Atmospheric
Chemistry
and Physics

An intercomparison of satellite, airborne, and ground-level observations with WRF–CAMx simulations of NO$_2$ columns over Houston, Texas, during the September 2021 TRACER-AQ campaign

M. Omar Nawaz[1], Jeremiah Johnson[2], Greg Yarwood[2], Benjamin de Foy[3], Laura Judd[4], and Daniel L. Goldberg[1]

[1]Department of Environmental and Occupational Health, George Washington University, Washington, DC 20052, USA
[2]Ramboll, Novato, CA 94945, USA
[3]Department of Earth and Atmospheric Sciences, Saint Louis University, St. Louis, MO 63103, USA
[4]NASA Langley, Hampton, VA 23666, USA

Correspondence: M. Omar Nawaz (nawaz.muhammad@email.gwu.edu)

Abstract. Nitrogen dioxide (NO$_2$) is a precursor of ozone (O$_3$) and fine particulate matter (PM$_{2.5}$) – two pollutants that are above regulatory guidelines in many cities. Bringing urban areas into compliance of these regulatory standards motivates an understanding of the distribution and sources of NO$_2$ through observations and simulations. The TRACER-AQ campaign, conducted in Houston, Texas, in September 2021, provided a unique opportunity to compare observed NO$_2$ columns from ground-, airborne-, and satellite-based spectrometers. In this study, we investigate how these observational datasets compare and simulate column NO$_2$ using WRF–CAMx with fine resolution (444×444 m^2) comparable to the airborne column measurements. We compare WRF-simulated meteorology to ground-level monitors and find good agreement. We find that observations from the GEOstationary Coastal and Air Pollution Events (GEO-CAPE) Airborne Simulator (GCAS) instrument were strongly correlated ($r^2 = 0.79$) to observations from Pandora spectrometers with a slight high bias (normalized mean bias (NMB) $= 3.4$ %). Remote sensing observations from the TROPOspheric Monitoring Instrument (TROPOMI) were generally well correlated with Pandora observations ($r^2 = 0.73$) with a negative bias (NMB $= -22.8$ %). We intercompare different versions of TROPOMI data and find similar correlations across three versions but slightly different biases (from -22.8 % in v2.4.0 to -18.2 % in the NASA MINDS product). Compared with Pandora observations, the WRF–CAMx simulation had reduced correlation ($r^2 = 0.34$) and a low bias (-21.2 %) over the entire study region. We find particularly poor agreement between simulated NO$_2$ columns and GCAS-observed NO$_2$ columns in downtown Houston, an area of high population and roadway densities. These findings point to a potential underestimate of NO$_x$ emissions (NO$_x = $ NO $+$ NO$_2$) from sources associated with the urban core of Houston, such as mobile sources, in the WRF–CAMx simulation driven by the Texas state inventory, and further investigation is recommended.

1 Introduction

Nitrogen dioxide (NO$_2$) is a critical precursor to criteria air pollutants (i.e., ozone or "O$_3$" and fine particulate matter or "PM$_{2.5}$") that are above regulatory thresholds in many urban areas. Exposure to NO$_2$ is also directly associated with asthma exacerbation in vulnerable groups (Achakulwisut et al., 2019; Anenberg et al., 2022) and premature death (Huang et al., 2021). Due to its short atmospheric lifetime (de Foy et al., 2014), observations of NO$_2$ can reveal fine-scale patterns associated with sources. A major source of NO$_2$ is fossil-fuel combustion (McDuffie et al., 2020), and in many urban airsheds this is the dominant contributor to NO$_2$; however, other natural sources – like lightning (Murray, 2016) and soil microbes (Hudman et al., 2012) – along with fires (Jin et al., 2021) and tropospheric–stratospheric NO$_2$ exchange also contribute to tropospheric NO$_2$ levels. The health burden, sources, and short atmospheric lifetime of NO$_2$ all compound in urban environments where there are large populations, diverse contributors, and unique fine-scale patterns in NO$_2$ levels.

In the US city of Houston, Texas – the fifth most populous metropolitan region in the United States (U.S. Census Bureau, 2023) – NO$_2$ is a major concern (Mazzuca et al., 2016) due to its role as a precursor of the formation of O$_3$ and PM$_{2.5}$. While NO$_2$ itself nor PM$_{2.5}$ exceeds their respective US EPA National Ambient Air Quality Standards (NAAQS), Houston is in moderate nonattainment of the 8 h ozone (2015) NAAQS. The large petrochemical industry in Houston emits NO$_2$ in addition to other common heavy emitting sources associated with coastal urban environments like vehicles, power stations, and shipping channels (Kim et al., 2011). The co-location of this large population with high levels of NO$_2$ presents a major public health concern that motivates research to better understand the sources that are most culpable in contributing to air pollution. Major interstate highways like I-610 and I-10, as well as Beltway 8, have heavy vehicle traffic that is responsible for elevated NO$_2$ concentrations (Miller et al., 2020). Large power stations and industrial facilities operate within and around the Houston metropolitan area, and these point sources – along with a large shipping channel – are responsible for NO$_2$ plumes (Luke et al., 2010). Characterizing the unique imprints of these disparate sources remains a question of scientific concern. There is also evidence that low-income and non-white populations in Houston are disproportionally affected by air pollutants such as NO$_2$ (Demetillo et al., 2020).

Synchronous observations of NO$_2$ column densities from aircraft, ground-based, and satellite spectrometers coincided in September 2021 during the Tracking Aerosol Convection Interactions Experiment – Air Quality (TRACER-AQ). This campaign provided a unique opportunity to investigate the fine-scale patterns in NO$_2$ levels in Houston. One of the devices employed during the TRACER-AQ campaign across its 12 flight days was the Geostationary Coastal and Air Pollution Events (GEO-CAPE) Airborne Simulator (GCAS) instrument that has been discussed in many previous studies (e.g., Judd et al., 2020; Kowalewski and Janz, 2014; Leitch et al., 2014; Nowlan et al., 2018). The GCAS instrument is an ultraviolet–visible (UV–VIS) spectrometer. Its data are used to retrieve NO$_2$ columns over a limited number of flight days. This made its observational average more sensitive to meteorological conditions than an instrument with a longer time record; however, this tool observes NO$_2$ patterns with uniquely fine-scale resolution (on average $560 \times 250 \, \text{m}^2$) and performed comprehensive measurements of NO$_2$ columns across large swaths of the city repeatedly up to three times per day. This differs with observations from the TROPOMI instrument on board the Copernicus Sentinel-5 Precursor (S5P) satellite that is in a near-polar sun synchronous orbit (van Geffen et al., 2022) that only observes NO$_2$ once per day in the early afternoon at a coarser resolution of $3.5 \times 5.5 \, \text{km}^2$ at nadir. TROPOMI and GCAS spectra are used to retrieve slant NO$_2$ columns that are converted into vertical columns using an air mass factor (AMF) (Palmer et al., 2001) which is the largest source of uncertainty in the tropospheric vertical column retrieval algorithm (Lorente et al., 2019). Comparing TROPOMI data to other observations – like those from aircraft or ground-based monitors – can serve as a useful diagnostic tool in characterizing its performance and potential biases. These characterizations have large-scale implications since TROPOMI measures NO$_2$ columns globally and is useful in areas that lack the observational infrastructure of other instruments. The Pandonia Global Network (PGN) is a network of Pandora instruments (Herman et al., 2009). These instruments are UV–VIS spectrometers that measure spectrally resolved radiance data that are used to retrieve total vertical NO$_2$ columns. A total of seven Pandora instruments were operational during the TRACER-AQ campaign across three separate sites in and around downtown Houston.

The Comprehensive Air Quality Model with Extensions (CAMx) is a multi-scale photochemical model that can simulate air pollutants including ozone, fine particulate matter, and NO$_2$ (Ramboll, 2022b). CAMx has been used extensively to investigate Texas air quality by leveraging model input data created by the Texas Commission on Environmental Quality (TCEQ) for air quality planning (Ge et al., 2021; Goldberg et al., 2022) with strong performance compared to that of remote sensing column concentrations in Texas (Goldberg et al., 2022; Li et al., 2023; Soleimanian et al., 2023). CAMx can be coupled with meteorological models like the Weather Research and Forecasting Model (WRF) which provide the meteorological inputs necessary to simulate fine-scale atmospheric conditions (Jia et al., 2017). This coupled modeling system is denoted as WRF–CAMx. Fine-scale simulations from WRF–CAMx are useful in helping to understand biases in simulated NO$_2$ and in identifying under- or overestimates of emissions from sectors and regions in the inventories that drive the model.

In this study, we leverage the unique coincidence of ground-based Pandora spectrometers as well as high-resolution airborne and TROPOMI-based remote sensing observations of column NO$_2$ during the September 2021 TRACER-AQ campaign (Judd et al., 2021). We assess the capabilities of these different data sources through cross-comparisons and then compare observed NO$_2$ to simulated values from a WRF–CAMx simulation to evaluate its performance. Additionally, we consider the impact of different TROPOMI algorithms on performance against Pandora measurements. Our comparisons across the three observational datasets clarify the range of expected values of NO$_2$ column concentrations in Houston and characterize potential deficiencies and biases in observational products and simulated CAMx values. We investigate weekday and weekend performance of the model and consider differences in the spatial distributions of tropospheric NO$_2$ columns to qualitatively identify the sources that may be under- or overestimated in local inventories and identify the regions in Houston that are most likely impacted by these incorrectly attributed emissions. Additionally, we compare diurnal profiles in column and surface concentrations of NO$_2$ across relevant products.

2 Data and methods

2.1 Pandora observations

During the TRACER-AQ campaign, a total of seven Pandora instruments operated across three sites in Houston (Table 1). Pandora instruments are ground-based UV–VIS spectrometers that measure spectrally resolved radiances, and this work only utilizes those collected via direct-sun observations (Herman et al., 2009). Trace gas spectral fitting routines are employed similarly to remote sensing and aircraft observations (Judd et al., 2020) to characterize column concentrations of gases (e.g., NO$_2$). Details on the Pandora instruments and their fitting routines are discussed in detail in past studies (Cede, 2021; Herman et al., 2009). The study was designed to have two Pandoras operating coincidently at each site during the campaign; however, due to instrument failures, an uneven number of observations were obtained at each site. In order to evenly weigh the observations between the three sites, we selected data from a single Pandora instrument at each site. Pan no. 58 at La Porte, no. 61 at Aldine, and no. 25 at the University of Houston were chosen for the following reasons: as indicated in Table 1, Pan no. 61 and Pan no. 58 clearly have the largest temporal coverage during the TRACER-AQ time period. While Pan no. 188 measured more frequently at the University of Houston than Pan no. 25, Pan no. 188 was operated on a tower about 70 m above the surface, which results in missing portions of the tropospheric column when operated in direct-sun mode.

Locations of the three sites are presented in Fig. 2f. These three chosen instruments are bolded in Table 1. Pandora direct-sun retrievals represent the "total vertical column" of NO$_2$ which differs from the aircraft measurements that only measure the tropospheric column. We directly compare these disparate sources by adding a "stratospheric NO$_2$ column component" derived from TROPOMI estimates to the aircraft measurements (see Sects. 2.2 and 2.3) to compare total column amounts.

2.2 GCAS observations

The GCAS instrument was installed on the NASA G-V aircraft. The GCAS instrument employs charge-coupled device array detectors to observe backscattered light. These data can be used to retrieve column densities of gases like NO$_2$ below the aircraft using DOAS computing software (Danckaert et al., 2017). During TRACER-AQ, GCAS collected data over the Houston metropolitan area across 12 d during late August and throughout September 2021. The flight strategy of the aircraft included flying the plane in a "lawnmower" fashion with flight lines spaced 6.3 km apart, ensuring overlap at flight altitude (FL280) with the instrument field of view of 45° creating one gapless map of NO$_2$ up to three times per flight day with an average differential slant column pixel size of 250 m × 250 m. NO$_2$ observations from GCAS are publicly available at the NASA Atmospheric Sciences Data Center (NASA/LARC/SD/ASDC, 2022a). Observations from 2 of the flight days – a test flight (30 August) and a flight over the Gulf of Mexico (27 September) – are excluded from this study because they provided no meaningful data over Houston. Given the relatively short time frame of flight data collection, meteorological conditions have an influence on the fine-scale patterns in NO$_2$ columns observations. Owing to this, we summarize some basic conditions and information of the 10 flight days that focused on Houston (Table 2). Wind and meteorological conditions were determined by review of historical weather archives taken at Houston Hobby Airport (EOSDIS Worldview, 2023; Weather Underground, 2023).

The publicly available GCAS measurements (version R2) include a version of the dataset with reprocessed AMFs to include NO$_2$ vertical profile estimates from the fine-scale (444 × 444 m^2) WRF–CAMx simulation used in this analysis (Sect. 2.4). Air mass factors use this vertical profile information to account for altitude-dependent sensitivities in remote sensing observations. The original vertical profiles in the dataset were derived from a global model, GEOS-CF (Keller et al., 2021), which had a coarser spatial resolution (0.25° × 0.25°). Lastly, to directly compare GCAS measurements to other NO$_2$ column concentrations, we regrid them to a common grid; in this study, we chose the fine-scale WRF–CAMx grid. Only cloud-free GCAS data are considered in this analysis.

To characterize the accuracy and precision of GCAS measurements we compare them to observations from the Pandora instruments (Sect. 3.1). This comparison requires both spatial and temporal screening. Spatially, we restrict our comparison to only the GCAS pixels that contain Pandora

Table 1. Details on Pandora instrument locations and measurements per day. Bolded instruments Pan no. 58 at La Porte, no. 61 at Aldine, and no. 25 at the University of Houston were those selected for evaluation in this study.

Inst. no.	Location	Lat.	Long.	September 2021 flight days with observations (number of high- and medium-quality measurements per day)									
				1	3	8	9	10	11	23	24	25	26
11	La Porte	29.67	−95.06	322	347	0	0	0	0	0	0	0	0
58	La Porte	29.67	−95.06	132	190	412	319	415	362	92	439	414	401
63	La Porte	29.67	−95.06	0	0	0	0	0	0	0	0	265	207
61	Aldine	29.90	−95.33	168	253	400	391	419	367	420	420	405	420
148	Aldine	29.90	−95.33	5	1	3	1	3	3	17	17	10	17
25	U of H	29.72	−95.34	213	256	300	299	392	273	400	382	400	344
188	U of H	29.72	−95.34	528	610	1184	957	1137	722	372	749	225	95

Table 2. Basic meteorological conditions and notes during the GCAS flight days at the Houston Hobby (KHOU) airport as obtained from https://www.wunderground.com/history/monthly/us/tx/houston/KHOU/date/2021-9 (last access: June 2023).

Day of September 2021	Day of the week	High temp	Wind direction	Additional note
1	Wednesday	96 °F	Weak SW winds	Thunderstorms from S to N, 11:00–16:00 LT
3	Friday	93 °F	Weak S winds	Scattered thunderstorms 12:00–16:00 LT
8	Wednesday	94 °F	N turning NE	Clear skies and no rain
9	Thursday	95 °F	N turning NE	Afternoon fair weather clouds, no rain
10	Friday	93 °F	NE turning E	Clear skies, no rain, some long-range smoke aloft
11	Saturday	93 °F	E winds	Afternoon fair weather clouds, no rain
23	Thursday	83 °F	E winds	Clear skies, no rain, cold front overnight 21 September
24	Friday	84 °F	E turning SE	Clear skies, no rain
25	Saturday	87 °F	NE turning E	Clear skies, no rain
26	Sunday	83 °F	Calm then SE	Clear skies, afternoon fair weather clouds

instruments. Temporally, we screen out all Pandora measurements that are more than 15 min removed from a GCAS overpass and then identify the Pandora measurement time within this 30 min window that most closely matches the GCAS overpass time. While we choose this 30 min window as an upper-bound cut-off, 96 % and 90 % of all Pandora closest matches occur within a 20 and 15 min window of GCAS overpasses, respectively, indicating that this choice of window will have a minimal impact on our results. After screening the data, we also account for the fact that GCAS only measures the tropospheric component of the NO$_2$ column. There is a substantial but predictable "above-aircraft" column that is not reflected in the GCAS measurements. This is primarily associated with stratospheric NO$_2$. To account for this, we approximate the above-aircraft component of the GCAS NO$_2$ columns using the stratospheric NO$_2$ column component of TROPOMI measurements (Sect. 2.3) and add this to GCAS observations. Additionally, we add an "above-aircraft" but below-troposphere partial column amount based on the CAMx simulation, and then we calculate the column. The three highest levels of CAMx amount to 0.57×10^{15} molec. cm^{-2}, and we add this amount to GCAS.

2.3 TROPOMI observations

The TROPOMI instrument, on board the Sentinel-5P satellite, has measured total slant columns of NO$_2$ daily at approximately 13:30 local time (LT) globally from 30 April 2018 to the present (Copernicus Sentinel-5P, 2021). The slant column measurements were converted into tropospheric vertical column amounts by subtracting off a stratospheric NO$_2$ component and transforming the remaining tropospheric slant column to vertical column using an air mass factor. We downloaded the publicly available data (https://dataspace.copernicus.eu/explore-data/data-collections/sentinel-data/sentinel-5p, last access: May 2023; https://dataspace.copernicus.eu/, last access: September 2024) coincident with the TRACER-AQ campaign in September 2021 for overpasses of Houston. In this study, we primarily consider measurements from the latest version (2.4.0) (Eskes et al., 2023); however, we additionally consider measurements processed using the version 2.3.1 algorithm (van Geffen et al., 2022) and the NASA Multi-Decadal Nitrogen Dioxide and Derived Products from Satellites (MINDS) product (Lamsal et al., 2022) and intercompare these different versions (Fig. 3). All product versions

stem from the same slant column retrieval but differ in the calculation of the air mass factor for slant to vertical column conversions and, in the case of NASA MINDS, separation of the stratosphere and troposphere (Bucsela et al., 2013). The main difference between versions 2.3.1 and 2.4.0 is the use of the $0.125° \times 0.125°$ directional Lambertian equivalent reflectivity (DLER) climatology derived from TROPOMI observations which replaces an old $0.5° \times 0.5°$ Lambertian equivalent reflectivity (LER) dataset used in v2.3.1 (Eskes et al., 2023). NASA MINDS uses a geometry-dependent surface Lambertian equivalent reflectivity (GLER) product for its surface reflectivity input into the AMF calculation based on MODIS observations. The other main difference in these products includes use of different a priori NO$_2$ profiles ($1° \times 1°$ TM5-MP for v2.3.1 and v2.4.0 vs. $0.25° \times 0.25°$ GMI simulation for NASA MINDS). A comparison between TROPOMI version 2.4.0 and a MAX-DOAS network found that in moderately polluted locations TROPOMI had a median bias of -35%. A comparison between TROPOMI version 2.4.0 and PGN found a median bias of -18% over polluted stations (Lambert et al., 2023).

These publicly available TROPOMI data are further processed for this study. We screen TROPOMI measurements to consider cloud coverage and erroneous data using the recommended qa_value filter (> 0.75). We regrid the TROPOMI NO$_2$ observations (resolution of 3.5×5.5 km^2 at nadir) onto the WRF–CAMx grid (444×444 m^2). When comparing TROPOMI observations to Pandora instruments we follow the same spatial and temporal screening approach as discussed for GCAS. Spatially, we identify the CAMx grid cell in which each Pandora instrument is located and only consider TROPOMI measurements that were regridded to these grid cells. We intercompare GCAS, TROPOMI, and CAMx at this resolution but also compare the three datasets at a coarser resolution (Sect. 3.4) to account for resolution-dependent errors. Temporally, we screen out all Pandora measurements that are more than 15 min removed from a TROPOMI overpass time and then identify the Pandora measurement time within this 30 min window that most closely matches the TROPOMI overpass time. While we choose this 30 min window as an upper-bound cut-off, 100 % and 97 % of all Pandora closest matches occur within a 20 and 15 min window of TROPOMI overpasses, respectively, indicating that this choice of window will have little impact on our results. Using WRF–CAMx vertical profile information we calculate both a total and tropospheric NO$_2$ column from TROPOMI v2.4.0 measurements using new AMF derived from the WRF simulation, and we difference the total and tropospheric values to calculate a stratospheric NO$_2$ column component from TROPOMI. We take the spatial and temporal average of this stratospheric component in Houston during the TRACER-AQ campaign to calculate a constant bias correction to convert tropospheric NO$_2$ columns – from GCAS and WRF–CAMx – to quasi-total NO$_2$ columns when comparing them to total NO$_2$ column measurements
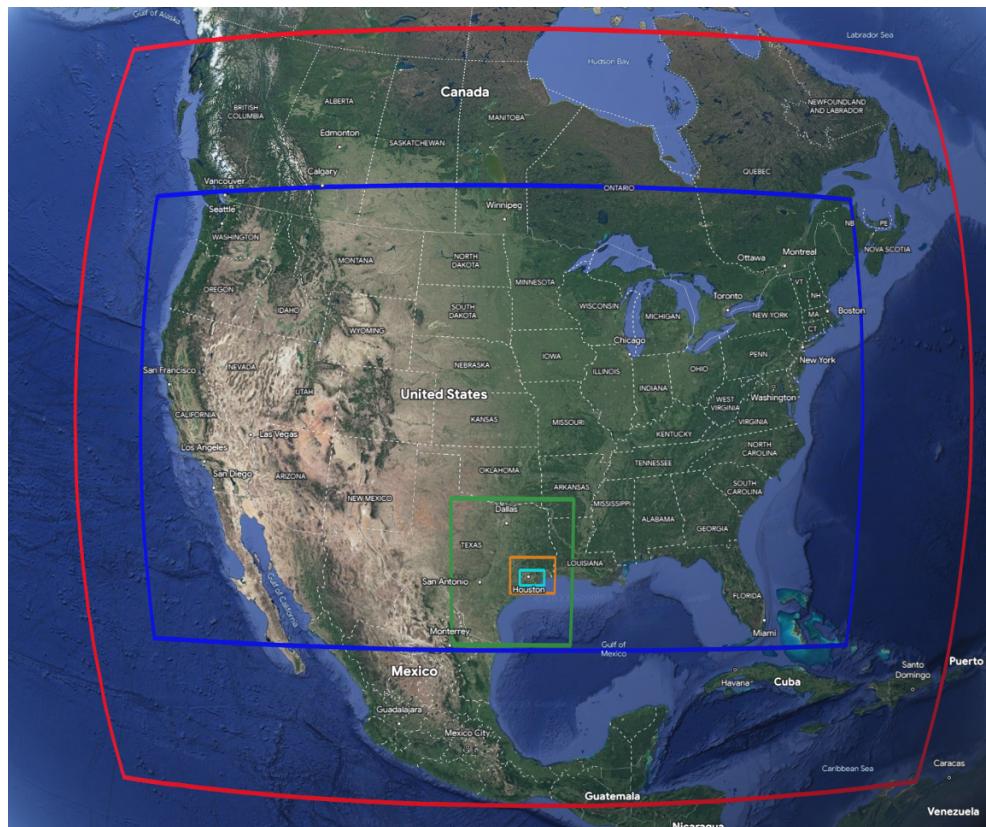
from Pandora instruments. This stratospheric vertical column NO$_2$ amount of 3.0×10^{15} is typical for Houston during summer (Geddes et al., 2018). Boersma et al. (2018) suggest that 0.5×10^{15} molec. cm^{-2} is the upper limit of structural uncertainty in the stratospheric estimate. This uncertainty should be considered when reviewing results that compare total column amounts (i.e., results comparing GCAS and CAMx to Pandora). We additionally account for diurnal variation in the stratospheric column by applying the results from work by Li et al. (2021); they calculate a daytime stratospheric NO$_2$ column increase rate of 1.34×10^{14} molec. cm^{-2} starting at 07:00 LT. We apply this increase rate by calculating the difference in hours between the dataset times – either the GCAS overpass times or CAMx simulation hours – and 13:30 LT, i.e., the approximate TROPOMI overpass time, and then multiply this difference by the increase rate. In doing so, total column values before the TROPOMI overpass are decreased and total column values after the overpass are increased.

2.4 WRF–CAMx-simulated NO$_2$

For this study, a set of simulations were conducted employing version 4.3.3 of the Advanced Research Weather Research and Forecasting (WRF) model (Skamarock et al., 2021) jointly with the Comprehensive Air Quality Model with Extensions (CAMx) v7.20 with the CB6r5 chemical mechanism for a simulation period that matched the September 2021 TRACER-AQ time frame. A new high-resolution modeling platform was designed specifically for this study that adopted prior approaches used in Texas Commission on Environmental Quality (TCEQ) state implementation plan (SIP) modeling (TCEQ, 2021) to update emissions.

The WRF model is a mesoscale numerical weather prediction system designed to serve both operational forecasting and atmospheric research needs (Skamarock et al., 2005, 2008). We define the WRF modeling domains as slightly larger than the corresponding CAMx domains (Fig. 1) to avoid possible numerical artifacts near domain boundaries when transferring WRF meteorology to CAMx. The 36 km CAMx domain (red) includes the continental United States, Mexico, and parts of Central America and Canada. The 36 km, 12 km (blue), and East Texas 4 km (green) domains are also used by the TCEQ for State Implementation Plan (SIP) modeling. The higher-resolution domains (1.333 km (orange) and 0.444 km (cyan)) were selected to include the most relevant GCAS flight tracks while considering computational expense.

Additional information on the WRF–CAMx modeling is included in the Supplement including the WRF physics options (Table S1 in the Supplement), vertical layer mapping from WRF to CAMx (Table S2), and CAMx science options (Table S3). We used 0.25° Global Forecasting System (GFS) data assimilation system (GDAS) analysis data (DOC/NOAA/NWS/NCEP/EMC, 2023) as initial conditions

Figure 1. Modeling domains used in the CAMx simulation for the 36 km resolution (red), 12 km resolution (blue), 4 km resolution (green), 1.333 km resolution (orange), and 0.444 km resolution (cyan). Map data provided by Google © 2020, Landsat/Copernicus Data SIO, NOAA, US Navy, NGA, GEBCO, IBCAO, INEGI, and US Geological Survey.

for the WRF meteorological model; these GDAS data are also used for boundary conditions and data assimilation. We configured the output time steps of WRF to 15 min for the higher-resolution domains. Conducting WRF simulations at fine spatial resolutions (i.e., 4, 1.333, and 0.444 km) requires careful consideration of physical schemes that are sensitive to grid spacing. We turn off the convective cumulus parameterization scheme for the fine grids because WRF can explicitly simulate convection for them. For coarser grids we turn on the cumulus parameterization to account for sub-grid-scale convection. The other physics options (Table S1) are kept consistent across the different resolutions. The CAMx simulation was first performed over the coarser domains (36, 12, and 4 km) from which initial and boundary conditions were extracted for the higher-resolution domains. TCEQ developed the 2019 modeling emissions inventory for the Dallas–Fort Worth (DFW) and Houston–Galveston–Brazoria (HGB) attainment demonstration (AD) SIP revisions (TCEQ, 2021). Starting with this inventory we implement further changes as discussed in the next paragraph.

First, we update the CAMx modeling emissions inventory from the TCEQ platform to incorporate 2021 hourly continuous emissions monitoring systems (CEMS) (EPA, 2023) data

for the 11 major electric generating units (EGUs) listed in Table S4. We download hourly data from Clean Air Markets Program Data (CAMPD) for the 11 EGUs for the 30 August to 27 September 2021 period, and stack parameters were based on the TCEQ 2019 emissions platform (TCEQ, 2021). Second, we update shipping emissions to incorporate MARINER v2 (Ramboll, 2022a) emissions built with 2021 automatic identification system (AIS) data for the higher-resolution domains. Third, we reprocess link-based on-road mobile emissions for the higher-resolution domains. Fourth, we update biogenic emissions and lightning NO$_x$ (LNO$_x$) based on WRF meteorology. Specifically, we use the Model of Emissions of Gases and Aerosols from Nature (MEGAN) version 3.2 (Guenther et al., 2012) for biogenic emissions, the Fire Inventory of NCAR (FINN) version 2.2 (Wiedinmyer et al., 2011) for fire emissions, and lightning NO$_x$ emissions derived by applying the CAMx LNO$_x$ processor to the 2021 meteorological data from the WRF simulation. Considering that wildfires in the Houston area are rare and that LNO$_x$ emissions are associated with convective clouds that obscure remote sensing column observations, we excluded these two emission sources from the finer resolution domains (the 1.333 and 0.444 km domains) but included them in the

Table 3. CAMx 444×444 m^2 domain-wide summary of average September weekday emissions by sector in units of metric tonnes per day (t d^{-1}).

Emission sector		Spatial resolution	NO$_x$ (t d^{-1})	VOC (t d^{-1})
EGUs		Point locations	25.5	0.2
On-road mobile		Line source	70.9	34.7
Railyards		444 m gridded	4.2	0.3
Shipping		Line source	63.9	4.3
KIAH airport		444 m gridded	6.4	0.8
KHOU airport		444 m gridded	1.8	0.4
Other	Off-road mobile	4 km gridded	33.1	31.4
	Non-EGU point sources	Point locations	47.9	27.8
	Oil and gas	4 km gridded	0.2	0.0
	Area	4 km gridded	92.8	623.2
	MEGAN biogenic	444 m gridded	25.9	319.7

larger domains. These two sources represent a small fraction of emissions in the local Houston area that are the primary focus of the finer resolution simulations. Lastly, we regrid all other gridded emissions from the coarser domains to the high-resolution domains without refining their spatial resolution. Specifically, all point sources are geolocated to the grid cell containing the source. On-road mobile source emissions and shipping emissions were provided for individual links which we allocated to 444 m grid cells and are based on known roadway networks, ship tracks, and traffic patterns. Airport and railyard emissions were allocated to 444 m grid cells within the property boundary. Other sources retained the 4 km grid resolution provided by the TCEQ. Daily emissions of NO$_x$ and volatile organic compounds (VOCs) in metric tonnes per day (t d^{-1}) for a September weekday are presented in Table 3.

We evaluate the WRF simulation meteorology by comparing surface-level wind speed, direction, temperature, and water vapor mixing ratio to observations from 16 ground-level monitors (Figs. S2–S11; Tables S5–S8) and calculate the mean bias error (MBE), mean absolute error (MAE), and Pearson R-squared (R^2) statistics as defined in Table S2. Circular statistics are calculated using the Astropy circular statistics package for Python (The Astropy Collaboration et al., 2022). We obtain integrated surface data from NCDC in the DS3505 format (https://doi.org/10.7289/V5M32STR, Vose et al., 2014). These data consist mainly of airport locations and have good meteorological siting and quality assurance procedures.

Generally, meteorological conditions simulated by WRF agree with ground-level observations especially on the more data-rich non-cloudy days, which are the most important for our intercomparison; however, performance depends on the specific measure of meteorology considered. Across all days, the WRF wind direction was well correlated ($R^2 = 0.76$) and had minimal bias (MBE $= 8°$) but some unsystematic errors (MAE $= 26°$) compared with observations. This indicates that the model generally captures variability in wind direction without a notable bias; however, considering any individual observation the simulated direction may differ by 20–30°. For non-cloudy days – which are more relevant for our intercomparisons due to more data – correlation for wind direction was similar ($R^2 = 0.73$), and the bias and error were reduced (MBE $= -5°$ and MAE $= 21°$). Simulations of wind speed were more poorly correlated ($R^2 = 0.26$) and had some unsystematic error (MAE $= 1.20$ m s^{-1}); however, there was very little systematic bias in the wind speed simulation (MBE $= -0.02$ m s^{-1}). Correlation and unsystematic errors improve on the non-cloudy days ($R^2 = 0.37$ and MAE $= 1.08$), while there is still no notable systematic bias (MBE $= -0.13$ m s^{-1}). Considering wind speeds at 09:00 and 13:00 LT (Figs. S2–S11), it appears that observations in the afternoon degrade correlation compared with the morning and that, generally, simulated wind speeds are better correlated with observations in downtown Houston than in the southeastern part of the domain near Galveston Bay. Comparisons between GCAS observations and WRF–CAMx simulations show that the model represents the dominant direction and dispersion of identifiable plumes from known sources. The wind speed bias is sufficiently low that model uncertainty will not lead to systematic errors in plume advection. Across the 8 non-cloudy days, hourly and site-specific – across the 16 monitors – WRF wind direction ($R^2 = 0.3$–0.8; MAE $= 14$–$32°$), wind speed ($R^2 = 0.1$–0.5; MAE $= 0.94$–1.35 m s^{-1}), temperature ($R^2 = 0.69$–0.81; MAE $= 0.93$–1.18 K), and water vapor mixing ratio ($R^2 = 0.28$–0.78; MAE $= 0.87$–3.11 g kg^{-1}) performed moderately compared with observations given the fine spatial and temporal resolution. Additionally, we compared simulated hourly NO$_2$ (Fig. S12 in the Supplement) and maximum daily 8 h average or "MDA8" O$_3$ (Fig. S13) to observations from 17 TCEQ continuous air monitoring stations (CAMS) operating in

Houston. We found poor performance and a strong negative bias in the simulated surface-level NO$_2$ (normalized mean bias (NMB) $= -59$ %), while simulated surface-level MDA8 O$_3$ had a much weaker bias (NMB $= -2.5$ %) compared with observations. Comparisons to ozonesondes (Figs. S14–S18) suggest that WRF simulates more aggressive vertical mixing than what is observed. This is consistent with our findings of a stronger negative bias at the surface level than for the columns, as emitted NO$_2$ at the surface is advected vertically quicker in WRF–CAMx than in reality.

2.5 Diurnal comparison

We further intercompare these data by grouping them at locations and then calculating their average diurnal profiles during the TRACER-AQ campaign for both column- and surface-level NO$_2$. Specifically we compare GCAS, CAMx, Pandora, and GEOS-CF (Keller et al., 2021) NO$_2$ columns at the three Pandora sites during TRACER-AQ flight days. We include NO$_2$ data from GEOS-CF – that will be used for processing NO$_2$ remote sensing observations from the NASA Tropospheric Emissions: Monitoring of Pollution (TEMPO) mission – to characterize differences between a global simulation and our regional WRF–CAMx modeling. Simulated surface and NO$_2$ columns from GEOS-CF are obtained through the GMAO OPeNDAP interface (https://opendap.nccs.nasa.gov/dods/) for all of 2021 and filtered to the specific Pandora instrument locations and during TRACER-AQ flight days. We apply both spatial and temporal screening. Spatially, we identify the CAMx grid cell for GCAS and CAMx, as well as the GEOS-CF grid cell in which the Pandora instrument is located. Temporally, for GCAS, we round all overpass times to the nearest hour and calculate the median value for each hour across all overpasses and days. For CAMx, GEOS-CF, and Pandora we identify the simulated and observed NO$_2$ column concentration closest to the hour and calculate the median value across all flight days and locations.
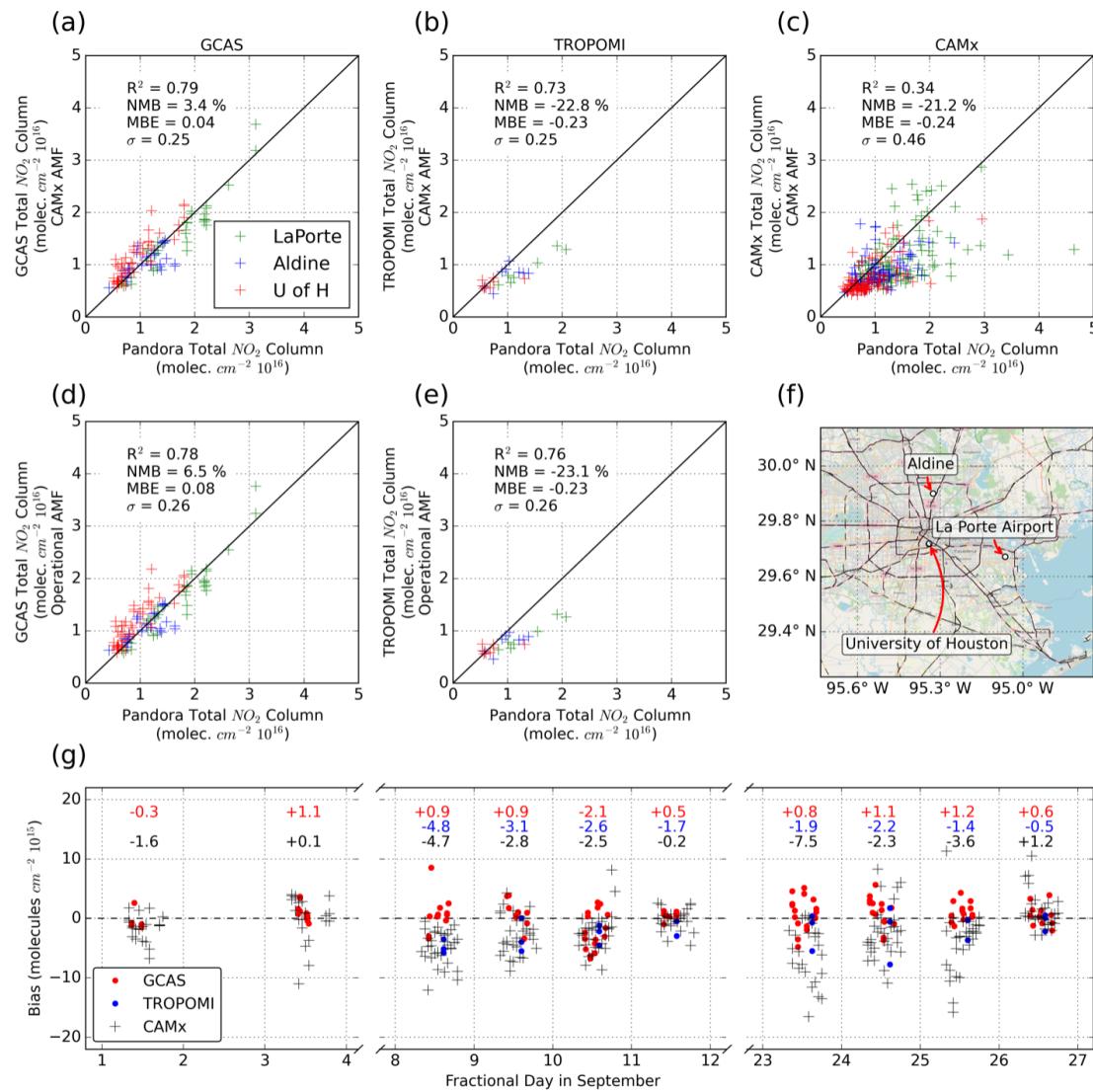
For diurnal comparisons at the surface, we use surface-level NO$_2$ concentrations from CAMx and GEOS-CF and apply the same temporal screening. Spatially, for the surface level we consider concentrations at a point in between the three Pandora instruments that is representative of downtown concentrations (29.7° N, 95.3° W). We choose this point to represent the temporal behavior of the wider regions rather than individual sites. Additionally, we download hourly NO$_2$ concentrations from the US Environmental Protection Agency (EPA) Air Quality System (AQS) (https://aqs.epa.gov/aqsweb/airdata/download_files.html, last access: July 2023). We download all hourly data for 2021 for the United States and filter the TRACER-AQ flight days and for monitors in Harris County. We identify the median hourly concentrations across these monitors and the TRACER-AQ flight days.

3 Results

3.1 Comparisons with Pandora observations

The observations from ground-based Pandora instruments are considered the most accurate of all observational platforms measuring column NO$_2$ presented in this project due to low uncertainties in their air mass factors (Herman et al., 2009) when operating in direct-sun mode. The air mass factor in this mode is calculated from simple solar geometry – unlike TROPOMI and GCAS, which rely on a priori assumptions like the vertical NO$_2$ profile and surface reflectivity. Pandora AMFs are not reliant on an a priori profile as the data we are using are only in direct-sun mode in cloud-free scenarios. AMF for Pandora is analogous to pathlength through the atmosphere relative to the vertical path. Since all of the signal is from a direct-sun path (with extremely minimal scattering), this is purely geometric. Given this, we use Pandora observations as our reference dataset to characterize the performance of the two observational datasets – GCAS and TROPOMI – along with the WRF–CAMx simulation across three sites (Table 1). These three sites (Aldine, La Porte, and the University of Houston) are located in the heavily polluted inner region of Houston that we denote as "urban Houston" (Fig. 2f). Background observations from Pandora instruments in less polluted sites were unavailable during the TRACER-AQ campaign, so there is less certainty about the performance of GCAS, TROPOMI, and CAMx outside of urban Houston. We consider the performance of GCAS processed with a CAMx-based AMF (Fig. 2a) and TROPOMI processed with a CAMx-based AMF (Fig. 2b) and CAMx (Fig. 2c), and we also consider the performance of GCAS and TROPOMI with the operational AMFs (Fig. 2d and e) individually and then intercompare the three datasets across the 10 GCAS observation days (Fig. 2g).

When comparing the observational and simulated datasets with Pandora observations, we consider the total column NO$_2$ and we add a stratospheric component from TROPOMI to the tropospheric column NO$_2$ of GCAS and CAMx to total column as discussed in the methodology. For TROPOMI, we use an AMF derived from the CAMx simulation to calculate a tropospheric NO$_2$ column from TROPOMI: following the TROPOMI user guide, we multiply the total averaging kernel by the ratio of the total air mass factor to the tropospheric air mass factor. We difference the total column NO$_2$ from TROPOMI with the tropospheric column to estimate a constant stratospheric NO$_2$ column amount that we add to GCAS and CAMx when we compare them with Pandora. This corresponds to a mean value of 3.0×10^{15}. For GCAS, we apply an additional amount, above the aircraft and below the tropopause, to account for the NO$_2$ column in the upper troposphere. We calculate the column of levels 27–29 which correspond to 9400–18 100 m above sea level – that extends roughly from the height at which GCAS flies of around 9100 m to the tropopause – and apply this to the GCAS re-

Figure 2. Comparison of Pandora total column NO$_2$ to GCAS using CAMx-based AMFs **(a)**; TROPOMI v2.4.0 using CAMx-based AMFs **(b)**; and CAMx **(c)**, GCAS **(d)**, and TROPOMI v2.4.0 **(e)** with their operational AMFs. Tropospheric columns from GCAS and CAMx are bias corrected with a TROPOMI-derived stratospheric column factor as discussed in the methodology. Data from all possible overpasses coincident within 15 min of a Pandora observation are considered. GCAS flight times generally ranged from 08:00 to 16:00 LT. TROPOMI overpasses occurred around 13:30 LT. Color coding indicates which of the Pandora instruments NO$_2$ column concentrations are being compared against as indicated in the legend in **(a)**, but statistics are presented across all locations. Map of Pandora instrument sites in urban Houston **(f)**. Bias between the three datasets and Pandora across GCAS flight days **(g)** with the overall average daily bias indicated above the points for all three datasets. The data are color-coded based on the observed or simulated source that is being compared against Pandora measurements. © OpenStreetMap contributors 2023. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.

sults; this corresponds to a value of 0.57×10^{15} molec. cm^{-2}. For all results that include comparison with Pandora we present total NO$_2$ columns, and for results where we only intercompare GCAS, TROPOMI, and CAMx, we compare the tropospheric column. All statistical measures (e.g., R^2) are defined in the Supplement.

In Fig. 2a–c we characterize the performance of the observational and simulated datasets of NO$_2$ column concen-

trations across the three sites in Houston. For each of the GCAS flight days, we compare GCAS and TROPOMI observations against Pandora measurements for every overpass that was not obstructed by cloud coverage; for CAMx we compare simulated columns for every daytime hour of each GCAS flight day.

Observations from GCAS were both well correlated ($r^2 = 0.79$) and slightly high biased (NMB $= +3.4$ %) when

compared with measurements from Pandora. Use of the CAMx AMF in place of the operational AMF had a minimal impact on comparisons with Pandora (from $r^2 = 0.78$ and NMB = +6.5 %). Observations from TROPOMI on GCAS flight days were also well correlated with Pandora measurements ($r^2 = 0.73$), but there was a negative bias (NMB = −22.8 %) in v2.4.0. This bias was worse for more NO$_2$ polluted scenarios. This negative bias may be attributable to the coarser resolution of TROPOMI compared with GCAS that weakens its ability to capture fine-scale plumes (Wagner et al., 2023) of NO$_2$ associated with road systems, airports, power stations, and industrial facilities. Similarly to GCAS, use of the CAMx AMF in place of the operational AMF for TROPOMI had a minimal impact on comparisons with Pandora (from $r^2 = 0.76$ and NMB = −23.1 %).

We calculate the ratios of the TROPOMI v2.4.0 product with the CAMx AMF compared with the operational AMF (Fig. S1) in September 2021 throughout the domain and note that tropospheric column NO$_2$ increases in the urban core and decreases in the city outskirts. The areas with Pandora instruments, in suburban Houston, have roughly equivalent values. Given that Pandora instruments were not located at either the most or least polluted areas of the metropolitan area, the benefit of the CAMx AMF may be underrepresented by our findings at the Pandora sites.

We compare simulated NO$_2$ columns from CAMx with Pandora measurements; however, in this comparison there are more points to intercompare as columns were simulated for each hour of every flight day by CAMx and observed multiple times per hour from Pandora. The CAMx-simulated columns were less correlated with Pandora measurements ($r^2 = 0.34$) than TROPOMI and GCAS, and they had a consistent negative bias (NMB = −21.2 %). This poor correlation could partially be explained by differences in WRF-simulated meteorology and observed meteorology specifically from differences in wind speed and direction and an inability to fully capture the bay breeze in Houston. We find that the WRF-simulated wind direction ($R^2 = 0.76$ and MBE = 8°), temperature ($R^2 = 0.71$ and MBE = 0.39 K), and water vapor mixing ratio ($R^2 = 0.86$ and MBE = −1.45 g kg^{-1}) (Figs. S2–S11; Tables S5–S8) are generally well correlated and minimally biased compared with observations; however, there are some unsystematic errors in wind direction (MAE = 26°) and poor correlation in wind speed ($R^2 = 0.26$) that would likely degrade correlation between observed and simulated NO$_2$ columns. While there are errors in the meteorological conditions, the biases at the surface are all small, including minimal bias in the wind speed (MBE = −0.02 m s^{-1}), indicating that the negative biases in NO$_2$ columns are likely attributable to an underestimate of NO$_x$ emissions; however, the WRF meteorological performance could partially explain the poor correlation and absolute errors in simulated NO$_2$ columns. We also note that generally, the model per-

formance is stronger on windier days, when speeds exceed 4 m s^{-1} ($R^2 = 0.5$ and 0.32), than on calmer days, when speeds are below 3 m s^{-1} ($R^2 = 0.07$, 0.1, and 0.25). Additionally, there can be substantial differences in vertical mixing coefficients in different schemes in the models, and these can impact the biases in column concentrations (de Foy et al., 2007; Riess et al., 2023). We briefly compare meteorology and the ozone mixing ratio in the WRF–CAMx simulation with ozonesonde data (https://www-air.larc.nasa.gov/cgi-bin/ArcView/traceraq.2021, last access: February 2024) and find that while temperature and pressure are captured well, there is variable performance in the vertical structure for the ozone mixing ratio, wind speed, and wind direction (Figs. S14–S18).

In Fig. 2g, we intercompare the daily variability in biases across the 10 GCAS flight days. There were no TROPOMI data for the first 2 flight days because cloud coverage blocked TROPOMI observations at the Pandora sites during its overpass time. The daily average biases of GCAS observations were consistently small throughout the entire period: they ranged from −2.1 to +1.2 × 10^{15} molec. cm^{-2} on 10 September and 3 and 24 September, respectively. TROPOMI observations were consistently biased systematically low: they ranged from −4.8 to −0.5 × 10^{15} molec. cm^{-2}; however, on all days except 26 September, the daily averaged TROPOMI biases were more negatively biased than −1.3 × 10^{15} molec. cm^{-2} compared with Pandora measurements. Unlike the two observational datasets, the bias in simulated CAMx NO$_2$ columns had much higher daily variability. On some days, such as 3 September, there was little bias in simulated columns compared with Pandora measurements, and on other days, such as 26 September, there was a minor high bias (+1.2 × 10^{15} molec. cm^{-2}); however, on most days there was a negative bias that was the strongest on 23 September when NO$_2$ columns were biased as low as −7.5 × 10^{15} molec. cm^{-2}. Generally, simulated CAMx columns perform better on weekend days (11, 25, and 26 September), which is investigated in greater detail in Sect. 3.4.

3.2 Comparisons of different TROPOMI algorithms with Pandora observations

We intercompare TROPOMI observations to Pandora measurements across three different algorithms: version 2.3.1 (Fig. 3a, d), version 2.4.0 (Fig. 3b, e), and the NASA MINDS product (Fig. 3c, f) using both the CAMx AMF (top row) and the operational AMF (bottom row) for the same Pandora instruments in Houston during the TRACER-AQ campaign. Overall, the choice of algorithm and AMF does affect the performance of TROPOMI compared with Pandora, albeit slightly. Regardless of AMF, version 2.4.0 appears to have the worst normalized mean bias in Houston during TRACER-AQ ($r^2 = 0.73$ and NMB = −22.8 %), ver-

sion 2.3.1 is improved ($r^2 = 0.72$ and NMB $= -18.3\%$), and the NASA MINDS product performs comparably ($r^2 = 0.69$ and NMB $= -18.2\%$) to version 2.3.1. Notably, NASA MINDS data for 11 September are missing, so these data are excluded from Fig. 3c and f. For version 2.3.1 and version 2.4.0 the CAMx AMF slightly improves the bias; however, for the MINDS product the CAMx AMF slightly worsens the bias compared with the operational AMF. The correlation is generally unaffected by the choice of AMF. We choose TROPOMI version 2.4.0 for the intercomparison in the following sections as it is the most recent version.

3.3 Comparisons of GCAS, TROPOMI, and CAMx data on the CAMx grid

The comparisons between Pandora measurements and the datasets indicate that GCAS observations are in best agreement with Pandora. While TROPOMI performs worse than GCAS, it still decisively outperforms simulated NO₂ columns from CAMx at Pandora sites in both correlation and bias despite its coarser resolution. With the above in mind, in this section we present NO₂ columns observed from GCAS and TROPOMI and simulated from CAMx at the $444 \times 444\,\text{m}^2$ resolution of the CAMx grid. We extend the prior comparison beyond focusing on three discrete points in urban Houston to the entire CAMx domain to get a more complete picture of the spatial components of these datasets. For each dataset we consider observations across all 10 GCAS flight days. We begin by comparing GCAS observations only with CAMx-simulated columns across all GCAS overpasses as these data are less limited temporally than TROPOMI observations (Fig. 4).

When considering data from all GCAS overpasses (Fig. 4a, b, c) we observe a consistent negative bias in the CAMx product compared with GCAS observations throughout the domain that worsens in the downtown (DT) area ($-2.7 \times 10^{15}\,\text{molec.}\,\text{cm}^{-2}$) compared with background levels in the rural East Galveston Bay (RB) ($-1.2 \times 10^{15}\,\text{molec.}\,\text{cm}^{-2}$). Near the W. A. Parish power station in the southwestern area of the domain there is a mixture of positive and negative biases in the CAMx-simulated columns that are likely indicative of errors in wind speeds or directions in the CAMx simulation. Overall, the CAMx-simulated columns were well correlated with GCAS observations ($r^2 = 0.78$), but the negative bias was substantial (NMB $= -30.6\%$) (Fig. 4d).
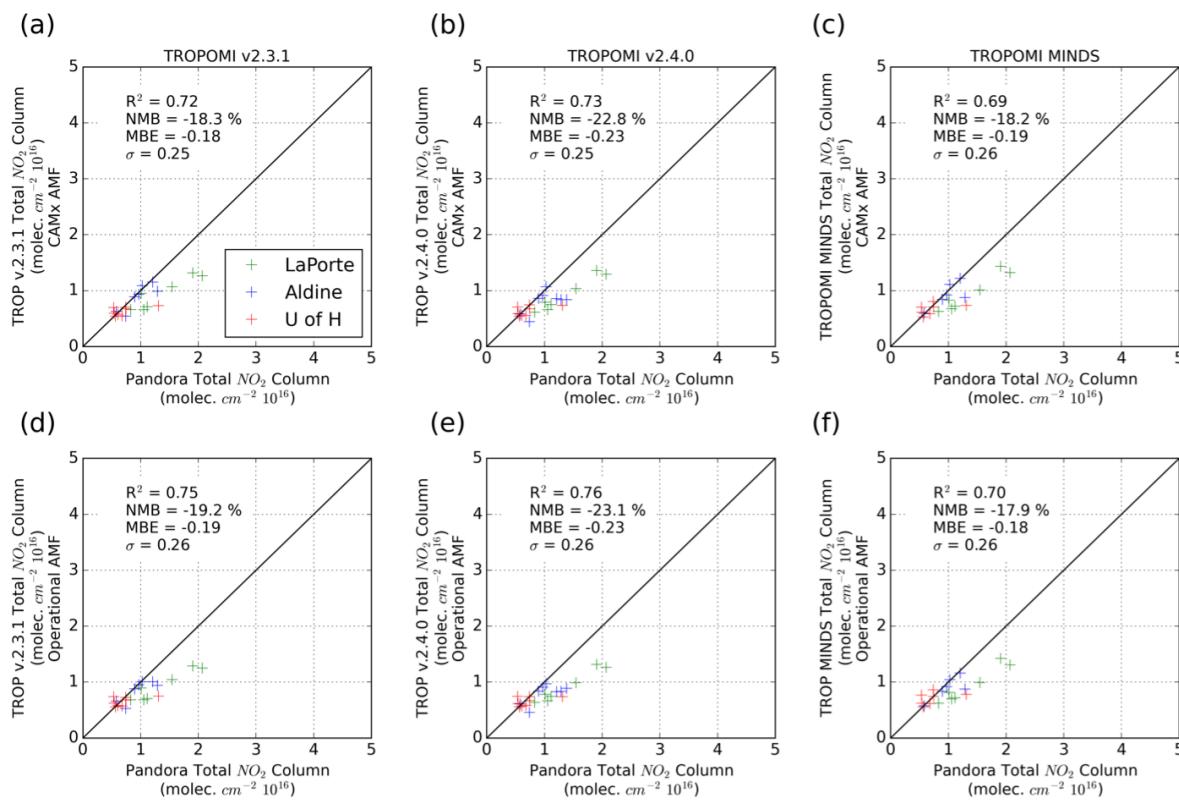
We continue this comparison in Fig. 5 where we limit the GCAS and CAMx values temporally around TROPOMI overpasses. For Fig. 5, we screen out all observations that are ± 90 min from TROPOMI overpass for each day and then temporally average the observations across the GCAS flight days (Fig. 5a, b, c). We difference, both absolutely (Fig. 5d, e, f) and relatively (Fig. 5g, h, i), the three pairs of datasets and present them in scatter density plots (Fig. 5j, k, l). We focus on three regions: downtown Houston (DT) (red), the ru-

ral East Galveston Bay (RB) (blue), and all other areas (OA) (green), and we calculate the mean values and differences for these areas in the top left of each of the plots. The results presented in Fig. 5 are the temporal average across all flight days; however, similar figures for individual flight days are presented in Figs. S19–S28.

First, we consider the spatial distribution of NO₂ columns from GCAS (Fig. 5a), TROPOMI (Fig. 5b), and CAMx (Fig. 5c) independently. For all three datasets, NO₂ columns are higher in downtown Houston than in the rural East Galveston Bay; generally, they are between 3 and 5 times as large. The two finer-resolution datasets, GCAS and CAMx, also capture NO₂ peaks associated with point sources like those from W. A. Parish, Texas City, as well as Baytown and the ship channel. A map of the major point sources discussed in this work is included in Fig. S29. The coarser resolution of TROPOMI leads to fewer identifiable peaks associated with point sources; however, there are slightly elevated observed values near the W. A. Parish and Texas City power plants as well as the ship channel. Observations from GCAS and TROPOMI reveal a more diffuse peak in NO₂ columns in and around downtown Houston that includes elevated levels of NO₂ in the western part of the city. Simulated columns from CAMx, on the other hand, primarily estimate higher NO₂ values in the eastern area of downtown Houston and have lower NO₂ values in the western area of the city.

We next consider the three products compared with one another through three methods: absolute difference (Fig. 5d, e, f), relative difference (Fig. 5g, h, i), and scatter density plots (Fig. 5j, k, l). We intercompare these three products by isolating three sets of pairs: GCAS and TROPOMI, GCAS and CAMx, and TROPOMI and CAMx.

First, considering GCAS and TROPOMI, there appears to be a systematic low bias in TROPOMI observations throughout nearly the entire domain. Regardless of the spatial subset, the low bias in TROPOMI was consistent and ranged from -27% in downtown to -32% in the rural bay (Fig. 5g). In an absolute sense, on average TROPOMI was between 2.1 and $0.7 \times 10^{15}\,\text{molec.}\,\text{cm}^{-2}$ lower than GCAS (Fig. 5d) across the three locations. Throughout the entire domain, observations from GCAS and TROPOMI were well correlated ($r^2 = 0.85$), but TROPOMI had an overall negative normalized mean bias of -31.6% (Fig. 5j). We note that this low bias is slightly greater than what we would expect from considering the biases of these products relative to Pandora measurements as we do in Sect. 3.1; doing this we would expect TROPOMI to be low biased relative to GCAS by around 23%. This slight additional negative bias indicates that either the three Pandora sites are unable to capture the full extent of the negative TROPOMI bias, and that TROPOMI may be lower biased outside of these sites (e.g., areas outside of downtown Houston), or that GCAS observations may be biased additionally high outside of these sites. Notably, there are a few areas surrounding point sources in the eastern area of downtown and around the W. A. Parish plant in which
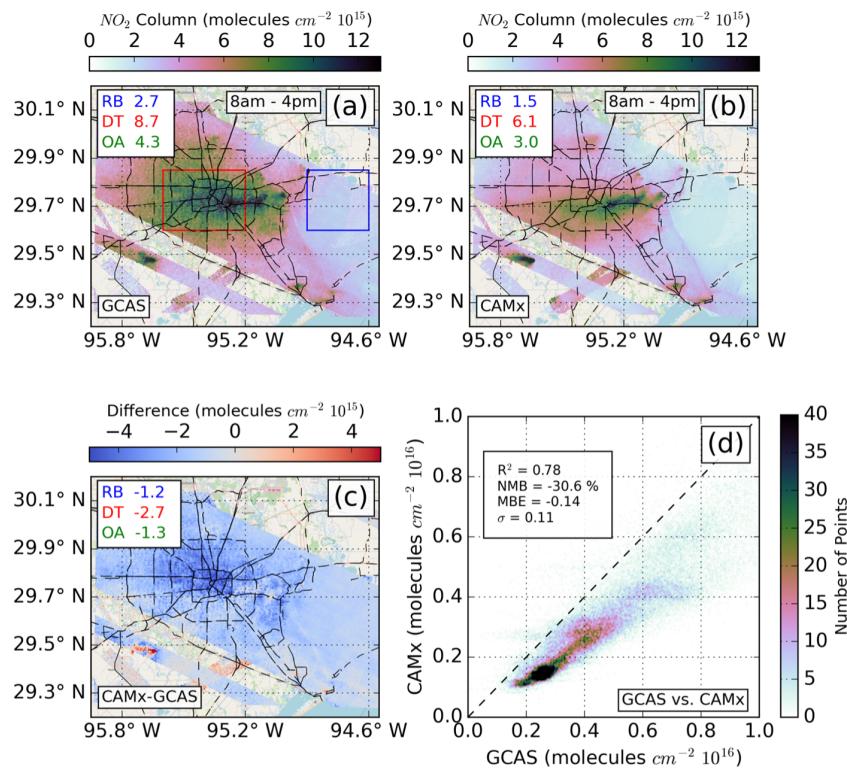
Figure 3. Comparison between Pandora measurements and TROPOMI observations using the CAMx AMF for version 2.3.1 **(a)**, 2.4.0 **(b)**, and NASA MINDS **(c)** and the same respective versions using the operational AMF **(d–f)**. Data from all possible overpasses coincident within 15 min of a Pandora observation are considered with one exception: data from 11 September 2021 were missing from the NASA MINDS product, so values in **(c)** and **(f)** exclude this day.

TROPOMI observes higher NO_2 columns than GCAS. This is likely attributable to the coarser resolution of TROPOMI that results in peaks of NO_2 spreading into surrounding areas that are in the same TROPOMI grid cell.

Second, comparing GCAS with CAMx we again find a low bias relative to GCAS, albeit one with a higher degree of spatial variability. In the remote bay, CAMx-simulated columns are lower than GCAS compared with elsewhere in the domain (-50%) (Fig. 5h), while downtown and background levels are similarly biased at 32 % and 39 %, respectively. This lower bias in the low emissions rural East Galveston Bay is indicative of an underestimation of background NO_2 columns in the CAMx simulation. Across these three regions the mean absolute differences range from -2.4 to -1.2×10^{15} molec. cm^{-2} (Fig. 5e). Visually, the negative bias in CAMx appears to be stronger in downtown and to the west, east, and northwest of downtown, and less to the south and southwest of downtown. Overall, GCAS and CAMx are well correlated ($r^2 = 0.74$) (Fig. 5k); however, simulated columns from CAMx have a worse negative bias (NMB $= -38.1\%$) against GCAS than what is captured at the Pandora sites of approximately -21%. Around some point sources CAMx columns are positively biased against

GCAS observations. This high bias in CAMx is likely attributable to differences in wind speed and direction in the WRF simulation than in reality. These differences could contribute to NO_2 plumes being advected in incorrect directions.

Lastly, when comparing observed columns from TROPOMI to simulated columns from CAMx, biases have a great degree of spatial variability; however, in general, CAMx is negatively biased. In a relative sense (Fig. 5i), the CAMx-simulated columns are lowest compared with TROPOMI in the rural bay (-26%) and similar in downtown (-7%) and in other areas (-11%). There are a few areas where this pattern does not hold: both in the area southwest of downtown Houston and near point sources, CAMx is biased high compared with TROPOMI. These results indicate that simulated columns from CAMx are underestimated in downtown Houston and that this underestimation could potentially be attributable to an incorrect advection of NO_2 from some downtown source to the southwest perhaps in conjunction with an underestimate of emissions in this downtown area. Overall, TROPOMI and CAMx are well correlated ($r^2 = 0.73$), and there is a spatially heterogeneous low bias when considering the two products throughout the domain (NMB $= -9.7\%$) (Fig. 5l).

Figure 4. Comparison of GCAS observations with CAMx-simulated NO$_2$ columns across all data during GCAS overpasses (generally 08:00–16:00 LT). Temporally averaged GCAS NO$_2$ columns **(a)**, temporally averaged simulated CAMx NO$_2$ columns **(b)**, the absolute difference between GCAS and CAMx **(c)**, and a scatter density plot comparing all observations between GCAS and CAMx **(d)**. We identify three distinct areas: downtown (DT; red), the low emissions rural East Galveston Bay (RB; blue), and all other areas (OA; green), and we calculate the averages in the top left corner of each chart. © OpenStreetMap contributors 2023. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.

3.4 Comparisons of GCAS, TROPOMI, and CAMx data at a coarser resolution
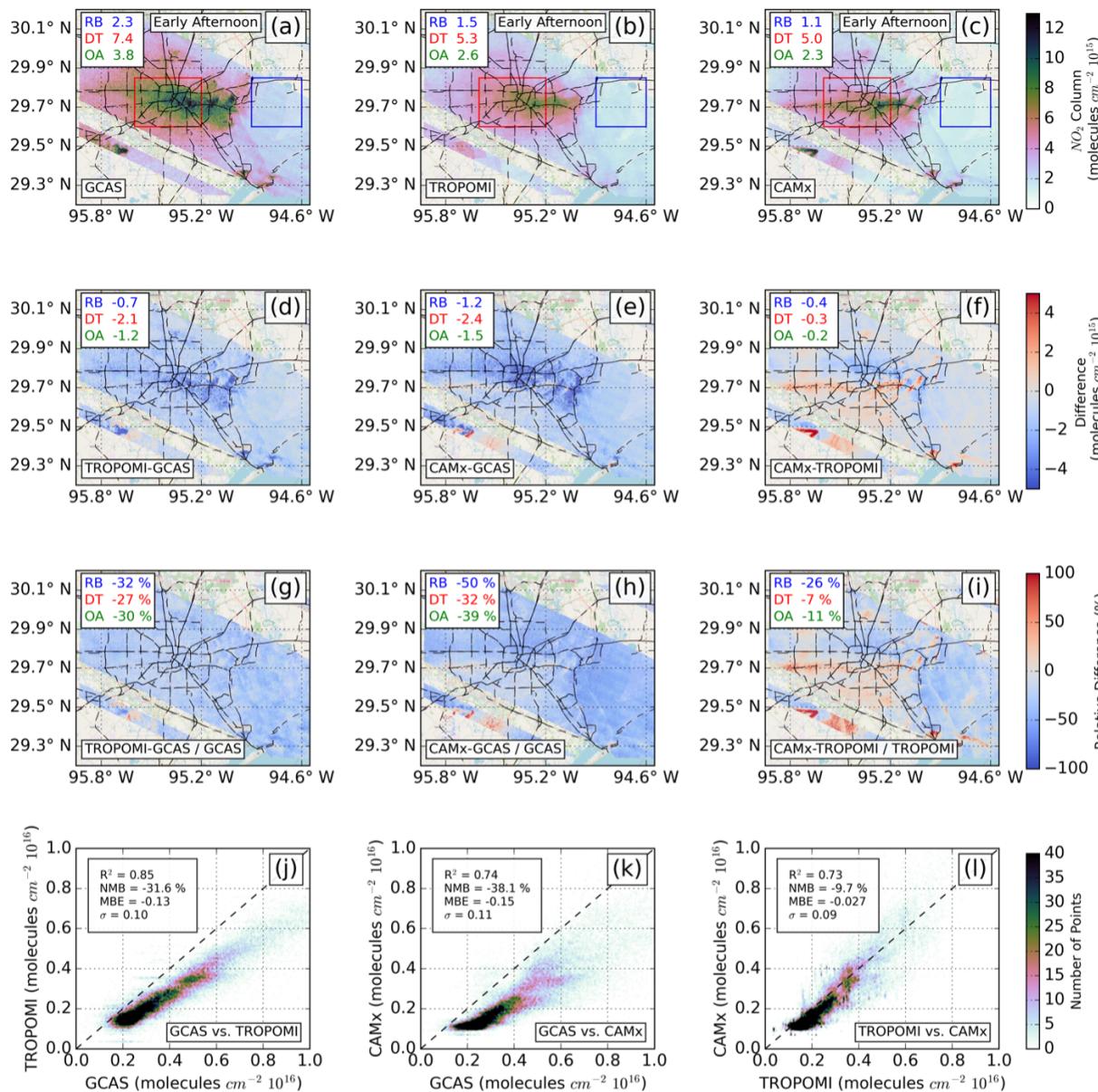
The comparisons presented in the prior section are done at the high resolution of the CAMx grid (444×444 m^2). Here, we characterize the effect of the coarser resolution of TROPOMI by performing an additional comparison of the three datasets at the $0.05° \times 0.05°$ resolution (approximately 5.5×5.5 km^2) (Fig. 6). We average all of the NO$_2$ columns from this finer resolution to the coarser resolution based on the centroid of the fine resolution grid cells. This new coarser resolution is comparable to that of the TROPOMI observations at nadir (on average 3.5×5.5 km^2). We additionally present comparisons at two further coarser resolutions in the Supplement: $0.25° \times 0.25°$ (Fig. S30) and $0.1° \times 0.1°$ (Fig. S31).

Generally, this change in resolution has only a minor effect on the trends discussed in the prior section. Observed NO$_2$ columns from GCAS and TROPOMI have a collocated peak in downtown Houston and NO$_2$ columns from TROPOMI are still systematically biased lower compared with GCAS. Simulated NO$_2$ columns from CAMx are clearly lower than GCAS in the area directly west of downtown and slightly

higher southwest of downtown compared with TROPOMI (Fig. 6a, b, c). Considering the spatial distribution of absolute (Fig. 6d, e, f) and relative (Fig. 6g, h, i) differences between the three products, the low bias in TROPOMI compared with GCAS is generally homogenous throughout the domain. On the other hand, there are clear peaks in negative biases in downtown and western Houston when comparing CAMx to GCAS, and in some areas southwest of downtown biases are small and positive. Averaging observations to this coarser resolution improved the correlation for all three pairs ($r^2 = 0.93$, 0.82, and 0.83 for GCAS and TROPOMI, GCAS and CAMx, and TROPOMI and CAMx, respectively), while the biases remained comparable to what was found in the comparison at a finer resolution (Fig. 5j, k, l).

3.5 Weekend vs. weekday patterns across the datasets

Three of the 10 GCAS flight days occurred on weekends (11, 25, and 26 September), and observations from GCAS and TROPOMI – along with simulated NO$_2$ columns from CAMx – exhibited different patterns on weekends vs. weekdays (1, 3, 8–10, 23, and 24 September). This difference in observed and simulated patterns is explored in greater detail
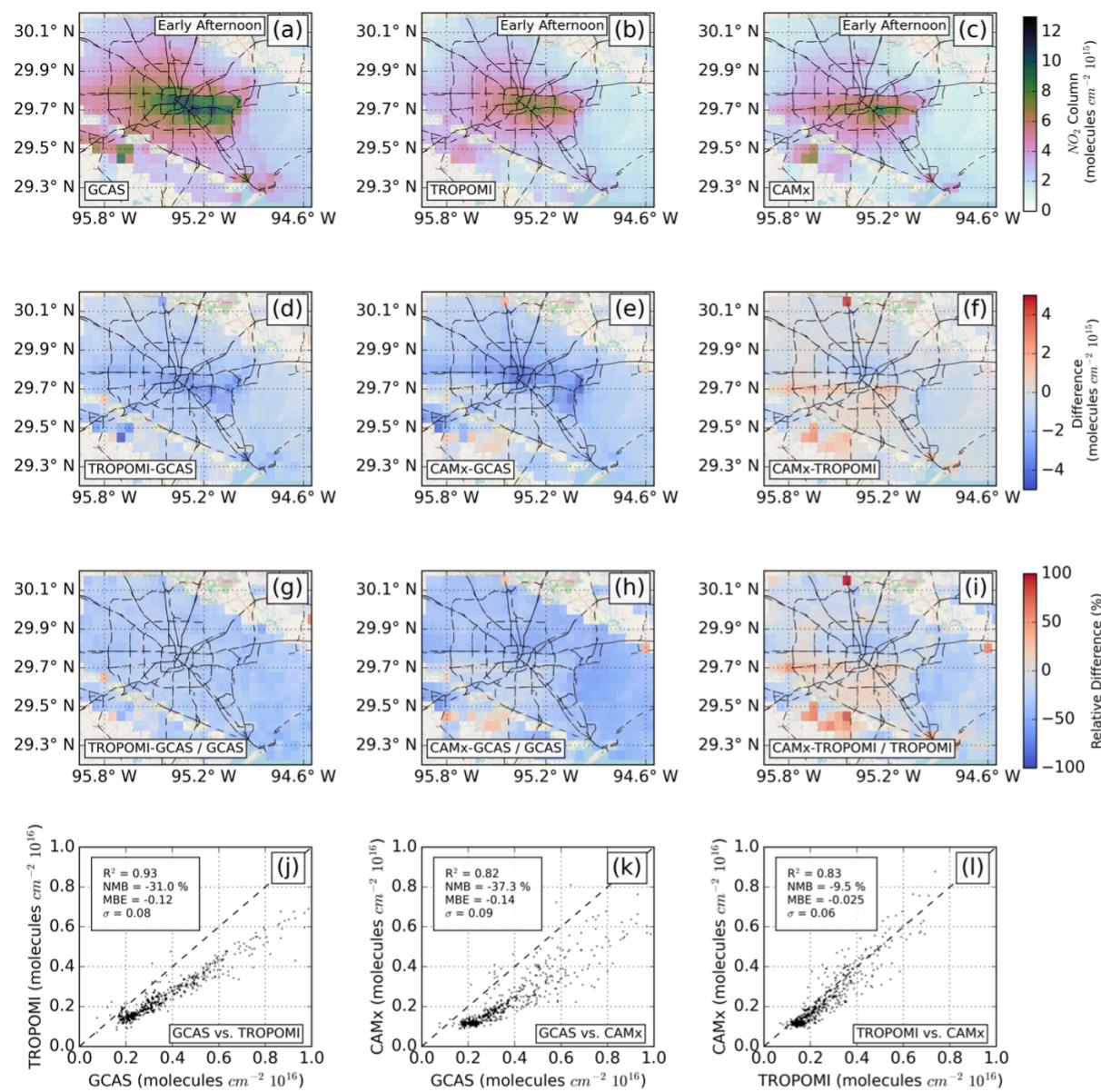
Figure 5. Spatial distribution of GCAS (a), TROPOMI (b), and CAMx (c) NO$_2$ columns averaged across the 10 GCAS flight days when within 90 min of each TROPOMI overpass representing early afternoon NO$_2$ columns. We identify three distinct areas: downtown (DT; red), the low emissions rural East Galveston Bay (RB; blue), and all other areas (OA; green), and we calculate the averages in the top left corner of each chart. Absolute differences between GCAS and TROPOMI (d), GCAS and CAMx (e), and TROPOMI and CAMx (f). Relative differences between GCAS and TROPOMI (g), GCAS and CAMx (h), and TROPOMI and CAMx (i). Scatter density plots of GCAS vs. TROPOMI (j), GCAS vs. CAMx (k), and TROPOMI vs. CAMx (l). © OpenStreetMap contributors 2023. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.

in this section, first through comparisons with Pandora measurements (Fig. 7) and then through spatial comparisons of the products on weekdays vs. weekends (Fig. 8). When interpreting these results, it should be considered that weekend data are limited to only 3 d. This data sparsity introduces a high degree of uncertainty in conclusions derived from this analysis. Day-to-day changes in meteorological conditions are likely responsible for some of the exhibited differences,

so they cannot solely be attributed to differences in emission patterns.

First, we consider how comparisons of the observational datasets – GCAS and TROPOMI – with Pandora change on weekends compared with weekdays. Biases for both GCAS and TROPOMI become more positive on weekends, NMB $= 10.2\%$ and NMB $= -15.7\%$, respectively, than on weekdays, NMB $= 1.5\%$ and NMB $= -25.2\%$. GCAS ob-
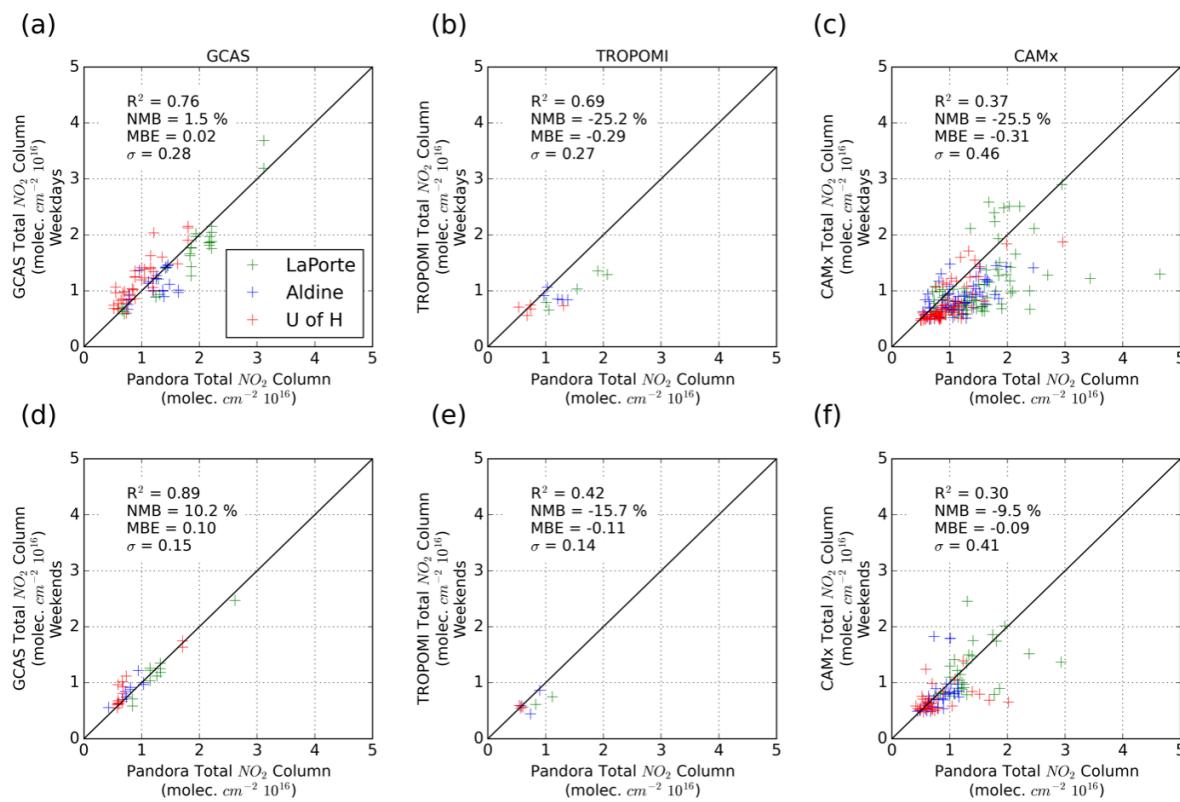
Figure 6. Spatial distribution of GCAS **(a)**, TROPOMI **(b)**, and CAMx **(c)** at the 0.05° × 0.05° resolution averaged across the 10 GCAS flight days when within 1.5 h of each TROPOMI overpass representing early afternoon NO$_2$ columns. Absolute differences between GCAS and TROPOMI **(d)**, GCAS and CAMx **(e)**, and TROPOMI and CAMx **(f)**. Relative differences between GCAS and TROPOMI **(g)**, GCAS and CAMx **(e)**, and TROPOMI and CAMx **(f)**. Scatter density plots of GCAS vs. TROPOMI **(g)**, GCAS vs. CAMx **(h)**, and TROPOMI vs. CAMx **(i)**. © OpenStreetMap contributors 2023. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.

servations are slightly better correlated to Pandora measurements on weekends ($r^2 = 0.89$ vs. $r^2 = 0.76$); however, TROPOMI observations are worse correlated ($r^2 = 0.42$ vs. $r^2 = 0.69$), which is likely attributable to a limited number of observations that are at lower NO$_2$ column levels with limited dynamic range. Overall, biases are slightly worse for GCAS and better for TROPOMI on weekends; however, given the small number of measurements it is unclear whether this pattern is attributable to meteorological condi-

tions or if it is attributable to some systematic bias in the instruments.

Simulated NO$_2$ columns from CAMx exhibit clearer weekday vs. weekend patterns, and since these simulated columns are available for every hour of the day there is a greater number of measurements to support these findings than for the two observational datasets. While the correlation is slightly degraded on weekends ($r^2 = 0.30$ vs. $r^2 = 0.37$), the negative bias in simulated columns com-

Figure 7. Comparison of GCAS (**a**), TROPOMI (**b**), and CAMx (**c**) to Pandora on weekdays and of GCAS (**d**), TROPOMI (**e**), and CAMx (**f**) to Pandora on weekends. Data from all possible overpasses coincident within 15 min of a Pandora observation are considered. GCAS flight times generally ranged from 08:00 to 16:00 LT. TROPOMI overpasses occurred around 13:30 LT.

pared with Pandora measurements is reduced on weekends (NMB = −25.5 % vs. NMB = −9.5 %).
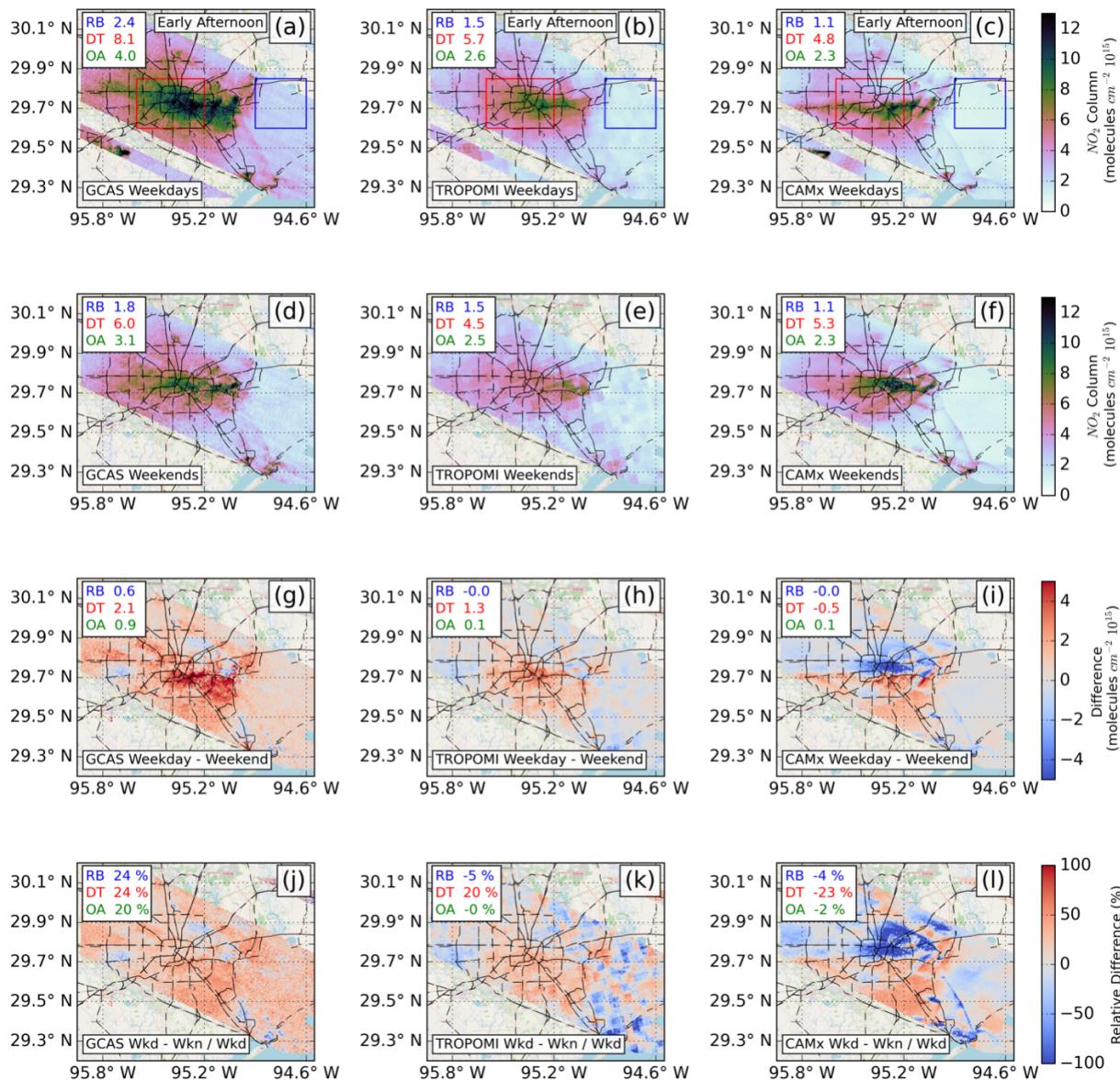
GCAS and TROPOMI observations of NO₂ column concentrations are higher on weekdays (Fig. 8a, b) than on weekends (Fig. 8d, e). This is true in downtown Houston and the rural bay where weekday GCAS observations are 2.1×10^{15} molec. cm^{-2} (24 %) and 0.6×10^{15} molec. cm^{-2} (24 %) higher, respectively, on weekdays than on weekends. In other areas of Houston, GCAS observations on weekdays are higher than on weekends but not to the same degree (20 %). A similar pattern occurs for TROPOMI: in downtown Houston TROPOMI columns are 20 % higher on weekdays than on weekends but comparable in other areas and −5 % lower in the rural bay. This comparison again implicates some underestimated weekday source of NO₂ in CAMx that is of great importance in the western area of Houston; however, due to the lack of data on weekends – that is apparent in the discontinuities in the weekend NO₂ column concentrations of TROPOMI – it is difficult to examine this quantitatively.

Comparing weekday columns simulated from CAMx with weekend columns, we find that the mean concentrations for the three defined areas are nearly identical (Fig. 8c and f), although columns on weekdays are higher south and south-west of downtown, while columns on weekends are higher within downtown. These spatial patterns are further revealed in the difference plots (Fig. 8i and l) where the difference in weekday vs. weekend values appears to be split right along interstate highway I-10: north of I-10 weekday values are much lower than weekend values, while south of I-10 the opposite is true. This difference is likely attributable to different meteorological conditions on these days. Overall, simulated CAMx columns are substantially lower than GCAS and TROPOMI on weekdays but more similar on weekends, implying that weekday emissions may be underestimated in the TCEQ inventory.

3.6   Relevance to TEMPO: diurnal patterns in column and surface NO₂
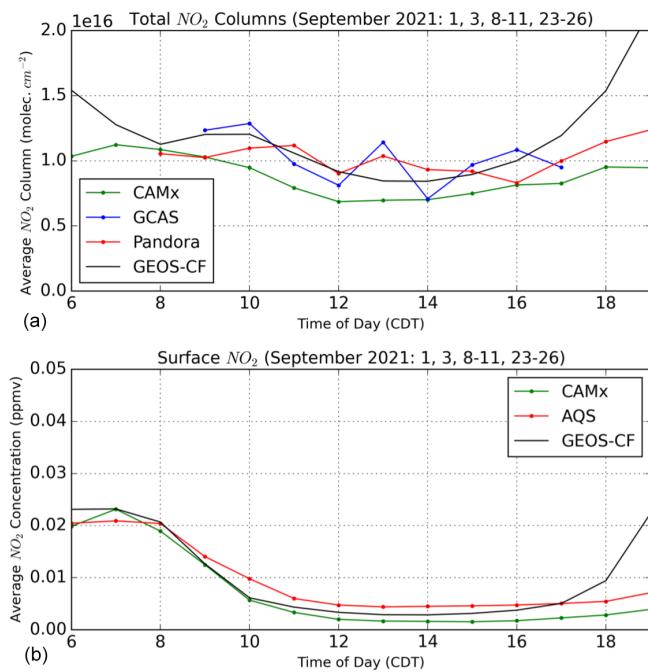
Lastly, we characterize the diurnal profiles of simulated and observed NO₂ columns during the TRACER-AQ campaign in downtown Houston (Fig. 9). First, considering column concentrations, we find generally good agreement during the early morning (08:00–10:00 LT) across the two simulated datasets (CAMx and GEOS-CF) and two observational datasets (GCAS and Pandora). Interestingly, between 09:00 and 11:00 LT, Pandora column measurements show a slight increase, while model simulations show a slight decrease

Figure 8. Spatial distribution of GCAS, TROPOMI, and CAMx NO$_2$ columns on weekdays **(a–c)** and weekends **(d–f)**, as well as the absolute difference **(g–i)** and relative difference **(j–l)** between weekdays and weekends. Data are averaged across the GCAS flight days corresponding to weekdays or weekends when within 1.5 h of each TROPOMI overpass representing early afternoon NO$_2$ columns. We identify three distinct areas: downtown (DT; red), the low emissions rural East Galveston Bay (RB; blue), and all other areas (OA; green), and we calculate the averages in the top left corner of each chart. © OpenStreetMap contributors 2023. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.

during the same time interval. During midday and the afternoon (11:00–16:00 LT) – that corresponds to the period with the most GCAS observations – GEOS-CF columns generally agree well with Pandora observations. In the evening (17:00–19:00 LT), GEOS-CF columns have a substantial high bias across these flight days. The GEOS-CF mismatch in the evening has implications for TEMPO NO$_2$ evening retrievals if this is a persistent bias in other urban areas since satellite instruments are especially sensitive to a priori assumptions at low sun angles.

Second, considering surface concentrations, we see a similar trend. Generally, there is great agreement across the three datasets (CAMx, AQS observations, and GEOS-CF) in the early morning (06:00–09:00 LT) before they begin to diverge with the two simulated produces maintaining comparable magnitudes with low biases compared with surface monitors. At around midday to the afternoon (12:00–17:00 LT) both simulated products have a low bias compared with observed surface-level NO$_2$; however, the bias in CAMx concentrations is worse. Some of the apparent low bias may be related

Figure 9. Diurnal patterns in total NO$_2$ columns **(a)** averaged across the three Pandora sites and 10 flight days from CAMx (green), GCAS (blue), Pandora (red), and GEOS-CF (black). Diurnal patterns in surface-level NO$_2$ concentrations **(b)** in downtown Houston for CAMx and GEOS-CF averaged across the 10 flight days and across all monitors in Harris County for AQS surface-level monitors (red).

to an artificial high bias in NO$_2$ chemiluminescence surface monitors (Dunlea et al., 2007; Lamsal et al., 2008). In the evening (18:00–19:00 LT), surface-level NO$_2$ from GEOS-CF climbs rapidly; however, observed NO$_2$ from the AQS and simulated NO$_2$ from CAMx increase only slightly. The large increase in NO$_2$ in GEOS-CF in the evening appears at both the surface and in the column, potentially indicating issues capturing boundary layer dynamics.

4 Conclusions

This study leveraged observational datasets of NO$_2$ column densities from three instruments: Pandora ground-based spectrometers, the airborne GCAS instrument, and the satellite TROPOMI instrument. These instruments were used to investigate NO$_2$ column densities in Houston, Texas, during the September 2021 TRACER-AQ campaign and to characterize strengths/weaknesses and uncertainties in the respective datasets. These observational datasets were then compared with simulated NO$_2$ columns from CAMx to characterize the performance of the simulation and to identify potential under- or overestimates of emissions in the simulation. We found that GCAS has strong agreement with Pandora instruments ($r^2 = 0.79$ and NMB = 3.4 %) during

its overpasses and that TROPOMI also has strong performance but an important low bias – consistent with validation by the European Space Agency (Verhoelst et al., 2021) – across the urban Houston locations ($r^2 = 0.73$ and NMB = −22.8 %). This low bias in TROPOMI observations persists despite the inclusion of an air mass factor derived from the CAMx simulation. When comparing different versions of TROPOMI we found differences between v2.3.1, v2.4.0, and the NASA MINDS product and found that the MINDS ($r^2 = 0.69$ and NMB = −18.2 %) and version 2.3.1 ($r^2 = 0.72$ and NMB = −18.3 %) products – with the CAMx AMF – perform comparably and that both outperform version 2.4.0 considering bias, albeit with slightly worse correlation. The performance of the CAMx simulation varied depending on the day, but overall, simulated NO$_2$ columns were more poorly correlated and more negatively biased, compared with Pandora measurements, than the observational datasets ($r^2 = 0.34$ and NMB = −21.2 %). Notably, this low bias in CAMx-simulated NO$_2$ columns improved on weekends (NMB = −9.5 %), albeit over a limited number of days. This improvement on weekends implies that a source that emits in greater amounts on weekdays (e.g., heavy-duty vehicles) could be underestimated in the TCEQ inventory; however, we cannot say this conclusively given the limited number of observations on weekends. The poor correlation in the simulated NO$_2$ columns is likely attributable to minor wind directional errors – simulated wind direction had an MAE that ranged from 14 to 32° when compared with observations – and spatial correlations over larger extents match well.

When we compare the spatial distribution of TROPOMI observations to GCAS (Figs. 3 and 4), we find that the low bias in TROPOMI NO$_2$ columns is perhaps stronger than the low bias implied at the three Pandora sites. This could be a resolution constraint of the coarser TROPOMI product which is unable to capture the fine-scale features in NO$_2$ column concentrations that GCAS is able to capture. If coarse resolution is responsible for this low bias, new instruments on geostationary satellites from missions like the NASA TEMPO mission could be leveraged to further improve satellite-derived estimates of urban NO$_2$ in cities like Houston. CAMx comparisons with GCAS, when extended beyond the limited number of Pandora sites, indicate that the CAMx-simulated low bias could be substantially worse than at the Pandora sites (−32 %) in downtown and west of downtown Houston. This overall underestimate in the CAMx simulations is potentially attributable to a number of confounding factors including an inability of the WRF simulation to capture local meteorology – WRF-simulated wind speeds had only modest correlation with observations ($R^2 = 0.26$), although there was little systematic bias (MBE = −0.02) – and an underestimate of emissions in sectors that are more spatially located in downtown and west Houston like on-road mobile emissions. We also consider differences in the diurnal profiles of surface and column NO$_2$ across multiple datasets and find that the performance of CAMx is at its worst in the

late morning and early afternoon (i.e., performance is better during other times of the day).

There is a clear negative bias in the CAMx-simulated NO_2 columns compared with GCAS observations. Although we primarily evaluate the performance of WRF meteorology at the surface, we also briefly investigate model vertical structure for five ozonesondes from different locations and days (Figs. S14–S18) and find great agreement in temperature and pressure; however, there is more mixed agreement in the ozone mixing ratio, wind speed, and wind direction. Future evaluation of 3D model simulated vertical structure for NO_2 using observations from NASA, such as measurements from the High Spectral Resolution Lidar 2 (HSRL-2) instrument, the Tropospheric Ozone Lidar Network (TolNet), or TRACER-AQ, may be helpful for diagnosing the distinct influences of emissions, meteorology, and chemistry on column NO_2. A previous study by Liu et al. (2023) has investigated this for the TRACER-AQ campaign in Houston, albeit with a different chemical transport model, and found generally good agreement in potential temperature but an underestimate of ozone in the free troposphere. We note that the YSU scheme used in the WRF–CAMx simulation (Table S1) has been shown to underestimate planetary boundary layer (PBL) height in the Houston area during the TRACER-AQ campaign (Liu et al., 2023) which would likely impact the vertical distribution of NO_2. Given the worse performance of WRF–CAMx at the surface (NMB = −59 %) than for the columns (−22 %), if the vertical mixing scheme has poor performance, we suspect it could be due to overmixing leading to the rapid removal of surface-level NO_2. Additionally, the low bias in the TROPOMI observations compared with Pandora and GCAS merits further investigation. The role of algorithm and resolution could be considered by comparing different versions and finer resolution geostationary observations in the future, beyond what is considered in this study. The reference background NO_2 from TROPOMI used in GCAS could also introduce error into these results that should be considered. Given the fine resolution of GCAS observations and CAMx-simulated column concentrations, there is potential for investigations into how air pollution is inequitably distributed across different populations in Houston and how specific sources contribute to these inequities. The findings presented here imply that TROPOMI-derived NO_2 column concentrations may be underestimated in Houston if not corrected for in applications such as exposure assessments and NO_x emissions derivations.

This analysis benefited from three independent measurement datasets (i.e., Pandora, TROPOMI, and GCAS) that were critical to isolate the negative biases in TROPOMI and CAMx, although we note that negative biases in TROPOMI have been mentioned in earlier literature (e.g., Verhoelst et al., 2021) and in the quarterly issued operational validation reports (available at https://mpc-vdaf.tropomi.eu/, last access: February 2024). It is common to consider TROPOMI measurements an accurate representation of NO_2 column concentrations; however, if we had done so in this study, we would have failed to identify the substantial negative bias in the CAMx simulation of column concentrations. Observations from multiple Pandora instruments and GCAS overpasses made it possible to isolate negative biases in TROPOMI and CAMx. While there are some errors in the meteorology – notably only a modest correlation between simulated and observed wind speed, albeit with little systematic bias, and mixed capturing of vertical structure compared with ozonesonde observations – these errors are unlikely to fully explain the low bias in simulated NO_2. Given the relatively minimal biases in WRF-simulated wind speed and direction at the surface compared with observations, low NO_2 biases in the simulated CAMx column concentrations imply that current TCEQ NO_x emission inventories in the Houston area used to drive the CAMx simulation may be underestimated, and that this underestimation is likely attributable to a source with weekday–weekend differences and correlated with roadways and/or population density.

References

Achakulwisut, P., Brauer, M., Hystad, P., and Anenberg, S. C.: Global, national, and urban burdens of paediatric asthma incidence attributable to ambient NO$_2$ pollution: estimates from global datasets, Lancet Planet. Health, 3, e166–e178, https://doi.org/10.1016/S2542-5196(19)30046-4, 2019.

Anenberg, S. C., Mohegh, A., Goldberg, D. L., Kerr, G. H., Brauer, M., Burkart, K., Hystad, P., Larkin, A., Wozniak, S., and Lamsal, L.: Long-term trends in urban NO$_2$ concentrations and associated paediatric asthma incidence: estimates from global datasets, Lancet Planet. Health, 6, e49–e58, https://doi.org/10.1016/S2542-5196(21)00255-2, 2022.

Boersma, K. F., Eskes, H. J., Richter, A., De Smedt, I., Lorente, A., Beirle, S., van Geffen, J. H. G. M., Zara, M., Peters, E., Van Roozendael, M., Wagner, T., Maasakkers, J. D., van der A, R. J., Nightingale, J., De Rudder, A., Irie, H., Pinardi, G., Lambert, J.-C., and Compernolle, S. C.: Improving algorithms and uncertainty estimates for satellite NO2 retrievals: results from the quality assurance for the essential climate variables (QA4ECV) project, Atmos. Meas. Tech., 11, 6651–6678, https://doi.org/10.5194/amt-11-6651-2018, 2018.

Bucsela, E. J., Krotkov, N. A., Celarier, E. A., Lamsal, L. N., Swartz, W. H., Bhartia, P. K., Boersma, K. F., Veefkind, J. P., Gleason, J. F., and Pickering, K. E.: A new stratospheric and tropospheric NO$_2$ retrieval algorithm for nadir-viewing satellite instruments: applications to OMI, Atmos. Meas. Tech., 6, 2607–2626, https://doi.org/10.5194/amt-6-2607-2013, 2013.

Cede, A.: Manual for Blick Software Suite 1.8, LUFT-BLICK, https://www.pandonia-global-network.org/wp-content/uploads/2020/12/BlickSoftwareSuite_Manual_v1-8-1.pdf (last access: March 2024), 2021.

Copernicus Sentinel-5P (processed by ESA): TROPOMI Level 2 Nitrogen Dioxide total column products, Version 02, European Space Agency [data set], https://doi.org/10.5270/S5P-9bnp8q8, 2021.

Danckaert, T., Fayt, C., Roozendael, M. V., Smedt, I. D., Letocart, V., Merlaud, A., and Pinardi, G.: QDOAS Software user manual, DOAS UV-VIS team at BIRA-IASB, https://uv-vis.aeronomie.be/software/QDOAS/index.php (last access: March 2024), 2017.

de Foy, B., Lei, W., Zavala, M., Volkamer, R., Samuelsson, J., Mellqvist, J., Galle, B., Martínez, A.-P., Grutter, M., Retama, A., and Molina, L. T.: Modelling constraints on the emission inventory and on vertical dispersion for CO and SO2 in the Mexico City Metropolitan Area using Solar FTIR and zenith sky UV spectroscopy, Atmos. Chem. Phys., 7, 781–801, https://doi.org/10.5194/acp-7-781-2007, 2007.

de Foy, B., Wilkins, J. L., Lu, Z., Streets, D. G., and Duncan, B. N.: Model evaluation of methods for estimating surface emissions and chemical lifetimes from satellite data, Atmos. Environ., 98, 66–77, https://doi.org/10.1016/j.atmosenv.2014.08.051, 2014.

Demetillo, M. A. G., Navarro, A., Knowles, K. K., Fields, K. P., Geddes, J. A., Nowlan, C. R., Janz, S. J., Judd, L. M., Al-Saadi, J., Sun, K., McDonald, B. C., Diskin, G. S., and Pusede, S. E.: Observing Nitrogen Dioxide Air Pollution Inequality Using High-Spatial-Resolution Remote Sensing Measurements in Houston, Texas, Environ. Sci. Technol., 54, 9882–9895, https://doi.org/10.1021/acs.est.0c01864, 2020.

DOC/NOAA/NWS/NCEP/EMC: Global Data Assimilation System (GDAS), National Centers for Environmental Information, https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ncdc:C00379 (last access: March 2023), 2023.

Dunlea, E. J., Herndon, S. C., Nelson, D. D., Volkamer, R. M., San Martini, F., Sheehy, P. M., Zahniser, M. S., Shorter, J. H., Wormhoudt, J. C., Lamb, B. K., Allwine, E. J., Gaffney, J. S.,

Marley, N. A., Grutter, M., Marquez, C., Blanco, S., Cardenas, B., Retama, A., Ramos Villegas, C. R., Kolb, C. E., Molina, L. T., and Molina, M. J.: Evaluation of nitrogen dioxide chemiluminescence monitors in a polluted urban environment, Atmos. Chem. Phys., 7, 2691–2704, https://doi.org/10.5194/acp-7-2691-2007, 2007.

EOSDIS Worldview: https://worldview.earthdata.nasa.gov/, last access: 29 October 2023.

EPA: Clean Air Markets Program Data, United States Environmental Protection Agency (EPA), Office of Atmospheric Protection, Clean Air Markets Division, Washington, DC, https://campd.epa.gov/ (last access: April 2023), 2023.

Eskes, H. J., Eichmann, K.-U., Lambert, J.-C., Loyola, D., Stein-Zweers, D., Dehn, A., and Zehner, C.: S5P MPC Product Readme Nitrogen Dioxide, ESA and Copernicus, https://sentinels.copernicus.eu/documents/247904/3541451/Sentinel-5P-Nitrogen-Dioxide-Level-2-Product-Readme-File (last access: October 2023), 2023.

Ge, S., Wang, S., Xu, Q., and Ho, T.: Source apportionment simulations of ground-level ozone in Southeast Texas employing OSAT/APCA in CAMx, Atmos. Environ., 253, 118370, https://doi.org/10.1016/j.atmosenv.2021.118370, 2021.

Geddes, J. A., Martin, R. V., Bucsela, E. J., McLinden, C. A., and Cunningham, D. J. M.: Stratosphere–troposphere separation of nitrogen dioxide columns from the TEMPO geostationary satellite instrument, Atmos. Meas. Tech., 11, 6271–6287, https://doi.org/10.5194/amt-11-6271-2018, 2018.

Goldberg, D. L., Harkey, M., de Foy, B., Judd, L., Johnson, J., Yarwood, G., and Holloway, T.: Evaluating NO$_x$ emissions and their effect on O$_3$ production in Texas using TROPOMI NO$_2$ and HCHO, Atmos. Chem. Phys., 22, 10875–10900, https://doi.org/10.5194/acp-22-10875-2022, 2022.

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

Herman, J., Cede, A., Spinei, E., Mount, G., Tzortziou, M., and Abuhassan, N.: NO$_2$ column amounts from ground-based Pandora and MFDOAS spectrometers using the direct-sun DOAS technique: Intercomparisons and application to OMI validation, J. Geophys. Res.-Atmos., 114, D13307, https://doi.org/10.1029/2009JD011848, 2009.

Huang, S., Li, H., Wang, M., Qian, Y., Steenland, K., Caudle, W. M., Liu, Y., Sarnat, J., Papatheodorou, S., and Shi, L.: Long-term exposure to nitrogen dioxide and mortality: A systematic review and meta-analysis, Sci. Total Environ., 776, 145968, https://doi.org/10.1016/j.scitotenv.2021.145968, 2021.

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

Jia, J., Cheng, S., Liu, L., Lang, J., Wang, G., Chen, G., and Liu, X.: An Integrated WRF-CAMx Modeling Approach for Impact Analysis of Implementing the Emergency PM$_{2.5}$ Control Measures during Red Alerts in Beijing in December 2015, Aerosol Air Qual. Res., 17, 2491–2508, https://doi.org/10.4209/aaqr.2017.01.0009, 2017.

Jin, X., Zhu, Q., and Cohen, R. C.: Direct estimates of biomass burning NO$_x$ emissions and lifetimes using daily observations from TROPOMI, Atmos. Chem. Phys., 21, 15569–15587, https://doi.org/10.5194/acp-21-15569-2021, 2021.

Judd, L. M., Al-Saadi, J. A., Szykman, J. J., Valin, L. C., Janz, S. J., Kowalewski, M. G., Eskes, H. J., Veefkind, J. P., Cede, A., Mueller, M., Gebetsberger, M., Swap, R., Pierce, R. B., Nowlan, C. R., Abad, G. G., Nehrir, A., and Williams, D.: Evaluating Sentinel-5P TROPOMI tropospheric NO2 column densities with airborne and Pandora spectrometers near New York City and Long Island Sound, Atmos. Meas. Tech., 13, 6113–6140, https://doi.org/10.5194/amt-13-6113-2020, 2020.

Judd, L. M., Sullivan, J. T., Lefer, B., Haynes, J., Jensen, M. P., and Nadkarni, R.: TRACER-AQ Science Plan, Version 1, NASA, https://www-air.larc.nasa.gov/missions/tracer-aq/docs/TRACERAQ_SciencePlan_v1.pdf (last access: March 2024), 2021.

Keller, C. A., Knowland, K. E., Duncan, B. N., Liu, J., Anderson, D. C., Das, S., Lucchesi, R. A., Lundgren, E. W., Nicely, J. M., Nielsen, E., Ott, L. E., Saunders, E., Strode, S. A., Wales, P. A., Jacob, D. J., and Pawson, S.: Description of the NASA GEOS Composition Forecast Modeling System GEOS-CF v1.0, J. Adv. Model. Earth Syst., 13, e2020MS002413, https://doi.org/10.1029/2020MS002413, 2021 (data available at: https://opendap.nccs.nasa.gov/dods/, last access: October 2023).

Kim, S.-W., McKeen, S. A., Frost, G. J., Lee, S.-H., Trainer, M., Richter, A., Angevine, W. M., Atlas, E., Bianco, L., Boersma, K. F., Brioude, J., Burrows, J. P., de Gouw, J., Fried, A., Gleason, J., Hilboll, A., Mellqvist, J., Peischl, J., Richter, D., Rivera, C., Ryerson, T., te Lintel Hekkert, S., Walega, J., Warneke, C., Weibring, P., and Williams, E.: Evaluations of NO$_x$ and highly reactive VOC emission inventories in Texas and their implications for ozone plume simulations during the Texas Air Quality Study 2006, Atmos. Chem. Phys., 11, 11361–11386, https://doi.org/10.5194/acp-11-11361-2011, 2011.

Kowalewski, M. G. and Janz, S. J.: Remote sensing capabilities of the GeoCAPE Airborne Simulator, in: Earth Observing Systems XIX, Earth Observing Systems XIX, Proceedings of SPIE (Society of Photo-Optical Instrumentation Engineers), 496–507, https://doi.org/10.1117/12.2062058, 2014.

Lambert, J.-C., Claas, J., Stein-Zweers, D., Ludewig, A., Loyola, D., Sneep, M., and Dehn, A.: Quarterly Validation Report of the Copernicus Sentinel-5 Precursor Operational Data Products #19, European Space Agency, https://sentinels.copernicus.eu/documents/d/sentinel/s5p-mpc-sron-prf-ch4_v02-06-00_2-5_20231129_signed (last access: March 2024), 2023.

Lamsal, L. N., Martin, R. V., van Donkelaar, A., Steinbacher, M., Celarier, E. A., Bucsela, E., Dunlea, E. J., and Pinto, J. P.: Ground-level nitrogen dioxide concentrations inferred from the satellite-borne Ozone Monitoring Instrument, J. Geophys. Res.-Atmos., 113, D16308, https://doi.org/10.1029/2007JD009235, 2008.

Lamsal, L. N., Krotkov, N. A., Marchenko, S. V., Joiner, J., Oman, L., Vasilkov, A., Fisher, B., Qin, W., Yang, E.-S., Fasnacht, Z., Choi, S., Leonard, P., and Haffner, D.: TROPOMI/S5P NO2 Tropospheric, Stratospheric and Total Columns MINDS 1-Orbit L2

Swath 5.5 km × 3.5 km, NASA Goddard Space Flight Center, Goddard Earth Sciences Data and Information Services Center, https://doi.org/10.5067/MEASURES/MINDS/DATA203, 2022.

Leitch, J. W., Delker, T., Good, W., Ruppert, L., Murcray, F., Chance, K., Liu, X., Nowlan, C., Janz, S. J., Krotkov, N. A., Pickering, K. E., Kowalewski, M., and Wang, J.: The Geo-TASO airborne spectrometer project, in: Earth Observing Systems XIX, Earth Observing Systems XIX, Proceedings of SPIE (Society of Photo-Optical Instrumentation Engineers), 487–495, https://doi.org/10.1117/12.2063763, 2014.

Li, K.-F., Khoury, R., Pongetti, T. J., Sander, S. P., Mills, F. P., and Yung, Y. L.: Diurnal variability of stratospheric column NO_2 measured using direct solar and lunar spectra over Table Mountain, California (34.38° N), Atmos. Meas. Tech., 14, 7495–7510, https://doi.org/10.5194/amt-14-7495-2021, 2021.

Li, W., Wang, Y., Liu, X., Soleimanian, E., Griggs, T., Flynn, J., and Walter, P.: Understanding offshore high-ozone events during TRACER-AQ 2021 in Houston: insights from WRF–CAMx photochemical modeling, Atmos. Chem. Phys., 23, 13685–13699, https://doi.org/10.5194/acp-23-13685-2023, 2023.

Liu, X., Wang, Y., Wasti, S., Li, W., Soleimanian, E., Flynn, J., Griggs, T., Alvarez, S., Sullivan, J. T., Roots, M., Twigg, L., Gronoff, G., Berkoff, T., Walter, P., Estes, M., Hair, J. W., Shingler, T., Scarino, A. J., Fenn, M., and Judd, L.: Evaluating WRF-GC v2.0 predictions of boundary layer height and vertical ozone profile during the 2021 TRACER-AQ campaign in Houston, Texas, Geosci. Model Dev., 16, 5493–5514, https://doi.org/10.5194/gmd-16-5493-2023, 2023.

Lorente, A., Boersma, K. F., Eskes, H. J., Veefkind, J. P., van Geffen, J. H. G. M., de Zeeuw, M. B., Denier van der Gon, H. A. C., Beirle, S., and Krol, M. C.: Quantification of nitrogen oxides emissions from build-up of pollution over Paris with TROPOMI, Sci. Rep., 9, 20033, https://doi.org/10.1038/s41598-019-56428-5, 2019.

Luke, W. T., Kelley, P., Lefer, B. L., Flynn, J., Rappenglück, B., Leuchner, M., Dibb, J. E., Ziemba, L. D., Anderson, C. H., and Buhr, M.: Measurements of primary trace gases and NOY composition in Houston, Texas, Atmos. Environ., 44, 4068–4080, https://doi.org/10.1016/j.atmosenv.2009.08.014, 2010.

Mazzuca, G. M., Ren, X., Loughner, C. P., Estes, M., Crawford, J. H., Pickering, K. E., Weinheimer, A. J., and Dickerson, R. R.: Ozone production and its sensitivity to NO_x and VOCs: results from the DISCOVER-AQ field experiment, Houston 2013, Atmos. Chem. Phys., 16, 14463–14474, https://doi.org/10.5194/acp-16-14463-2016, 2016.

McDuffie, E. E., Smith, S. J., O'Rourke, P., Tibrewal, K., Venkataraman, C., Marais, E. A., Zheng, B., Crippa, M., Brauer, M., and Martin, R. V.: A global anthropogenic emission inventory of atmospheric pollutants from sector- and fuel-specific sources (1970–2017): an application of the Community Emissions Data System (CEDS), Earth Syst. Sci. Data, 12, 3413–3442, https://doi.org/10.5194/essd-12-3413-2020, 2020.

Miller, D. J., Actkinson, B., Padilla, L., Griffin, R. J., Moore, K., Lewis, P. G. T., Gardner-Frolick, R., Craft, E., Portier, C. J., Hamburg, S. P., and Alvarez, R. A.: Characterizing Elevated Urban Air Pollutant Spatial Patterns with Mobile Monitoring in Houston, Texas, Environ. Sci. Technol., 54, 2133–2142, https://doi.org/10.1021/acs.est.9b05523, 2020.

Murray, L. T.: Lightning NOxand Impacts on Air Quality, Curr. Pollut. Rep., 2, 115–133, https://doi.org/10.1007/s40726-016-0031-7, 2016.

NASA/LARC/SD/ASDC: TRACER-AQ JSC G-V Aircraft Remotely Sensed GEOstationary Coastal and Air Pollution Events (GEO-CAPE) Airborne Simulator (GCAS) Data, NASA Langley Atmospheric Science Data Center DAAC [data set], https://doi.org/10.5067/ASDC/SUBORBITAL/TRACERAQ/DATA001/GV/AircraftRemoteSensing/GCAS_1, 2022a.

NASA/LARC/SD/ASDC: TRACER-AQ Pandora Column Observations, NASA Langley Atmospheric Science Data Center DAAC [data set], https://doi.org/10.5067/ASDC/SUBORBITAL/TRACERAQDATA001/Ground/Pandora_1, 2022b.

Nowlan, C. R., Liu, X., Janz, S. J., Kowalewski, M. G., Chance, K., Follette-Cook, M. B., Fried, A., González Abad, G., Herman, J. R., Judd, L. M., Kwon, H.-A., Loughner, C. P., Pickering, K. E., Richter, D., Spinei, E., Walega, J., Weibring, P., and Weinheimer, A. J.: Nitrogen dioxide and formaldehyde measurements from the GEOstationary Coastal and Air Pollution Events (GEO-CAPE) Airborne Simulator over Houston, Texas, Atmos. Meas. Tech., 11, 5941–5964, https://doi.org/10.5194/amt-11-5941-2018, 2018.

Palmer, P. I., Jacob, D. J., Chance, K., Martin, R. V., Spurr, R. J. D., Kurosu, T. P., Bey, I., Yantosca, R., Fiore, A., and Li, Q.: Air mass factor formulation for spectroscopic measurements from satellites: Application to formaldehyde retrievals from the Global Ozone Monitoring Experiment, J. Geophys. Res.-Atmos., 106, 14539–14550, https://doi.org/10.1029/2000JD900772, 2001.

Ramboll: Enhance MARINER Tool for Commercial Marine Emission Inventories Final Report, Ramboll US Consulting, Inc., https://www.tceq.texas.gov/downloads/air-quality/research/reports/emissions-inventory/5822231220030-20220624-ramboll-mariner-enhancement.pdf (last access: March 2023), 2022a.

Ramboll: User's Guide COMPREHENSIVE AIR QUALITY MODEL WITH EXTENSIONS Version 7.20, Ramboll US Consulting, Inc., https://www.camx.com/Files/CAMxUsersGuide_v7.20.pdf (last access: March 2023), 2022b.

Riess, T. C. V. W., Boersma, K. F., Van Roy, W., de Laat, J., Dammers, E., and van Vliet, J.: To new heights by flying low: comparison of aircraft vertical NO_2 profiles to model simulations and implications for TROPOMI NO_2 retrievals, Atmos. Meas. Tech., 16, 5287–5304, https://doi.org/10.5194/amt-16-5287-2023, 2023.

Skamarock, C., Klemp, B., Dudhia, J., Gill, O., Barker, M., Wang, W., and Powers, G.: A Description of the Advanced Research WRF Version 2 (No. NCAR/TN-468+STR), University Corporation for Atmospheric Research, https://doi.org/10.5065/D6DZ069T, 2005.

Skamarock, C., Klemp, B., Dudhia, J., Gill, O., Barker, D., Duda, G., Huang, X., Wang, W., and Powers, G.: A Description of the Advanced Research WRF Version 3 (No. NCAR/TN-475+STR), University Corporation for Atmospheric Research, https://doi.org/10.5065/D68S4MVH, 2008.

Skamarock, C., Klemp, B., Dudhia, J., Gill, O., Liu, Z., Berner, J., Wang, W., Powers, G., Duda, G., Barker, D., and Huang, X.: A Description of the Advanced Research WRF Model Version 4.3

(No. NCAR/TN-556+STR), University Corporation for Atmospheric Research, https://doi.org/10.5065/1dfh-6p97, 2021.

Soleimanian, E., Wang, Y., Li, W., Liu, X., Griggs, T., Flynn, J., Walter, P. J., and Estes, M. J.: Understanding ozone episodes during the TRACER-AQ campaign in Houston, Texas: The role of transport and ozone production sensitivity to precursors, Sci. Total Environ., 900, 165881, https://doi.org/10.1016/j.scitotenv.2023.165881, 2023.

TCEQ: SIP Revision: Regional Haze, TEXAS COMMISSION ON ENVIRONMENTAL QUALITY, https://www.tceq.texas.gov/downloads/air-quality/sip/haze/revisions/2021rhsip_ado.pdf (last access: March 2023), 2021.

The Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., Earl, N., Starkman, N., Bradley, L., Shupe, D. L., Patil, A. A., Corrales, L., Brasseur, C. E., Nöthe, M., Donath, A., Tollerud, E., Morris, B. M., Ginsburg, A., Vaher, E., Weaver, B. A., Tocknell, J., Jamieson, W., Kerkwijk, M. H. van, Robitaille, T. P., Merry, B., Bachetti, M., Günther, H. M., Authors, P., Aldcroft, T. L., Alvarado-Montes, J. A., Archibald, A. M., Bódi, A., Bapat, S., Barentsen, G., Bazán, J., Biswas, M., Boquien, M., Burke, D. J., Cara, D., Cara, M., Conroy, K. E., Conseil, S., Craig, M. W., Cross, R. M., Cruz, K. L., D'Eugenio, F., Dencheva, N., Devillepoix, H. A. R., Dietrich, J. P., Eigenbrot, A. D., Erben, T., Ferreira, L., Foreman-Mackey, D., Fox, R., Freij, N., Garg, S., Geda, R., Glattly, L., Gondhalekar, Y., Gordon, K. D., Grant, D., Greenfield, P., Groener, A. M., Guest, S., Gurovich, S., Handberg, R., Hart, A., Hatfield-Dodds, Z., Homeier, D., Hosseinzadeh, G., Jenness, T., Jones, C. K., Joseph, P., Kalmbach, J. B., Karamehmetoglu, E., Kałuszyński, M., Kelley, M. S. P., Kern, N., Kerzendorf, W. E., Koch, E. W., Kulumani, S., Lee, A., Ly, C., Ma, Z., MacBride, C., Maljaars, J. M., Muna, D., Murphy, N. A., Norman, H., O'Steen, R., Oman, K. A., Pacifici, C., Pascual, S., Pascual-Granado, J., Patil, R. R., Perren, G. I., Pickering, T. E., Rastogi, T., Roulston, B. R., Ryan, D. F., Rykoff, E. S., Sabater, J., Sakurikar, P., Salgado, J., Sanghi, A., Saunders, N., Savchenko, V., Schwardt, L., Seifert-Eckert, M., Shih, A. Y., Jain, A. S., Shukla, G., Sick, J., Simpson, C., Singanamalla, S., Singer, L. P., Singhal, J., Sinha, M., Sipőcz, B. M., Spitler, L. R., Stansby, D., Streicher, O., Šumak, J., Swinbank, J. D., Taranu, D. S., Tewary, N., Tremblay, G. R., de Val-Borro, M., Van Kooten, S. J., Vasović, Z., Verma, S., Cardoso, J. V. d. M., Williams, P. K. G., Wilson, T. J., Winkel, B., Wood-Vasey, W. M., Xue, R., Yoachim, P., Zhang, C., and Zonca, A.: The Astropy Project: Sustaining and Growing a Community-oriented Open-source Project and the Latest Major Release (v5.0) of the Core Package, Astrophys. J., 935, 167, https://doi.org/10.3847/1538-4357/ac7c74, 2022.

U.S. Census Bureau: Metropolitan and Micropolitan Statistical Areas Population Totals: 2020–2022, Census Bureau, https://www.census.gov/data/tables/time-series/demo/popest/2020s-total-metro-and-micro-statistical-areas.html, last access: 1 September 2023.

US EPA: US EPA Air Quality System (AQS), US Environmental Protection Agency [data set], https://aqs.epa.gov/aqsweb/airdata/download_files.html, last access: July 2023.

van Geffen, J., Eskes, H., Compernolle, S., Pinardi, G., Verhoelst, T., Lambert, J.-C., Sneep, M., ter Linden, M., Ludewig, A., Boersma, K. F., and Veefkind, J. P.: Sentinel-5P TROPOMI NO$_2$ retrieval: impact of version v2.2 improvements and comparisons with OMI and ground-based data, Atmos. Meas. Tech., 15, 2037–2060, https://doi.org/10.5194/amt-15-2037-2022, 2022.

Verhoelst, T., Compernolle, S., Pinardi, G., Lambert, J.-C., Eskes, H. J., Eichmann, K.-U., Fjæraa, A. M., Granville, J., Niemeijer, S., Cede, A., Tiefengraber, M., Hendrick, F., Pazmiño, A., Bais, A., Bazureau, A., Boersma, K. F., Bognar, K., Dehn, A., Donner, S., Elokhov, A., Gebetsberger, M., Goutail, F., Grutter de la Mora, M., Gruzdev, A., Gratsea, M., Hansen, G. H., Irie, H., Jepsen, N., Kanaya, Y., Karagkiozidis, D., Kivi, R., Kreher, K., Levelt, P. F., Liu, C., Müller, M., Navarro Comas, M., Piters, A. J. M., Pommereau, J.-P., Portafaix, T., Prados-Roman, C., Puentedura, O., Querel, R., Remmers, J., Richter, A., Rimmer, J., Rivera Cárdenas, C., Saavedra de Miguel, L., Sinyakov, V. P., Stremme, W., Strong, K., Van Roozendael, M., Veefkind, J. P., Wagner, T., Wittrock, F., Yela González, M., and Zehner, C.: Ground-based validation of the Copernicus Sentinel-5P TROPOMI NO$_2$ measurements with the NDACC ZSL-DOAS, MAX-DOAS and Pandonia global networks, Atmos. Meas. Tech., 14, 481–510, https://doi.org/10.5194/amt-14-481-2021, 2021.

Vose, R. S., Applequist, S., Squires, M., Durre, I., Menne, M. J., Williams Jr., C. N., Fenimore, C., Gleason, K., and Arndt, D.: NOAA Monthly U.S. Climate Divisional Database (NClimDiv), NOAA National Climatic Data Center [data set], https://doi.org/10.7289/V5M32STR, 2014.

Wagner, T., Warnach, S., Beirle, S., Bobrowski, N., Jost, A., Puķīte, J., and Theys, N.: Investigation of three-dimensional radiative transfer effects for UV–Vis satellite and ground-based observations of volcanic plumes, Atmos. Meas. Tech., 16, 1609–1662, https://doi.org/10.5194/amt-16-1609-2023, 2023.

Weather Underground: Houston, TX Weather History, Weather Underground, https://www.wunderground.com/history/daily/us/tx/houston/KHOU, last access: 29 October 2023.

Wiedinmyer, C., Akagi, S. K., Yokelson, R. J., Emmons, L. K., Al-Saadi, J. A., Orlando, J. J., and Soja, A. J.: The Fire INventory from NCAR (FINN): a high resolution global model to estimate the emissions from open burning, Geosci. Model Dev., 4, 625–641, https://doi.org/10.5194/gmd-4-625-2011, 2011.