sensitive to reducing on-land emissions than over-water emissions. This is different from WRF-GC, in which over-water ozone has a similar response to reductions in either on-land or over-water emissions.

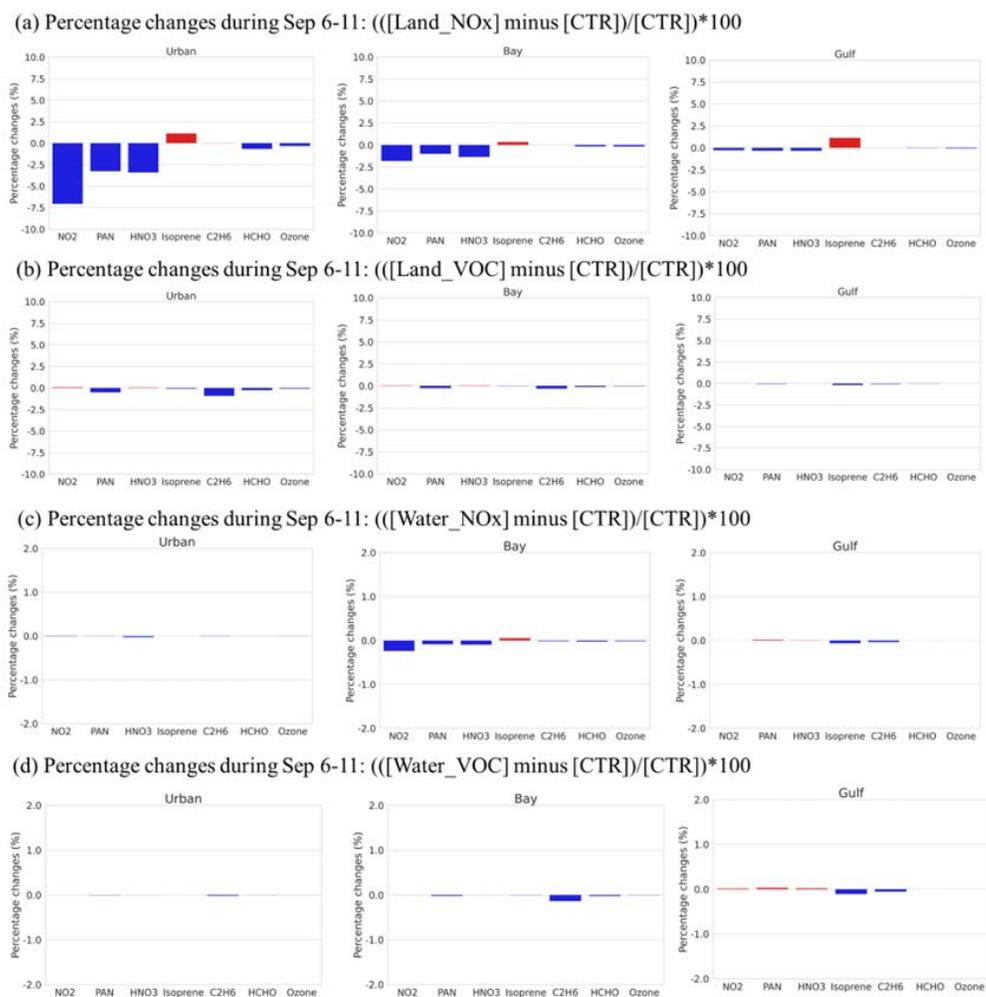


Figure 46. Same as Figure 42 but for CAMx.

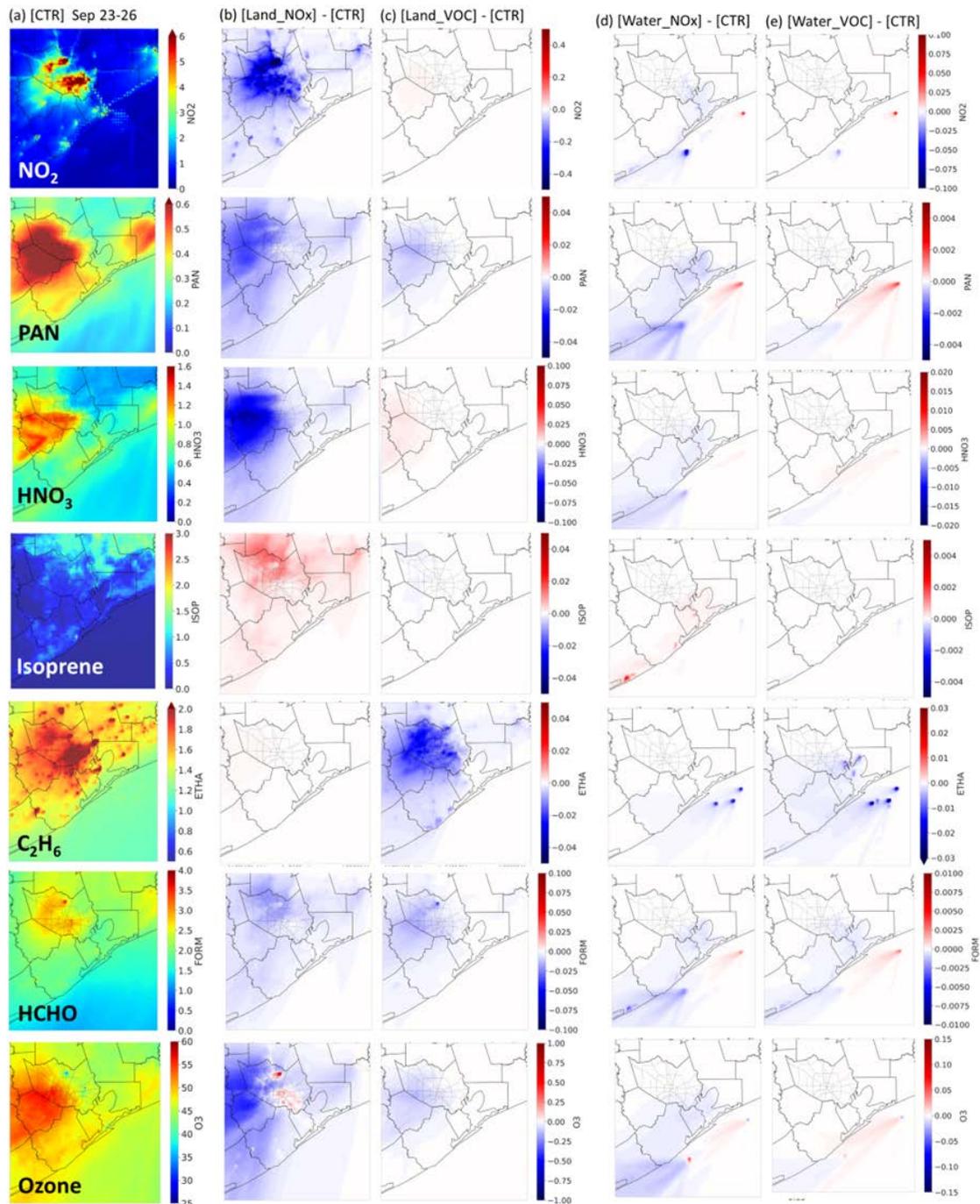


Figure 47. Same as Figure 43 but for CAMx.

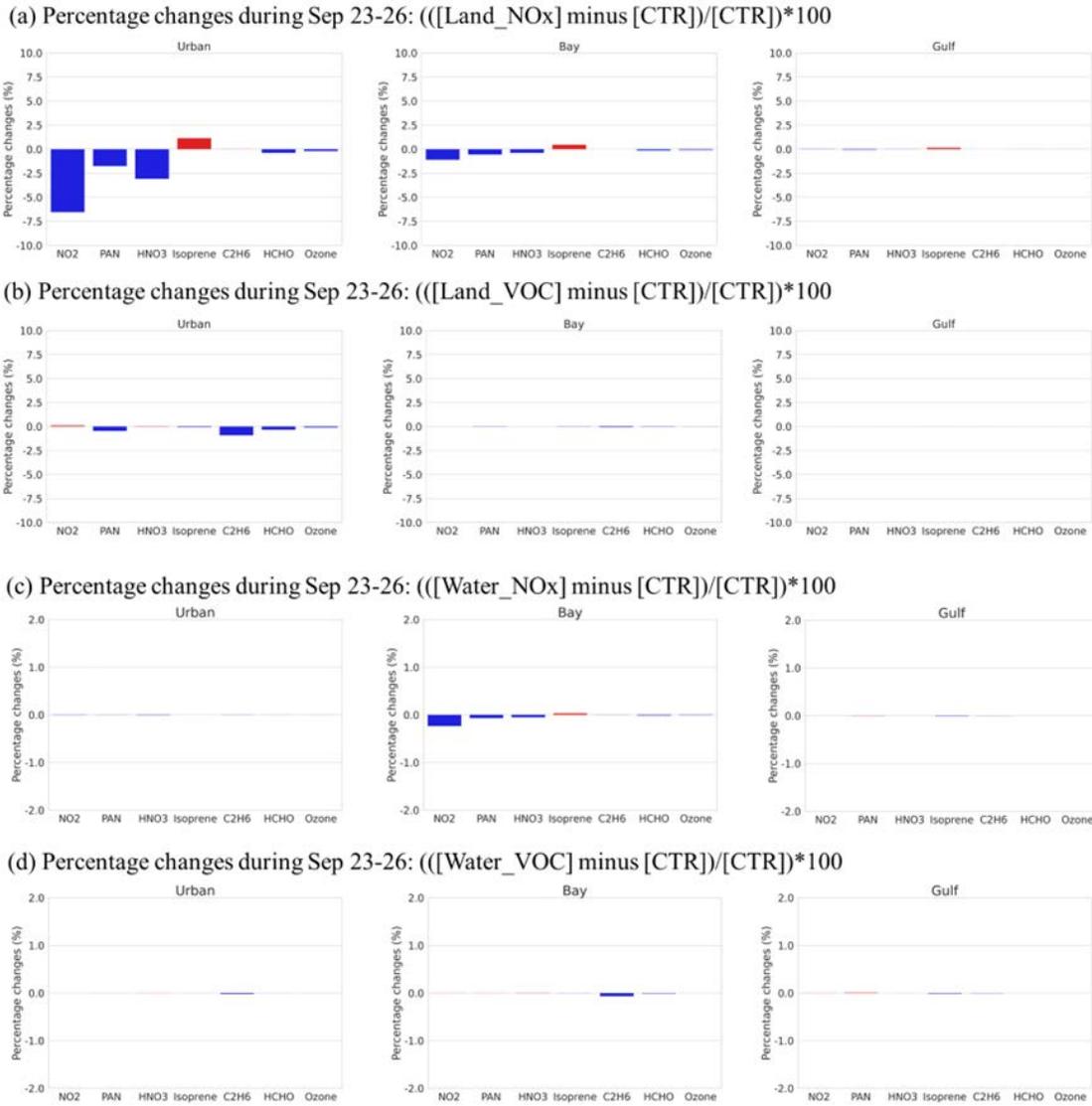


Figure 48. Same as Figure 44 but for CAMx.

5.3 CAMx Process Analysis

We conducted process analysis in the CAMx model to identify key processes which led to simulated O_3 change during high- O_3 episodes relative to clean days. The process analysis is calculated over a subregion of the Gulf of Mexico with high O_3 mixing ratios observed (**Figure 49**) and integrated across the lowest five model layers comparable to the morning PBL heights over water.

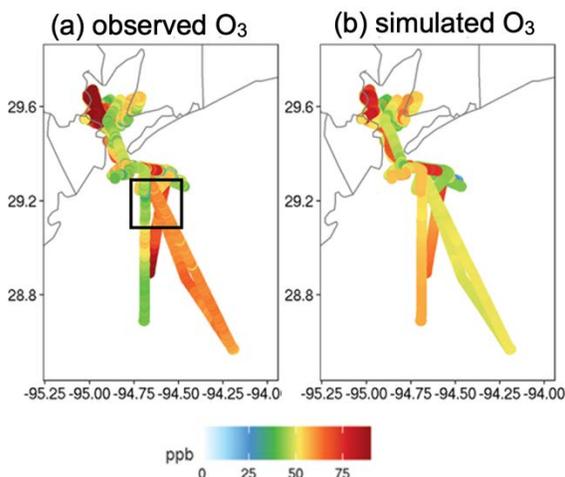


Figure 49. Maps of observed (a) and CAMx-simulated (b) surface ozone over the waters during ozone episode days. The black box shows the selected offshore region for process analysis.

The diurnal average of each process on clean and O_3 episode days is shown in **Figure 50**. Chemistry (CHEM) is the major O_3 source during daytime and becomes the primary O_3 sink after sunset. Advection (ADV) serves as a pathway for an O_3 sink for most hours, especially during the day, while vertical diffusion (DIF) mostly contributes as an O_3 source. Deposition (DEP) constantly removes O_3 from the atmosphere at all hours, yet with a marginal value of 0.1 ppb/hr. During high- O_3 events, CHEM is the most important process causing higher O_3 levels over water relative to clean days, followed by vertical DIF (**Figure 50b**). We found that O_3 across the entire profile is higher on episode days than on clean days, indicating an elevated O_3 background on high- O_3 days which is consistent with the [BGD] simulation from CAMx (c.f. Figure 38 and Figure 40). In addition, the O_3 gradient above and below the PBL is also higher on episode days, especially during morning hours, which can induce more vertical diffusion if downmixing occurs from above the PBL when the capping inversion is weak.

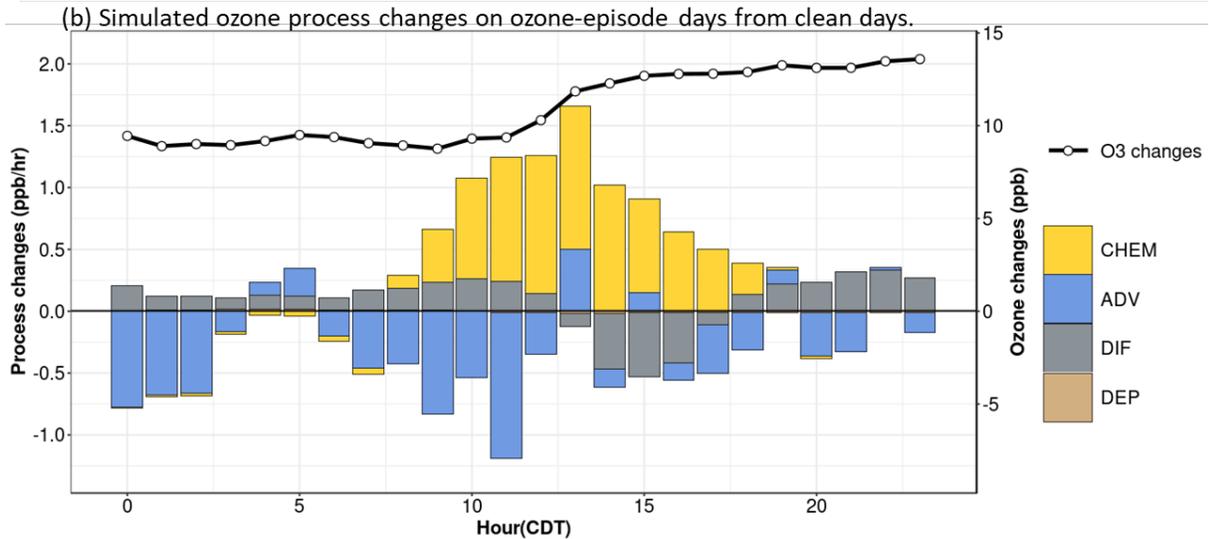
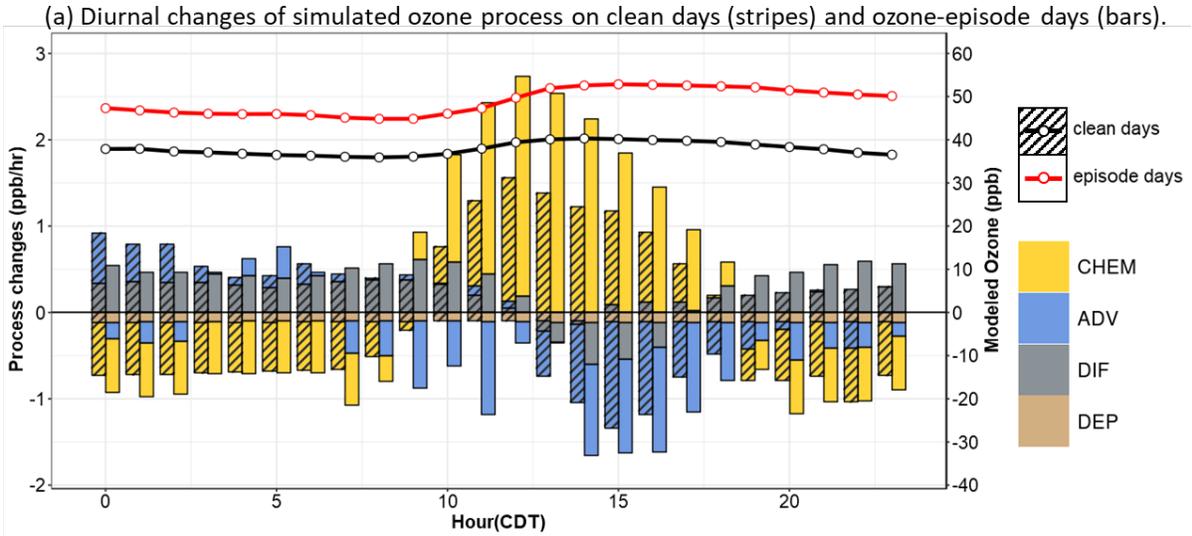


Figure 50. (a) Diurnal changes of simulated ozone processes over the Gulf of Mexico, including chemistry (CHEM), advection (ADV), vertical diffusion (DIF), and deposition (DEP) on clean days (stripes) and O₃-episode days (bars) integrated across the lowest five model layers. Overlaid lines and points are simulated hourly ozone on clean (black) and O₃-episode (red) days. (b) Process (filled bars) and O₃ (black line) changes during high-O₃ episodes relative to clean days.

Since transport patterns vary on an hourly and daily basis, the averages of multiple-day results shown in **Figure 50** may not be specific for individual hours/days with ozone exceedance, particularly for transport processes such as advection and diffusion. Therefore, we conducted the process analysis on a case-study day (September 9, 2021) over the Gulf of Mexico and the results are shown in **Figure 51**. It shows that ADV, in addition to CHEM, contributes to the enhanced O₃ levels at 10:00 and 13:00 (**Figure 51**), which respectively corresponds to northerly and easterly winds in the model and highlights the importance of regional transport. This demonstrates that the contributions from ADV to the increase of O₃ can be high in some specific cases, although its mean contributions over multiple days are averaged out in **Figure 50**.

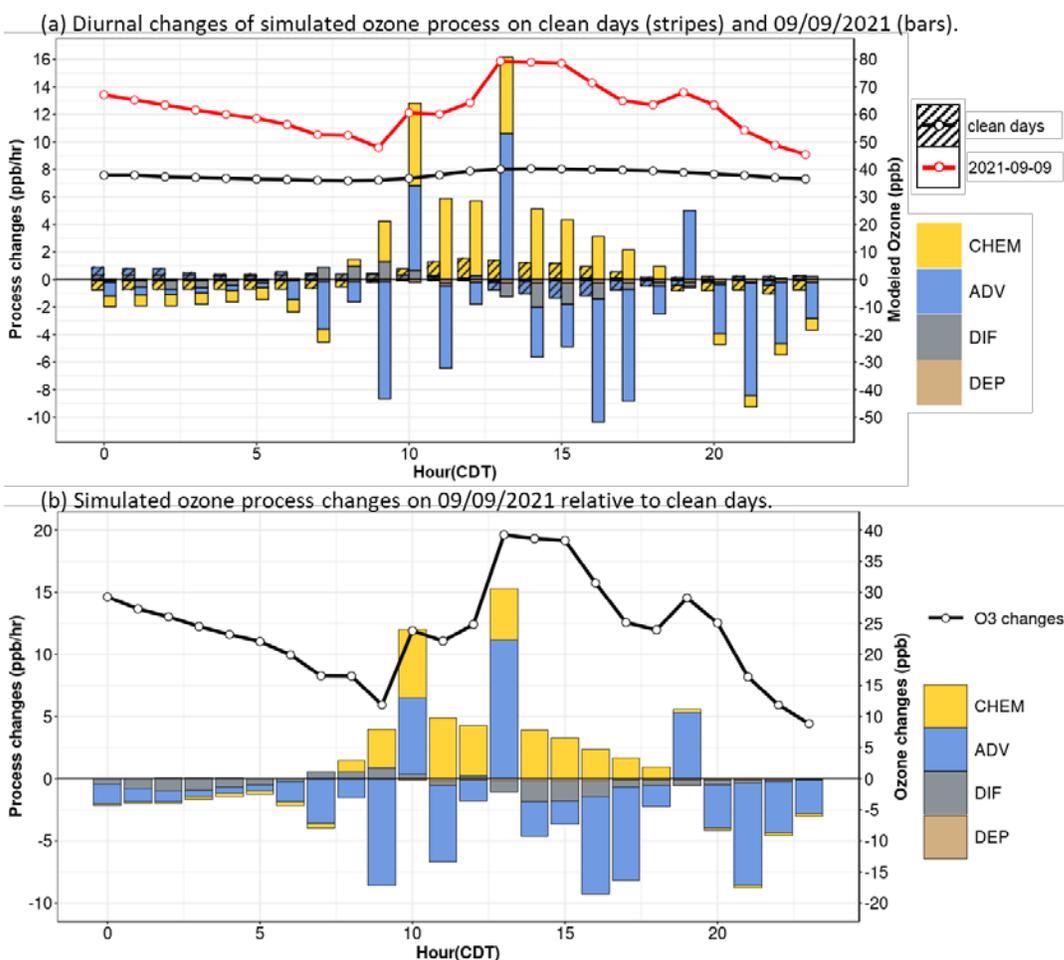


Figure 51. (a) Diurnal changes of simulated ozone processes over the Gulf of Mexico, including chemistry (CHEM), advection (ADV), vertical diffusion (DIF), and deposition (DEP) on clean days (stripes) and 09/09/2021 (bars) integrated across the lowest five model layers. Overlaid lines and points are simulated hourly ozone on clean (black) and 09/09/2021 (red) days. (b) Process (filled bars) and O₃ (black line) changes on 09/09/2021 relative to clean days.

5.4 Summary

Through zeroing-out local anthropogenic emissions in two models (WRF-GC and CAMx), we found local anthropogenic emissions contribute to less than 20% of surface ozone in Houston during the episode periods in September 2021. This contribution is smaller in WRF-GC (13%) than CAMx (19%), with the former predicting 15 ppbv high background ozone than the latter. Given the large contributions of non-local emissions to surface ozone in Houston, there is a need to better model natural emissions as well as regional transport dynamics for improving ozone simulation.

The 10% soft perturbation experiments of land emissions yielded less than 0.5% changes in surface ozone in WRF-GC and CAMx in urban Houston. The smaller response in WRF-GC is consistent with the fact that this model has a larger background contribution. In both models, offshore ozone does not decrease as much in response to 10% emission reductions of land-based emissions or even increase because of the offset effect of increasing the lifetime of primary

pollutants leading to more in situ ozone production over the waters. A 10% reduction in over-water NO_x and NMVOC emissions has essentially no impact on over-water ozone (less than 0.01% change). This finding reveals the resiliency of high ozone over water to small changes in land emissions or over-water emissions, making it difficult to control.

CAMx process analysis revealed that in situ production of ozone over the waters is the leading process responsible for net ozone increase at offshore locations during episode days relative to clean days, followed by vertical diffusion. Considering the small change of over-water ozone to 10% emission reductions of local emissions on land and over water, we suspect that ozone precursors originating from outside Houston are the key contributor to in situ ozone production at offshore locations. The contributions from direct advection can be high in some specific cases, but its mean contributions over multiple days are averaged out and present a small ozone sink at offshore locations.

6. Summary and Future Directions

The TRACER-AQ and GO_3 field campaigns in September 2021 provided unprecedentedly rich observations of ozone air pollution covering both offshore and onshore locations that are needed to validate current air quality models. Utilizing these datasets, the project revealed strengths and definitive gaps in the WRF model and WRF-driven photochemical models in replicating the meteorological and chemical observations in Houston and surrounding waters. The project activities were conducted to answer the following primary science questions:

Q1. Which configurations and simulation settings of WRF most accurately replicate the extensive meteorological data collected as part of TRACER-AQ?

Answer: The project evaluated ten WRF configurations of different meteorological inputs, physics options, and data assimilation options against both onshore and offshore observations. While different WRF configuration has its own advantage in reducing model biases, [HRRR], [Nudged2], and [Reinit] are found to be the three best simulations based on campaign-wide statistics. Considering that [Nudged2] requires additional efforts to prepare observational datasets and [Reinit] needs to automate the model running process, [HRRR] is the easiest but the most effective option to reproduce meteorology during the TRACER-AQ 2021 campaign.

Q2. How well do coupled mesoscale meteorological and photochemical grid modeling of coastal/maritime boundary layers replicate observations?

Answer: The WRF model, regardless of configuration settings, shows persistent low biases in PBL heights. Compared to PBL derived from the HSRL-2 flying over urban Houston and Galveston Bay for ten days, the model simulations capture spatial variabilities at noon ($R=0.62-0.77$) and in the afternoon ($R=0.71-0.76$) but have difficulties capturing that in the morning ($R=-0.1-0.14$) due to the presence of complex nocturnal PBL structure with both residual layer and stable surface layer. Despite less spatial correlation in the morning, land-water differences are well represented by the model throughout the day, with lower PBL heights over water than on land. Compare to PBL derived from a ceilometer on a mobile boat over Galveston Bay, the model captures the low PBL in the morning but has difficulties capturing the high PBL in the afternoon, leading to low correlation among different configurations ($R=0.25-0.41$) and a low bias of 184 m.

Q3. How well do photochemical grid models predict over-water concentrations and formation rates of ozone?

Answer: All three models evaluated captured high ozone over water during ozone episode periods. However, all models underestimate offshore ozone during episode periods. WRF-GC has the lowest bias, while CAMx has the highest correlation with offshore ozone. WRF-Chem is best at capturing free-troposphere ozone plumes. The model's performance decreases during clean days, with reduced R and a substantial high bias of 7-10 ppbv. CAMx process analysis predicts a 2-3 ppbv/hr chemical production of ozone at offshore locations between noon and 3 PM during episode days, compared to 1-1.5 ppbv/hr during clean days.

Q4. What are the spatial distributions of ozone and ozone precursors during TRACER-AQ on days with on-land monitors recording exceedances of the NAAQS and how well does the photochemical model predict such distributions?

Answer: TRACER-AQ observations show high surface ozone in southwest Houston, the Ship Chanel, and over water during days with on-land monitors recording exceedances. All three models successfully reproduce such patterns, predicting high ozone over water during exceedance days. However, there are large model-by-model differences in the magnitude of over-water ozone enhancement between clean and episode days. CAMx has the best performance for on-land distributions, while WRF-GC is best at offshore distributions. WRF-Chem does not capture high ozone in southwest Houston.

Q5. Which emission source categories affect ozone formation over Galveston Bay and the Gulf of Mexico?

Answer: Through zeroing-out local anthropogenic emissions in two models (WRF-GC and CAMx), we found local anthropogenic emissions contribute to less than 20% of surface ozone over Galveston Bay and the Gulf of Mexico during the episode periods in September 2021. The 10% soft perturbation experiments of local emissions from land and over water further confirm this finding. Offshore ozone decreases less than 0.5% in response to 10% emission reductions of land-based emissions or even increases because of the offset effect of increasing the lifetime of primary pollutants leading to more in situ ozone production over the waters. A 10% reduction in over-water NO_x and NMVOC emissions has essentially no impact on over-water ozone (less than 0.1% change). Ozone precursors originating from outside Houston are likely the main contributor to in situ ozone production at offshore locations. This finding reveals the resiliency of high ozone over water to small changes in land emissions or over-water emissions, making it difficult to control.

The project results suggest three broad directions for future improvements in meteorological and photochemical models. First, meteorological simulations can be further improved, as the challenges observed in simulating winds and PBL heights across all available WRF configurations. Second, background ozone concentrations in photochemical models need to be constrained. Our analysis demonstrates that background ozone between WRF-GC and CAMx differ by around 10 ppbv during ozone episode periods, which explains inter-model differences of ozone simulation. Furthermore, improving the simulations of natural emissions and regional transport is crucial to better understand their contributions to elevated background ozone during ozone episodes. For example, CAMx lacks representation of lightning NO_x and shows significantly lower isoprene and soil NO_x compared to WRF-GC, resulting in lower background ozone levels. Given the fact that WRF-GC overestimates ozone during clean days, we think the natural emissions could be overestimated in this model.

References

- Banta, R., Senff, C., Nielsen-Gammon, J., Darby, L., Ryerson, T., Alvarez, R., Sandberg, S., Williams, E., Trainer, M., 2005. A bad air day in Houston. *Bulletin of the American Meteorological Society* 86, 657-670.
- Berlin, S.R., Langford, A.O., Estes, M., Dong, M., Parrish, D.D., 2013. Magnitude, decadal changes, and impact of regional background ozone transported into the Greater Houston, Texas, area. *Environmental science & technology* 47, 13985-13992.
- Burkholder, J. B., Sander, S. P., Abbatt, J. P. D. A. D., Barker, J. R., Huie, R. E., Kolb, C. E., et al. (2019). *Chemical Kinetics and Photochemical Data for Use in Atmospheric Studies: Evaluation number 18*. JPL Publication 15-10, Jet Propulsion Laboratory, Pasadena, CA.
- Caicedo, V., Rappenglueck, B., Cuchiara, G., Flynn, J., Ferrare, R., Scarino, A., Berkoff, T., Senff, C., Langford, A., Lefer, B., 2019. Bay and sea-breeze circulations impacts on the planetary boundary layer and air quality from an observed and modeled DISCOVER-AQ Texas case study. *Journal of Geophysical Research: Atmospheres*.
- Chen, F. and Dudhia, J., 2001. 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), pp.587-604.
- Emmons, L. K., Walters, S., Hess, P. G., Lamarque, J.-F., Pfister, G. G., Fillmore, D., Granier, C., Guenther, A., Kinnison, D., Laepple, T., Orlando, J., Tie, X., Tyndall, G., Wiedinmyer, C., Baughcum, S. L., and Kloster, S.: Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4), *Geosci. Model Dev.*, 3, 43–67, <https://doi.org/10.5194/gmd-3-43-2010>, 2010.
- Gettelman, A., Mills, M. J., Kinnison, D. E., Garcia, R. R., Smith, A. K., Marsh, D. R., Tilmes, S., Vitt, F., Bardeen, C. G., McInerney, J., Liu, H. L., Solomon, S. C., Polvani, L. M., Emmons, L. K., Lamarque, J. F., Richter, J. H., Glanville, A. S., Bacmeister, J. T., Phillips, A. S., Neale, R. B., Simpson, I. R., DuVivier, A. K., Hodzic, A., and Randel, W. J.: The Whole Atmosphere Community Climate Model Version 6 (WACCM6), *J Geophys Res: Atmospheres*, 124, 12 380–12 403, doi:10.1029/2019JD030943, 2019.
- Goldberg DL, Harkey M, de Foy B, Judd L, Johnson J, Yarwood G, et al. Evaluating NO_x emissions and their effect on O₃ production in Texas using TROPOMI NO₂ and HCHO. *Atmospheric Chemistry and Physics* 2022; 22: 10875-10900.
- Hong, S.Y., Noh, Y. and Dudhia, J., 2006. A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly weather review*, 134(9), pp.2318-2341.
- Hong, S.Y. and Lim, J.O.J., 2006. The WRF single-moment 6-class microphysics scheme (WSM6). *Asia-Pacific Journal of Atmospheric Sciences*, 42(2), pp.129-151.
- Hudman, R.C., N.E. Moore, R.V. Martin, A.R. Russell, A.K. Mebust, L.C. Valin, and R.C. Cohen (2012), A mechanistic model of global soil nitric oxide emissions: implementation and space based-constraints, *Atm. Chem. Phys.*, 12, 7779-7795, doi:10.5194/acp-12-7779-2012.
- Jin, X., Fiore, A. M., Boersma, K. F., De Smedt, I., and Valin, L. (2020): Inferring changes in summertime surface ozone-NO_x-VOC chemistry over U.S. urban areas from two decades of satellite and ground-based observations, *Environ. Sci. Technol.*, 54, 6518– 6529, <https://doi.org/10.1021/acs.est.9b07785>.
- Kommalapati, R. R., Liang, Z., & Huque, Z. (2016). Photochemical model simulations of air quality for Houston–Galveston–Brazoria area and analysis of ozone–NO_x–hydrocarbon

- sensitivity. *International Journal of Environmental Science and Technology*, 13(1), 209–220.
- Li, W., Y. Wang, Bernier, C., & Estes, M., 2020. Identification of sea breeze recirculation and its effects on ozone in Houston, TX, during DISCOVER-AQ 2013. *Journal of Geophysical Research: Atmospheres*, 125, <https://doi.org/10.1029/2020JD033165>
- Morcrette, J.J., Barker, H.W., Cole, J.N.S., Iacono, M.J. and Pincus, R., 2008. Impact of a new radiation package, McRad, in the ECMWF Integrated Forecasting System. *Monthly weather review*, 136(12), pp.4773-4798.
- Morrison, H., Thompson, G. and Tatarskii, V., 2009. Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one-and two-moment schemes. *Monthly weather review*, 137(3), pp.991-1007.
- Murray, L. T., Jacob, D. J., Logan, J. A., Hudman, R. C., & Koshak, W. J. (2012). Optimized regional and interannual variability of lightning in a global chemical transport model constrained by LIS/OTD satellite data. *J. Geophys. Res.*, 117(D20), D20307. <https://doi.org/10.1029/2012jd017934>.
- Nakanishi, M. and Niino, H., 2009. Development of an improved turbulence closure model for the atmospheric boundary layer. *Journal of the Meteorological Society of Japan. Ser. II*, 87(5), pp.895-912.
- Pan, S., Choi, Y., Roy, A., Li, X., Jeon, W., & Souri, A. H. (2015). Modeling the uncertainty of several VOC and its impact on simulated VOC and ozone in Houston, Texas. *Atmospheric Environment*, 120, 404–416.
- Pleim, J.E., 2007. A combined local and nonlocal closure model for the atmospheric boundary layer. Part II: Application and evaluation in a mesoscale meteorological model. *Journal of Applied Meteorology and Climatology*, 46(9), 1396-1409.
- Rappenglück, B., Perna, R., Zhong, S., & Morris, G. A., 2008. An analysis of the vertical structure of the atmosphere and the upper-level meteorology and their impact on surface ozone levels in Houston, Texas. *Journal of Geophysical Research*, 113(D17). doi:10.1029/2007jd009745
- US Environmental Protection Agency, Air Quality Criteria for Ozone and Related Photochemical Oxidants (Final), Vols. I, II, and III, EPA 600/R-05/004aF-cF, 2006.
- Wang, Y., Jia, B., Wang, S.-C., Estes, M., Shen, L., Xie, Y., 2016. Influence of the Bermuda High on interannual variability of summertime ozone in the Houston–Galveston–Brazoria region. *Atmospheric Chemistry and Physics* 16, 15265-15276
- Wesely, M. L. (1989). Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models. *Atmospheric Environment*, 41, 52–63.
- Wiedinmyer, C. and Emmons, L.: Fire Inventory from NCAR version 2 Fire Emission, <https://doi.org/10.5065/XNPA-AF09>, 2022.
- Xiao, X., Cohan, D. S., Byun, D. W., & Ngan, F. (2010). Highly nonlinear ozone formation in the Houston region and implications for emission controls. *Journal of Geophysical Research: Atmospheres*, 115(D23).
- Zaveri, R. A., Easter, R. C., Fast, J. D., and Peters, L. K.: Model for Simulating Aerosol Interactions and Chemistry (MOSAIC), *J. Geophys. Res.-Atmos.*, 113, D13204, <https://doi.org/10.1029/2007jd008782>, 2008.