# An intercomparison of satellite, airborne, and ground-level observations with WRF-CAMx simulations of NO$_2$ columns over Houston, TX 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, California, 94945, USA

[3]Saint Louis University, St. Louis, Missouri, 63103, USA

[4]NASA Langley, Hampton, Virginia, 23666, USA

*Correspondence to* 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, TX 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 x 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 (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 to 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 US (United States Census Bureau, 2022) – $NO_2$ 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 $PM_{2.5}$ exceed their respective US EPA National Ambient Air Quality Standards (NAAQS), Houston is in moderate nonattainment of the 8-hour 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 highways like the I-610 interstate, the I-10 interstate, and Beltway 8 have heavy vehicle traffic that are 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, TX. One of the devices employed during the TRACER-AQ campaign across its twelve 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 & Janz, 2014; Leitch et al., 2014; Nowlan et al., 2018). The GCAS instrument is an ultraviolet-visible (UV-VIS) spectrometer. Its data is 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 $NO_2$ pattens with uniquely fine-scale resolution (on average 560 x 250 $m^2$) and performed comprehensive measurements of $NO_2$

70    columns across large swaths of the city repeatedly up to three time 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 x 5.5 $km^2$ 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

75    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 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

80    instruments are UV-VIS spectrometers that measure spectrally resolved radiance data that is 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

85    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 remote-sensing column concentrations in Texas (Goldberg et al., 2022; W. Li et al., 2023; Soleimanian et al., 2023). CAMx can be coupled with meteorological models like the Weather Research and

90    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 to understand biases in simulated $NO_2$ and to identify under- or over-estimates of emissions from sectors and regions in the inventories that drive the model.

95    In this study, we leverage the unique coincidence of ground-based Pandora spectrometers, 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 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

100   measurements. Our comparisons across the three observational datasets clarifies the range of expected values of $NO_2$ column concentrations in Houston, TX and characterizes potential deficiencies and biases in observational products and simulated CAMx values. We investigate weekday-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

105   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 to characterize column concentrations of gases (e.g., $NO_2$) similar to remote-sensing and aircraft observations (Judd et al., 2020). 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 select data from a single Pandora instrument at each site. Pan #58 at La Porte, #61 at Aldine, and #25 at University of Houston were chosen for the following reasons. As indicated in Table 1, Pan #61 and Pan #58 clearly have the largest temporal coverage during the TRACER-AQ time period. While Pan #188 measured more frequently at the University of Houston than Pan #25, Pan #188 was operated on a tower about 70 meters 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 shaded and bolded in the table below. 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 section 2.2 and 2.3) to compare total column amounts.

**Table 1: Details on Pandora instrument operational time**

| Inst. # | Location | Lat. | Lon. | September 2021 Flight Days with Observations (Number of high- and medium-quality measurements per day) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 1st | 3rd | 8th | 9th | 10th | 11th | 23rd | 24th | 25th | 26th |
| 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 |

**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 a DOAS computing software (Danckaert et al., 2017). During TRACER-AQ,

135 GCAS collected data over the Houston metropolitan area across 12 days 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 degrees creating one gapless map of NO2 up to three times per flight day. $NO_2$ observations from GCAS are publicly available at the NASA Atmospheric Sciences Data Center (NASA/LARC/SD/ASDC, 2022). Observations from two

140 of the flight days – a test flight (August 30) and a flight over the Gulf of Mexico (September 27) are excluded from this study because they provided no meaningful data over Houston. Given the relatively short timeframe 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

145 Houston Hobby Airport (NASA, 2023; Weather Underground, 2023).

**Table 2: Basic meteorological conditions and notes during GCAS flights**

| Day of Sept 2021 | Day of the Week | High Temp | Wind direction | Additional note |
|---|---|---|---|---|
| 1 | Wed | 96 F | Weak SW winds | Thunderstorms from S to N, 11 AM to 4 PM |
| 3 | Fri | 93 F | Weak S winds | Scattered thunderstorms 12 PM to 4 PM |
| 8 | Wed | 94 F | N turning NE | Clear skies and no rain |
| 9 | Thurs | 95 F | N turning NE | Afternoon fair weather clouds, no rain |
| 10 | Fri | 93 F | NE turning E | Clear skies, no rain, some long-range smoke aloft |
| 11 | Sat | 93 F | E winds | Afternoon fair weather clouds, no rain |
| 23 | Thurs | 83 F | E winds | Clear skies, no rain, cold font overnight Sept 21 |
| 24 | Fri | 84 F | E turning SE | Clear skies, no rain |
| 25 | Sat | 87 F | NE turning E | Clear skies, no rain |
| 26 | Sun | 83 F | Calm then SE | Clear skies, afternoon fair weather clouds |

150 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 ($444 \times 444$ m$^2$) WRF-CAMx simulation used in this

analysis (section 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), that had a coarser spatial resolution ($0.25° \times 0.25°$). Lastly, to directly

155  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 is considered in this analysis.

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

160  Pandora measurements that are more than 15 minutes removed from a GCAS overpass and then identify the Pandora measurement time within this 30-minute window that most closely matches the GCAS overpass time. While we choose this 30-minute window as an upper-bound cut-off, 96% and 90% of all Pandora closest matches occur within a 20- and 15-minute 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

165  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 (section 2.3) and add this to GCAS observations. Additionally, we add a "above aircraft" but below troposphere partial column amount based on the CAMx simulation; we calculate the

170  column. In the 3 highest levels of CAMx that amounts to $0.57 \times 10^{15}$ molecules cm$^{-2}$ and 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 globally from April 30, 2018 to present (European Space Agency, 2021). The slant column measurements were converted into tropospheric vertical column amounts by subtracting off a stratospheric

175  $NO_2$ component and transforming the remaining tropospheric slant column to vertical column using an air mass factor. We download the publicly available data (https://data-portal.s5p-pal.com/products/no2.html; https://dataspace.copernicus.eu/) coincident with the TRACER-AQ campaign in September 2021 for overpasses of Houston, TX. 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.,

180  2021) 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 2.3.1 and 2.4.0 is the use of the $0.125° \times 0.125°$ Directional Lambertian Equivalent Reflectivity

185  (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 their surface reflectivity input into the AMF calculation based on MODIS observations. The other main difference in these products include 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

190    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

195    TROPOMI $NO_2$ observations (resolution of 3.5 x 5.5 $km^2$ at nadir) onto the WRF-CAMx grid ($444 \times 444$ $m^2$). 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. Temporally, we screen out all Pandora measurements that are more than 15 minutes removed from a TROPOMI

200    overpass time and then identify the Pandora measurement time within this 30-minute window that most closely matches the TROPOMI overpass time. While we choose this 30-minute window as an upper-bound cut-off, 100% and 97% of all Pandora closest matches occur within a 20- and 15-minute 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

205    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

210    of $3.0 \times 10^{15}$ is typical for Houston during summer (Geddes et al., 2018). Boersma et al. (2018) suggests that $0.5 \times 10^{15}$ molecules $cm^{-2}$ 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 K.-F. Li et al. (2021); they calculate a daytime stratospheric $NO_2$ column increase rate of $1.34 \times 10^{14}$

215    molecules $cm^{-2}$ starting at 7:00 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 – 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 NO₂**

220   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 timeframe. 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

225   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

230   when transferring WRF meteorology to CAMx. The 36 km CAMx domain (red) includes the continental US, 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 (1.333km (orange) and 0.444km (cyan)) were selected to include the most relevant GCAS flight tracks while considering computational expense.

235

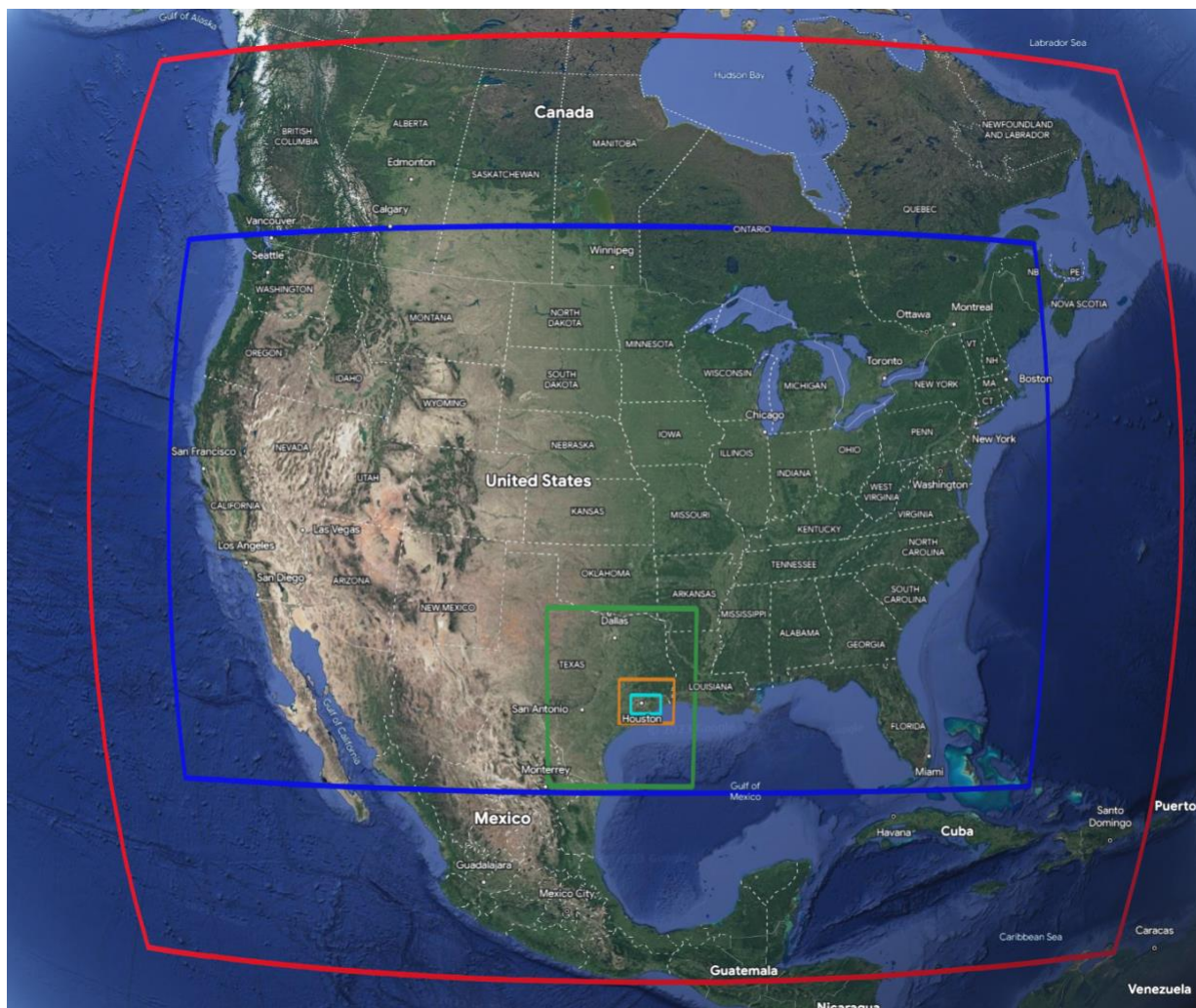



**Figure 1: Modeling domains used in the CAMx simulation for the 36 km resolution (red), 12 km resolution (blue), 4 km resolution (green), 1.333 km resolution (orange), and the 0.444 km resolution (cyan). Maps data provided by Google © 2020, Landsat / CopernicusData SIO, NOAA, U.S. Navy, NGA, GEBCO,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

Additional information on the WRF-CAMx modeling is included in the supplemental including the WRF physics options (Table S1), 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 for the WRF meteorological model; this GDAS data is also used for boundary conditions and data assimilation. We configured the output timesteps of WRF to 15 minutes for the higher resolution domains. Conducting WRF simulations at fine spatial resolutions (i.e., 4 km, 1.333 km, and 0.444 km) requires careful consideration of physical schemes that are sensitive to grid spacing. We turn off the convective cumulus parametrization scheme for the fine grids because WRF can explicitly simulate convection for them. For coarser grids we turn on the cumulus parametrization 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 km, 12km, 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, 255  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 eleven major electric generating units (EGUs) listed in Table S4. We download hourly data from Clean Air Markets Program Data (CAMPD) for the eleven EGUs for the August 30-Sep 27, 2021, period and stack parameters were based on the TCEQ 2019 emissions 260  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) (Guenther et al., 2012) version 3.2 for biogenic emissions, the Fire 265  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 larger domains. These two sources 270  represent a small fraction of emissions in the local Houston area that is 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 geo-located 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 275  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 tons per day (tpd) for a September weekday are presented in Table 3 below.

We evaluate the WRF simulation meteorology by comparing surface-level wind speed, direction, temperature, and water vapor mixing ratio to observations from sixteen ground-level monitors (Table S5-8 and Fig. S2-11) and 280  calculate the mean bias error (MBE), mean absolute error (MAE), and Pearson-R squared ($R^2$) statistics as defined in supplement table S2. Circular statistics are calculated using the Astropy circular statistics module for Python (The Astropy Collaboration, 2022). We obtain integrated surface data from NCDC in the DS3505 format (ftp://ftp.ncdc.noaa.gov/pub/data/noaa/); these data consist mainly of airport locations and have good meteorological siting and quality assurance procedures.

285     Generally, meteorological conditions simulated by WRF agree with ground-level observations especially on the
        more data rich non-cloudy days that are the most important for our intercomparison. Across all days, the WRF wind
        direction MBE was -8°, wind speed MBE was -0.02 m/s, temperature MBE was 0.39 K, and water vapor mixing
        ratio MBE was -1.45 g/kg. Across the eight non-cloudy days, hourly and site-specific – across the sixteen monitors –
        WRF wind direction ($R^2$=0.3 to 0.8; MAE=14° to 32°), wind speed ($R^2$=0.1 to 0.5; MAE=0.94 m/s to 1.35 m/s),
290     temperature ($R^2$=0.69 to 0.81; MAE=0.93 K to 1.18 K), and water vapor mixing ratio ($R^2$=0.28 to 0.78; MAE=0.87
        g/kg to 3.11 g/kg) performed moderately compared to observations given the fine spatial and temporal resolution.
        Additionally, we compare simulated hourly $NO_2$ (Fig. S12) and maximum daily eight-hour average or "MDA8" $O_3$
        (Fig. S13) to observations from seventeen TCEQ continuous air monitoring stations (CAMS) operating in Houston.
        We find poor performance and a strong negative bias in the simulated surface-level $NO_2$ (NMB=-59%) and
295     generally good performance in the simulated surface-level MDA8 $O_3$ (NMB=-2.5%) compared to observations.

**Table 3: CAMx 444 × 444 m² domain-wide summary of average September weekday emissions by sector in units of tons per day (tpd).**

| Emission Sector | Spatial Resolution | NO$_X$ (tpd) | VOC (tpd) |
|---|---|---|---|
| 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 |

**2.5 Diurnal Comparison**

300    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

305    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 – and GEOS-CF grid cell in which the Pandora instrument is located. Temporally,

310    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

315    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). We download all hourly data for 2021 for the US and filter the TRACER-AQ flight days and for monitors in Harris County. We identify the median hourly concentrations
320   across these monitors and the TRACER-AQ flight days.

## 3 Results

### 3.1 Comparisons to 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
325   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 is only in direct-sun mode in cloud-free scenes.  AMF for Pandora is analogous to pathlength through the atmosphere relative to the vertical path.  Since all the signal is from direct sun path (extremely minimal scattering), this is purely geometric.
330   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 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
335   of GCAS, TROPOMI, and CAMx outside of urban Houston. We consider the performance of GCAS processed with a CAMx-based AMF (Fig. 2A), TROPOMI processed with a CAMx-based AMF (Fig. 2B), and CAMx (Fig. 2C) and the performance of GCAS and TROPOMI with the operational AMFs (Fig. 2D and 2E) individually and then intercompare the three datasets across the ten GCAS observation days (Fig. 2G).

When comparing the observational and simulated datasets to Pandora observations we consider the total column
340   NO$_2$ – 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 users 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
345   amount that we add to GCAS and CAMx when we compare them to Pandora; this corresponds to a mean value of $3.0 \times 10^{15}$. For GCAS, we apply an additional amount to account for the NO$_2$ column in the upper troposphere – above the aircraft and below the tropopause. We calculate the column of levels 27-29 that correspond to 9400-18100 m above sea-level – that extends roughly from the height that GCAS flies at of around 9100 m to the tropopause – and apply this to the GCAS results; this corresponds to a value of $0.57 \times 10^{15}$ molecules cm$^{-2}$. For all results that

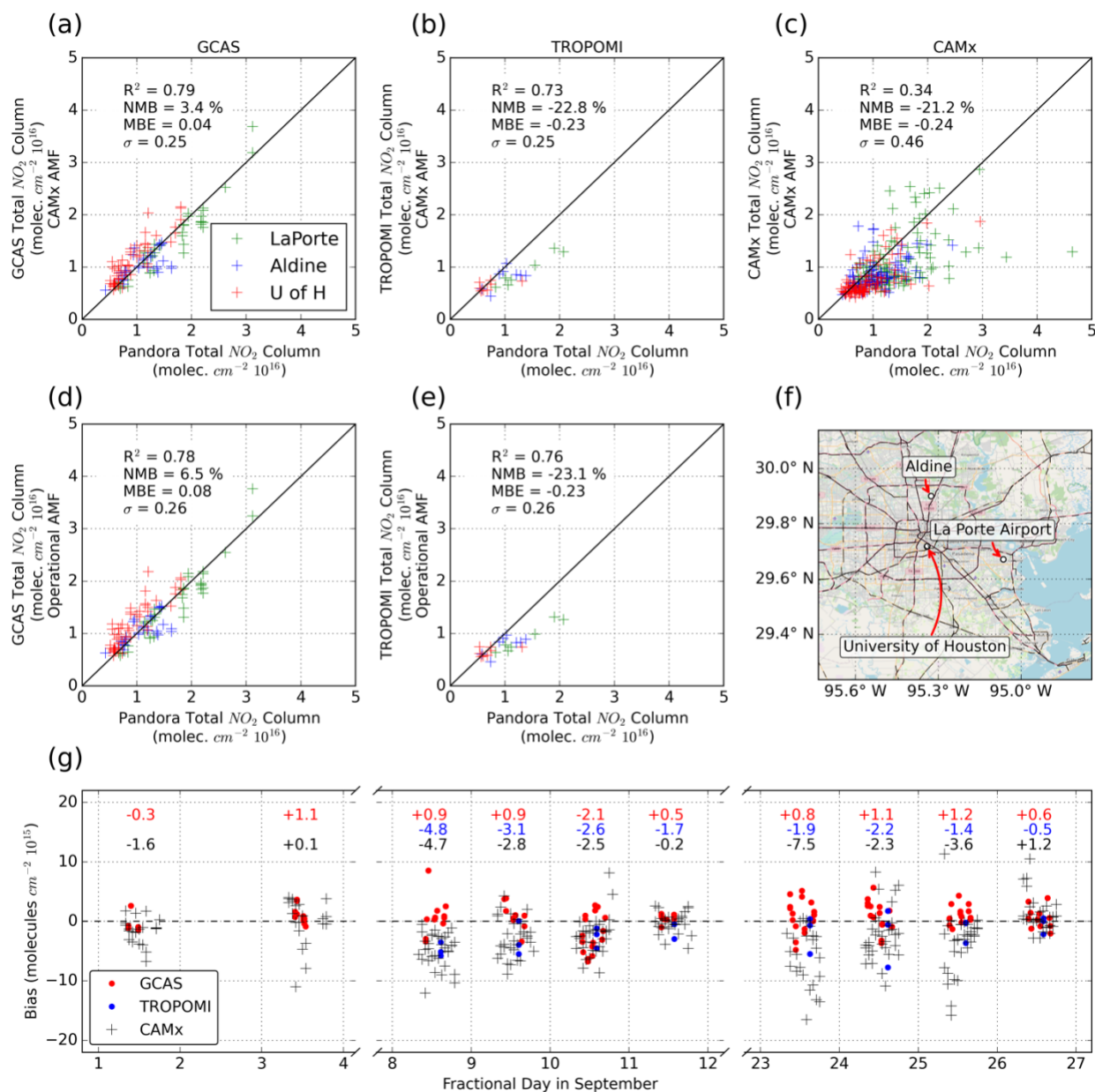
350     include comparison to Pandora we present total NO$_2$ columns; for results where we only intercompare GCAS, TROPOMI, and CAMx we compare the tropospheric column. All statistical measures (e.g., R$^2$) are defined in the supplement.

355

**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) and GCAS (D) and TROMPOMI 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 minutes of a Pandora observation are considered. GCAS flight times generally ranged from 8:00 AM-4:00 PM CDT. TROPOMI overpasses occurred around 1:30 PM local time. Color coding indicates which of the Pandora instruments $NO_2$ column concentrations are being compared against indicated in the legend in subplot 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.**

In Fig 2A-C we characterize the performance of the observational and simulated datasets of $NO_2$ column concentrations across the three sites in Houston. For each of the GCAS flight days, we compare GCAS and TROPOMI observation 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 to measurements from Pandora. Use of the CAMx AMF in place of the operational AMF had a minimal impact on comparisons to 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 scenes. This negative bias may be attributable to the coarser resolution of TROPOMI compared to 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. Similar to GCAS, use of the CAMx AMF in place of the operational AMF for TROPOMI had a minimal impact on comparisons to 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 to 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 compared to 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 generally find good agreement in the WRF simulated meteorology (Table S5-S8 and Fig.

S2-11); however, the non-systematic differences in wind direction on the order of 20° would likely degrade correlation between observed and simulated $NO_2$ columns. We do not find notable differences in correlation or error on windier days than on calmer days (Table S5-S6). Given that there is no apparent systematic bias in the

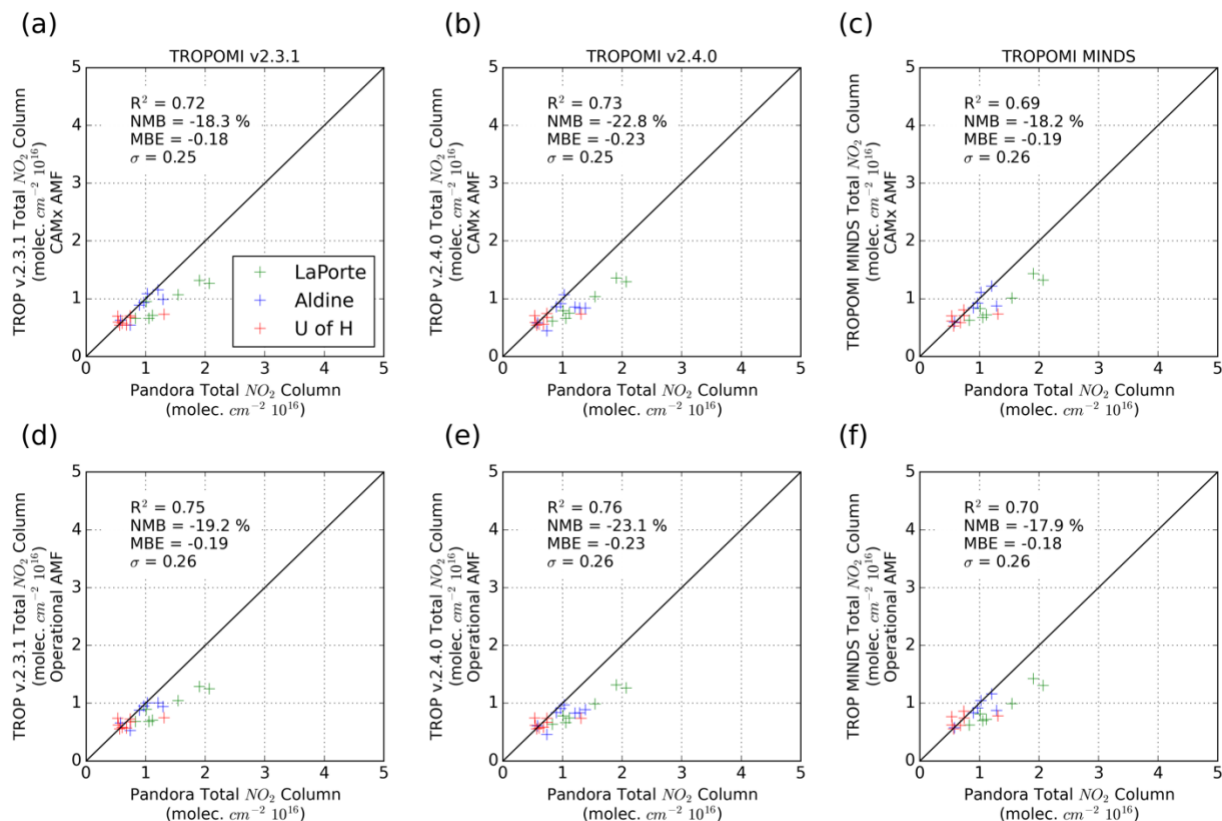
400    meteorology, the negative bias in $NO_2$ columns is likely attributable to an underestimate of $NO_X$ emissions. Additionally, there can be substantial differences in vertical mixing coefficients in different schemes in the models, and these can impact the biases in column concentrations (de Foy et al., 2007; Riess et al., 2023).

In Fig. 2G, we intercompare the daily variability in biases across the ten GCAS flight days. There were no

405    TROPOMI data for the first two flight days because cloud coverage blocked TROPOMI observations at the Pandora sites during its overpass time. The daily average bias of GCAS observations were consistently small throughout the entire period: they ranged from -2.1 to +1.2 molecules $cm^{-2}$ $10^{15}$ on September 10th and the 3rd and 24th, respectively. TROPOMI observations were consistently biased systematically low: they ranged from -4.8 to -0.5 molecules $cm^{-2}$ $10^{15}$; however, on all days except the 26th daily averaged TROPOMI biases were more negatively biased than -1.3

410    molecules $cm^{-2}$ $10^{15}$ compared to 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 September 3rd, there was little bias in simulated columns compared to Pandora measurements, and on other days such as September 26th, there was a minor high bias (+1.2 molecules $cm^{-2}$ $10^{15}$); however, on most days there was a negative bias that was the strongest on September 23rd when $NO_2$ columns were biased as low as -7.5 molecules $cm^{-2}$ $10^{15}$. Generally,

415    simulated CAMx columns perform better on weekend days (11th, 25th, and 26th) which is investigated in greater detail in section 3.4.

**3.2 Comparisons of different TROPOMI algorithms to Pandora Observations**

420    **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 minutes of a Pandora observation are considered with one exception: data from September 11$^{th}$, 2021, was missing from the NASA MINDS product and so values in plots C and F exclude this day.**

425

We intercompare TROPOMI observations to Pandora measurements across three different algorithms: version 2.3.1 (Fig. 3 A, 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 to 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%), version 2.3.1 is improved ($r^2$=0.72 and NMB=-18.3%) while the NASA MINDS product performs comparably ($r^2$=0.69 and NMB=-18.2%) to version 2.3.1. Notably, NASA MINDS data for September 11$^{th}$ are missing so these data are excluded from panels C 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 worsened bias compared to 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**

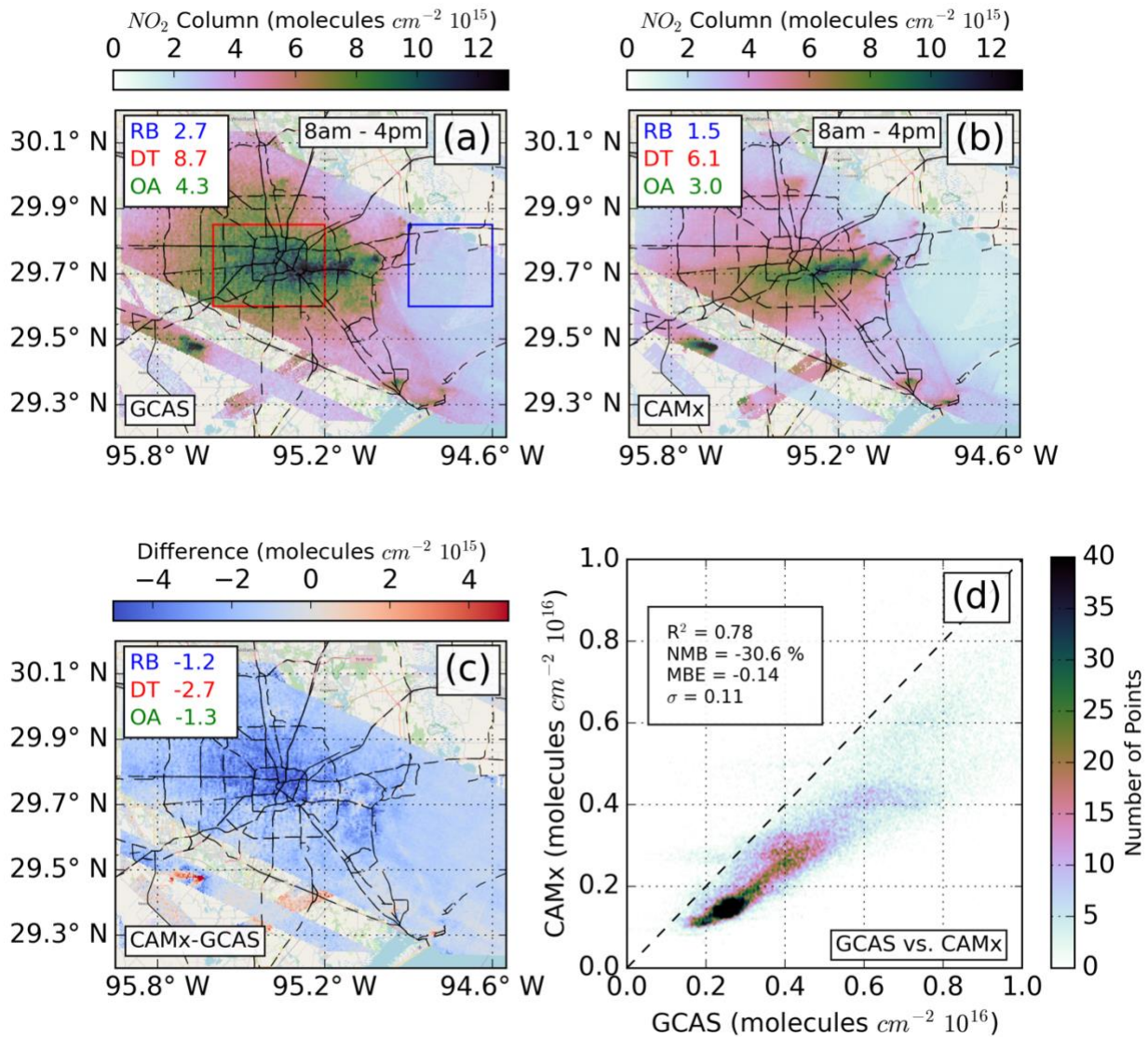
The comparisons between Pandora measurements and the datasets indicate that GCAS observations are in best agreement with Pandora. While TROPOMI performs worse than GCAS, it still decisively outperforms simulated $NO_2$ columns from CAMx at Pandora sites in both correlation and bias despite its coarser resolution. With the above in mind, in this section we present $NO_2$ columns observed from GCAS and TROPOMI and simulated from CAMx at the $444 \times 444$ 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 ten 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).

450

**Figure 4: Comparison of GCAS observations to CAMx simulated NO₂ columns across all data during GCAS overpasses (generally 8 am – 4 pm). Temporally averaged GCAS NO₂ columns (A), temporally averaged simulated CAMx NO₂ 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:**
455 **downtown or "DT" (red), the low emissions East Galveston rural Bay or "RB" (blue), and all other areas or "OA" (green) and calculate the averages in the top left of each chart. © OpenStreetMap contributors *2023*. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.**

460 When considering data from all GCAS overpasses (Fig. 4A, B, C) we observe a consistent negative bias in the CAMx product compared to GCAS observations throughout the domain that worsens in the downtown (DT) area (-2.7

molecules cm$^{-2}$ 10$^{15}$) compared to background levels in the rural East Galveston Bay (RB) (-1.2 molecules cm$^{-2}$ 10$^{15}$). Near the W A Parish power station in the southwestern area of the domain there are 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

465    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 minutes from TROPOMI overpass for each day

470    and then temporally average the observations across the GCAS flight days (Fig. 5A-C). We difference, both absolutely (Fig. 5D-F) and relatively (Fig. 5G-I), the three pairs of datasets and present them in scatter density plots (Fig. 5J-L). We focus on three regions: downtown Houston (DT) (red), the rural East Galveston Bay (RB) (blue), and all other areas (OA) (green) and 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

475    individual flight days are presented in the supplemental (Fig. S14-23)

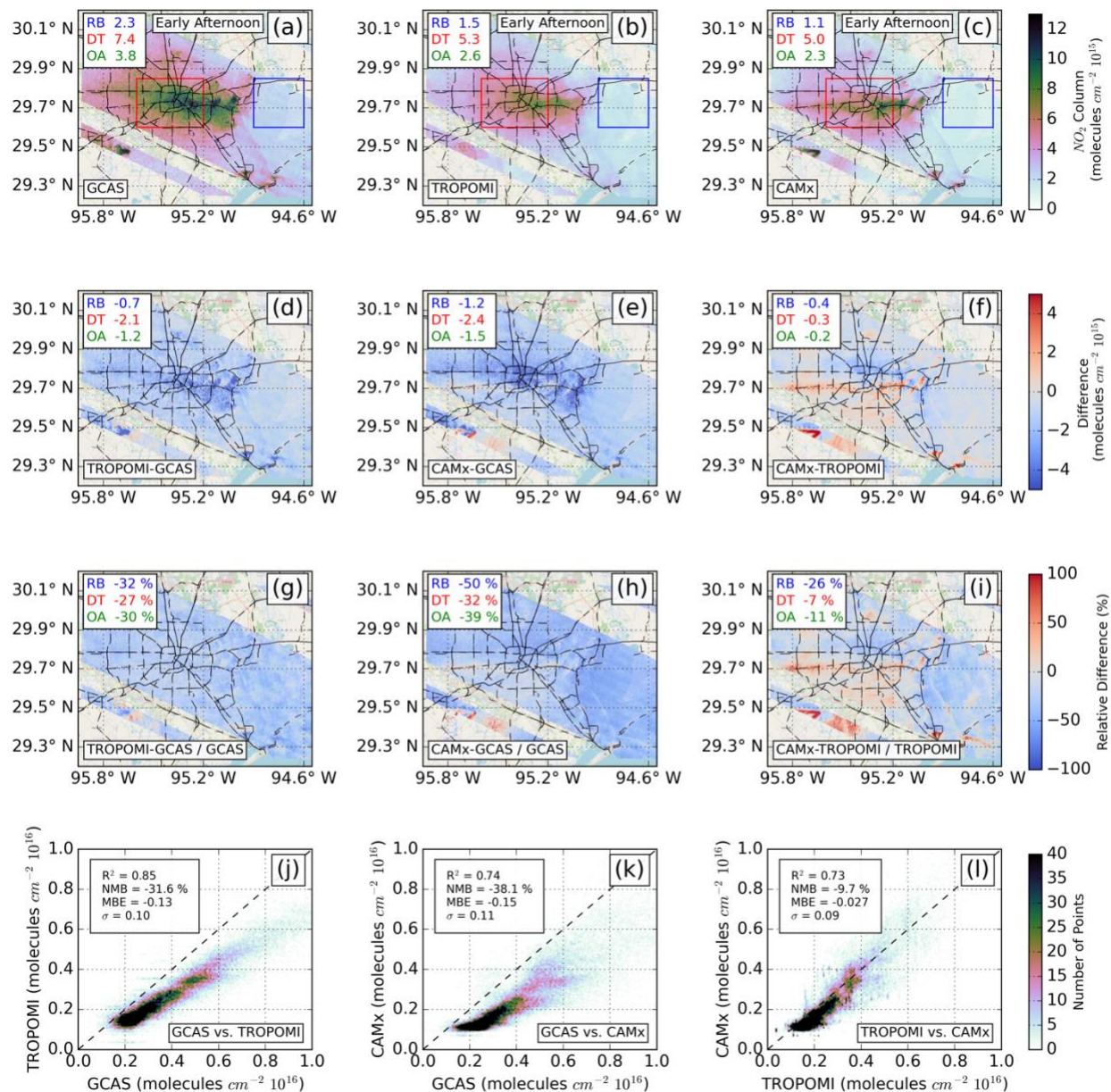



**Figure 5: Spatial distribution of GCAS (A), TROPOMI (B), and CAMx (C) NO₂ columns averaged across the ten GCAS flight days when within 90 minutes of each TROPOMI overpass representing early afternoon NO₂ columns. We identify three distinct areas: downtown or "DT" (red), the low emissions East Galveston Bay or "RB" (blue), and all other areas or "OA" (green) and calculate the averages in the top left 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.**

First, we consider the spatial distribution of $NO_2$ columns from GCAS (Fig. 5A), TROPOMI (Fig. 5B), and CAMx

490      (Fig. 5C) independently. For all three datasets, $NO_2$ 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 $NO_2$ peaks associated with point sources like those from W A Parish, Texas City, and Baytown

and in the Ship Channel. A map of the major point sources discussed in this work is included in the supplemental

(Fig. S24). The coarser resolution of TROPOMI leads to fewer identifiable peaks associated with point sources;

495      however, there are slightly elevated observed values near the W A Parish and Texas City power plants and the Ship

Channel. Observations from GCAS and TROPOMI reveal a more diffuse peak in $NO_2$ columns in and around

downtown Houston that includes elevated levels of $NO_2$ in the western part of the city. Simulated columns from

CAMx, on the other hand, primarily estimate higher $NO_2$ values in the eastern area of downtown Houston and have

lower $NO_2$ values in the western area of the city.

500

We next consider the three products compared to one another through three methods: absolute difference (Fig. 5D-

F), relative difference (Fig. 5G-I), and scatter density plots (Fig. 5J-L). We intercompare these three products by

isolating three sets of pairs: that is GCAS and TROPOMI, GCAS and CAMx, and TROPOMI and CAMx.

505      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 molecules cm$^{-2}$ 10$^{15}$ 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

510      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 section

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

515      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 TROPOMI observes higher $NO_2$ columns than GCAS. This is likely attributable to the coarser resolution of

TROPOMI that results in peaks of $NO_2$ to be spread into surrounding areas that are in the same TROPOMI grid cell.

520      Second, comparing GCAS to 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 to elsewhere in the

domain (-50%) (Figure 5H) while downtown and background levels are similarly biased at 32% and 39%,

respectively. This lower bias in the low emission 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

525      from -2.4 to -1.2 molecules cm$^{-2}$ 10$^{15}$ (Figure 5E). Visually, the negative bias in CAMx appears to be stronger in downtown and to the west, east, and north-west of downtown and less to the south and south-west of downtown. Overall, GCAS and CAMx are well correlated (r$^2$=0.74) (Figure 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

530      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

535      simulated columns are lowest compared to 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 to 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 south-west perhaps in conjunction

540      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).

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$^2$).

545      Here, we characterize the effect of the coarser resolution of TROPOMI by performing an additional comparison of the three datasets at the 0.05° × 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 supplemental: 0.25° × 0.25°

550      (Fig. S25) and 0.1° × 0.1° (Fig. S26).

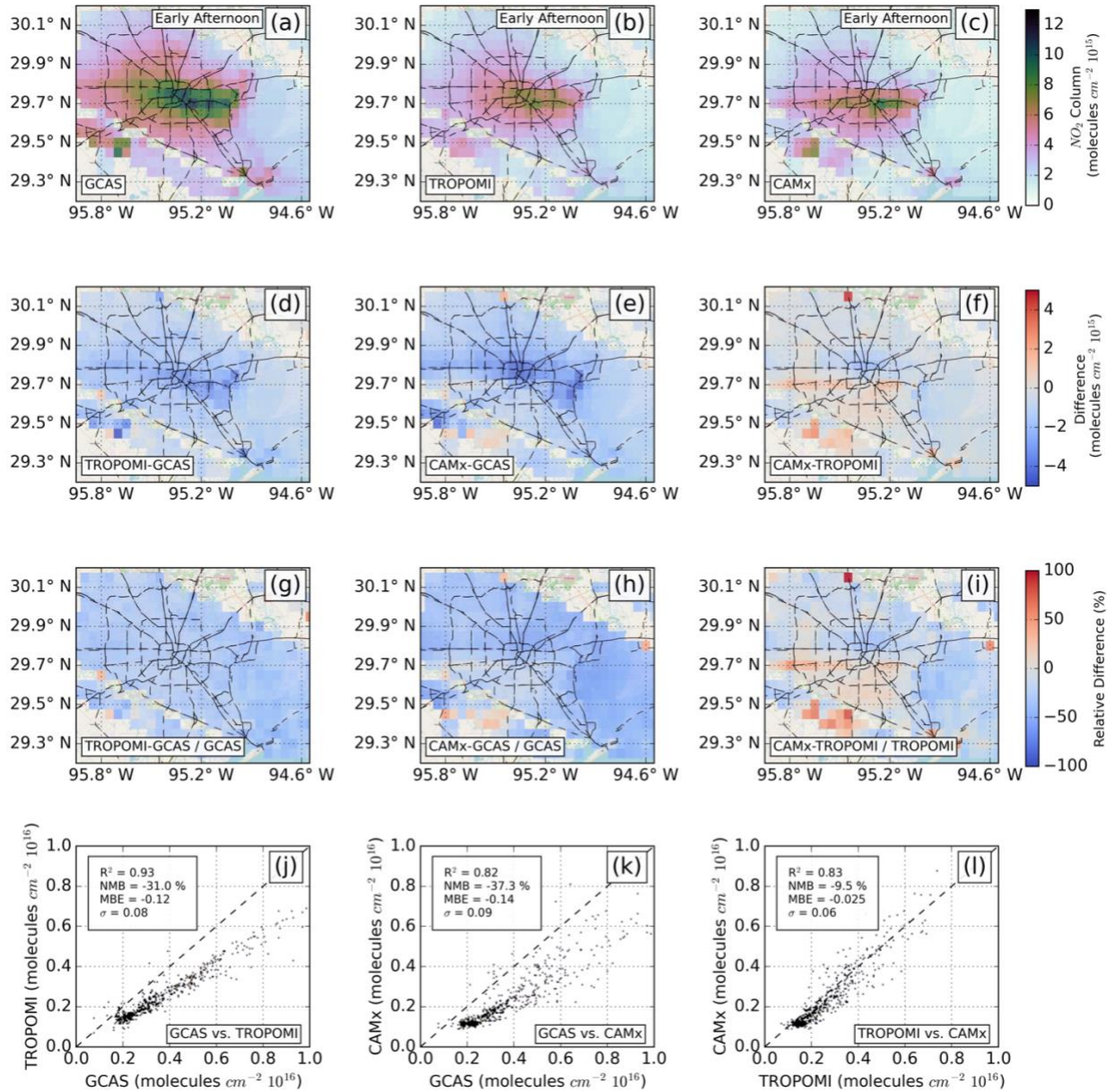



**Figure 6: Spatial distribution of GCAS (A), TROPOMI (B), and CAMx (C) at the 0.05° × 0.05° resolution averaged across the ten GCAS flight days when within 1.5 hours of each TROPOMI overpass representing early afternoon NO₂ columns. Absolute differences between GCAS and TROPOMI (D), GCAS and CAMx (E), 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.**

560

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 to 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

565 compared to TROPOMI (Fig. 6A-C). Considering the spatial distribution of absolute (Fig. 6D-F) and relative (Fig. 6G-I) differences between the three products, the low bias in TROPOMI compared to 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,

570 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-L).

**3.5 Weekend vs. weekday patterns across the datasets**

Three of the ten GCAS flight days occurred on weekends (September 11th, 25th, and 26th) and observations from GCAS and TROPOMI – along with simulated $NO_2$ columns from CAMx –exhibited different patterns on weekends

575 versus on weekdays (September 1st, 3rd, 8-10th, 23rd, 24th). This difference in observed and simulated patterns is explored in greater detail in this section, first through comparisons to Pandora measurements (Fig. 7) and then through spatial comparisons of the products on weekdays versus on weekends (Fig. 8). When interpreting these results, it should be considered that weekend data is limited to only three days. This data sparsity introduces a high degree of uncertainty in conclusions derived from this analysis. Day to day changes in meteorological conditions are

580 likely responsible for some of the exhibited differences so they cannot solely be attributed to differences in emission patterns.

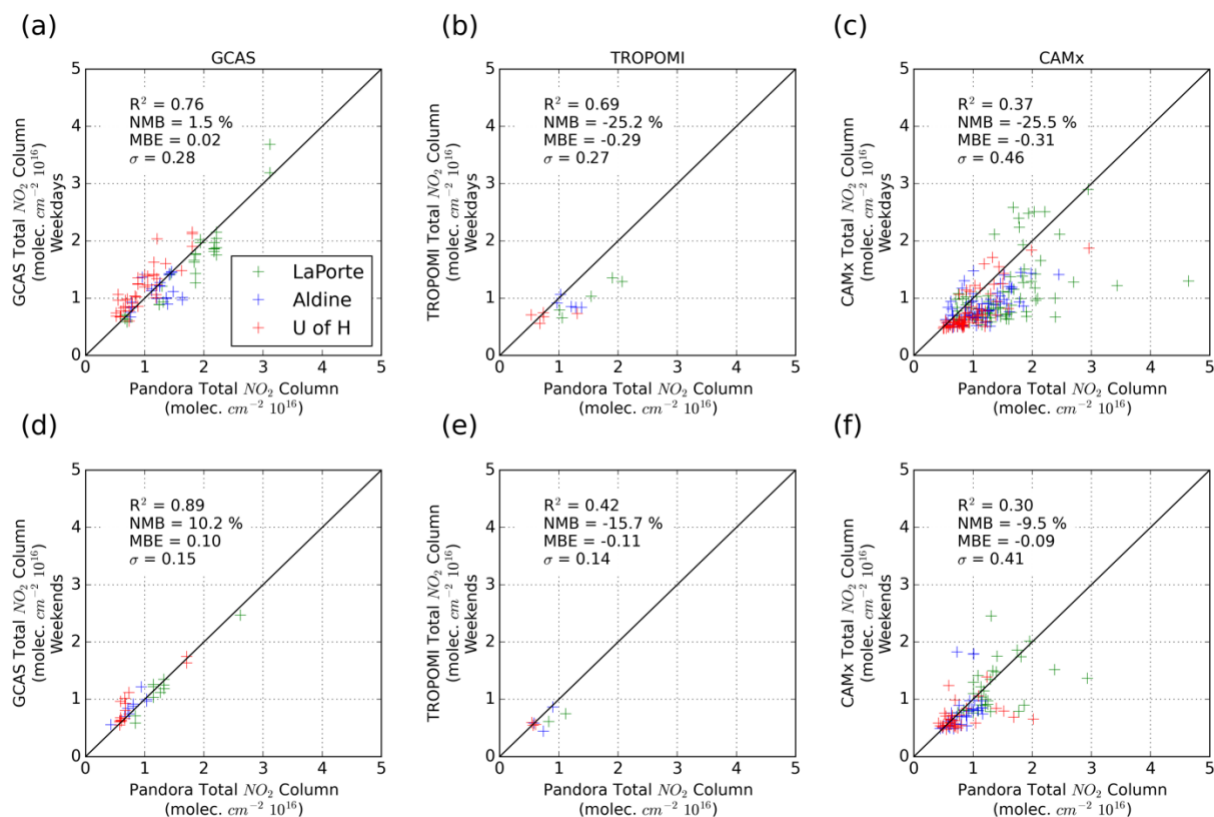



**Figure 7: Comparison of GCAS (A), TROPOMI (B), and CAMx (C) to Pandora on weekdays and of GCAS (D), TROPOMI (E), and CAMx (F) to Pandora on weekends. Data from all possible overpasses coincident within 15 minutes of a Pandora observation are considered. GCAS flight times generally ranged from 8:00 AM-4:00 PM CDT. TROPOMI overpasses occurred around 1:30 PM local time.**

First, we consider how comparisons of the observational datasets – GCAS and TROPOMI – with Pandora change on weekends compared to 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 observations are slightly better correlated to Pandora measurements on weekends ($r^2$=0.89 versus $r^2$=0.76); however, TROPOMI observations are worse correlated ($r^2$=0.42 versus $r^2$=0.69) that 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 conditions or if it is attributable to some systematic bias in the instruments.

Simulated $NO_2$ columns from CAMx exhibit clearer weekday versus 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 versus $r^2$=0.37) the negative bias in simulated columns compared to Pandora measurements is reduced on weekends (NMB=-25.5% versus NMB=-9.5%).

605

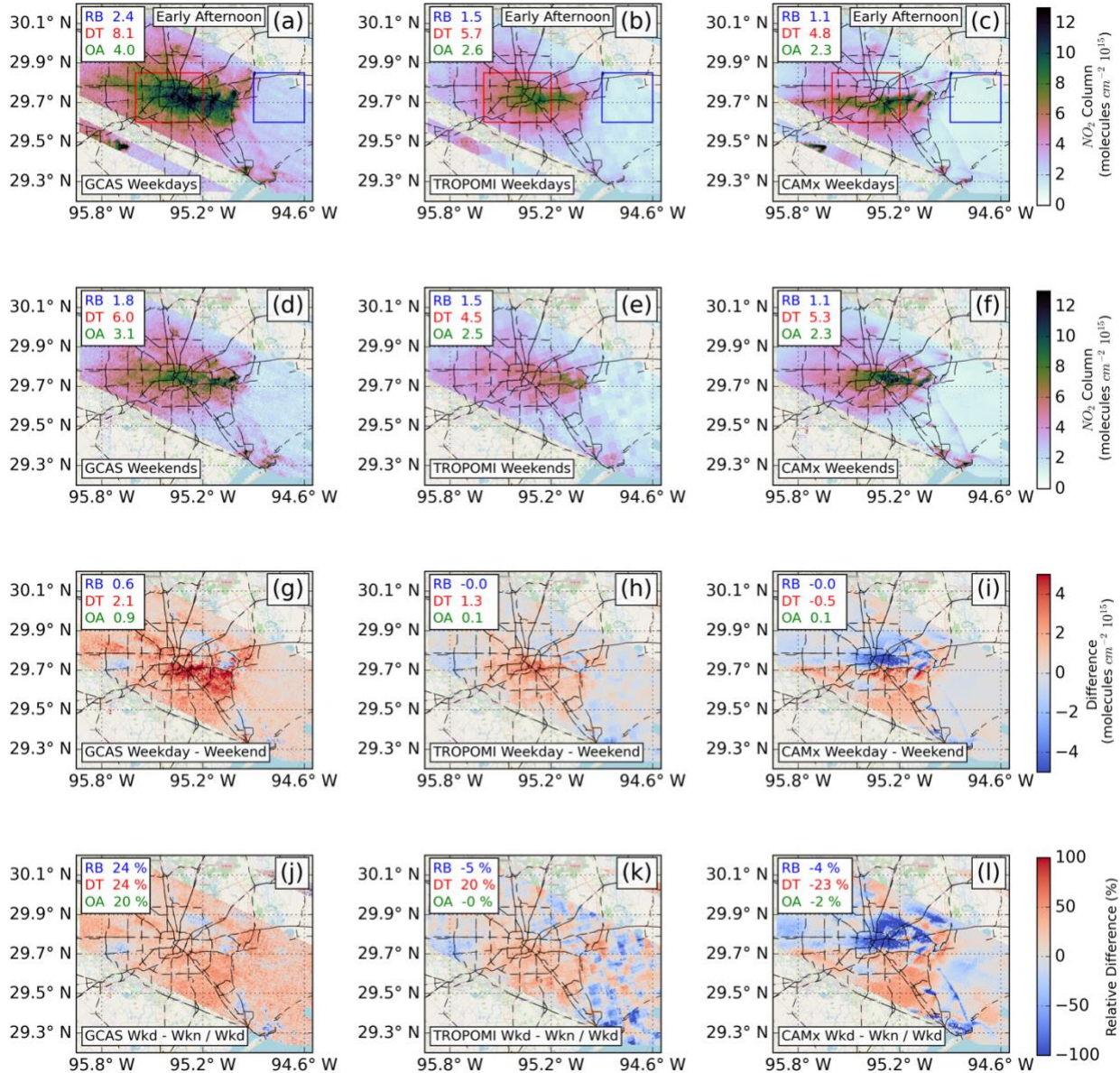



**Figure. 8: Spatial Distribution of GCAS, TROPOMI, and CAMx NO₂ columns on weekdays (A-C), weekends (D-F), and the absolute difference between weekdays and weekends (G-I) and relative difference (J-L). Data are averaged across the GCAS flight days corresponding to weekdays or weekends when within 1.5 hours of**
610 **each TROPOMI overpass representing early afternoon NO₂ columns. We identify three distinct areas: downtown or "DT" (red), the low emissions East Galveston Bay or "RB" (blue), and all other areas or "OA" (green) and calculate the averages in the top left of each chart. © OpenStreetMap contributors 2023. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.**

615      GCAS and TROPOMI observations of $NO_2$ 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 molecules $cm^{-2}$ $10^{15}$ (24%) and 0.6 molecules $cm^{-2}$ $10^{15}$ (24%) higher, respectively, on weekdays than on weekends. In other areas of Houston, GCAS observations on weekdays are higher than weekends but not to the same degree (20%). A similar pattern occurs for TROPOMI, in downtown Houston TROPOMI columns are 20%

620 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 $NO_2$ 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 $NO_2$ column concentrations of TROPOMI – it is difficult to examine this quantitatively.

625      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 southwest of downtown while columns on weekends are higher within downtown. These spatial patterns are further revealed in the difference plots (Fig. 8I and 8L) where the difference in weekday versus weekend values appear to be split right along I-10; north of I-10 weekday values are much lower than weekend values while south of I-10 the

630 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.

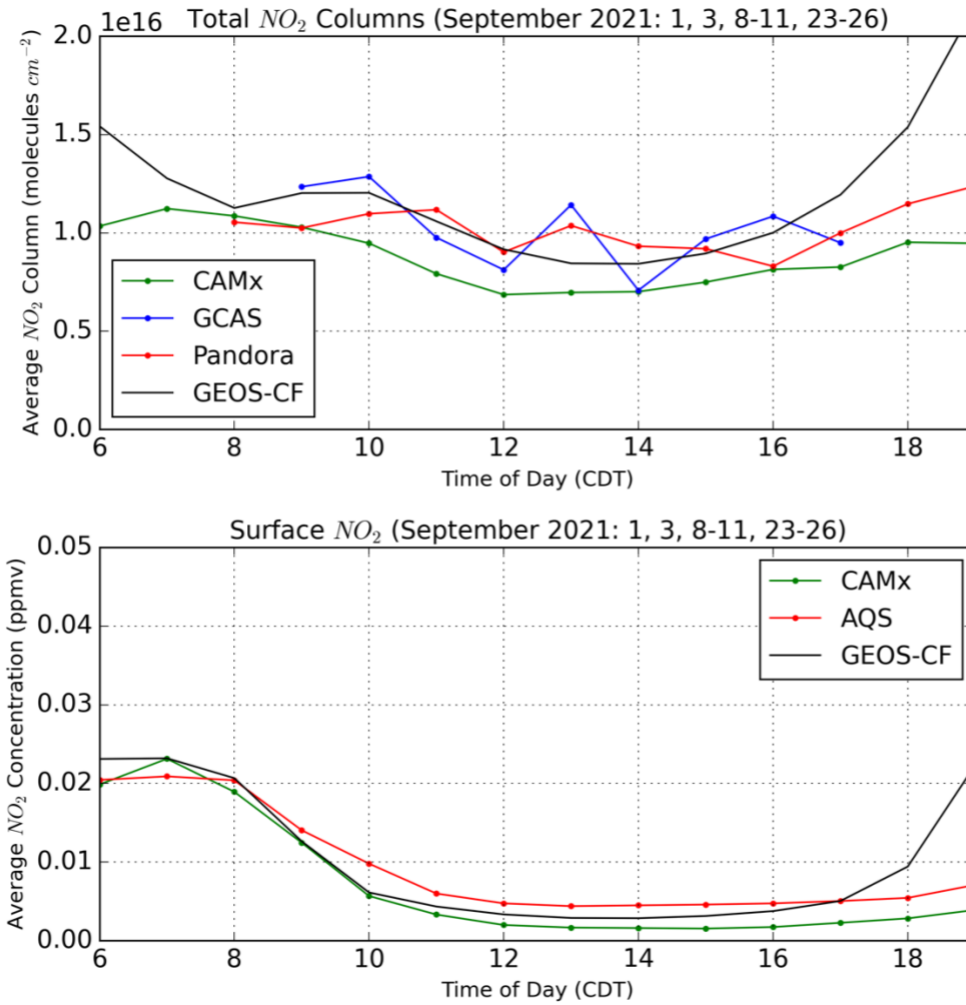
**3.6 Relevance to TEMPO: Diurnal patterns in column and surface NO₂**

635

**Figure 9: Diurnal patterns in Total NO₂ columns (top) 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 NO₂ concentrations (bottom) 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).**

640

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 (8:00 -10:00 local time) across the two simulated datasets (CAMx and GEOS-CF) and two observational datasets (GCAS and Pandora). Interestingly, between 9:00 and 11:00 local time, Pandora

645   column measurements show a slight increase, while model simulations show a slight decrease during the same time

interval. During midday and the afternoon (11:00 – 16:00 local time) – 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 local time), 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

650    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 (6:00 – 9:00 local time) before they begin to diverge with the two simulated produces maintaining comparable magnitudes with low biases compared to

655    surface monitors. At around midday to the afternoon (12:00 – 17:00 local time) both simulated products have a low bias compared to observed surface-level $NO_2$; however, the bias in CAMx concentrations is worse. Some of the apparent low bias may be related 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 local time), 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

660    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

665    were used to investigate $NO_2$ column densities in Houston, TX during the September 2021 TRACER-AQ campaign and to characterize strengths/weaknesses and uncertainties in the respective datasets. These observational datasets were then compared to 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 find 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

670    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 find differences between the v2.3.1, v2.4.0, and NASA MINDS product and find 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

675    AMF – performs comparably and 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 to 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

680    implicates 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 a MAE that ranged from 14° to 32° when compared to observations – and spatial correlations over larger extents match well.

685

When we compare the spatial distribution of TROPOMI observations to GCAS (Fig. 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 that is unable to capture the fine-scale features in $NO_2$ column concentrations that GCAS is able to. If coarse resolution is responsible for this low-bias, new instruments on

690 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 to 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

695 simulation to capture local meteorology and an underestimate of emissions in sectors that are more spatially located in downtown and western 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 and performance is better during other times of the day.

There is a clear negative bias in the CAMx simulated $NO_2$ columns compared to GCAS observations. Although we

700 evaluate the performance of WRF at the surface, we do not consider vertical advection and the vertical mixing scheme in WRF merits further investigation. 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. 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

705 we suspect it to be due to overmixing leading to the rapid removal of surface-level $NO_2$. Additionally, the low bias in the TROPOMI observations compared to 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

710 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 benefitted from three independent measurement datasets (i.e., Pandora, TROPOMI, and GCAS) that

715 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/). It is common to consider TROPOMI measurements as 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. Lastly, given the relatively minimal biases in WRF simulated meteorology compared to observations, low $NO_2$ biases in the simulated CAMx column concentrations imply that current TCEQ $NO_X$ emissions 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.

**Acknowledgements**

The authors of this paper acknowledge Elena Lind, Alex Kotsakis, the NASA Pandora Project and LuftBlick for their contribution in deploying, operating, and processing data from the Pandora spectrometers during TRACER-AQ, the NASA Tropospheric Composition Program and the Texas Commission on Environmental Quality for TRACER-AQ support, and the TRACER-AQ science team for their useful contributions. We also acknowledge the use of Google Earth for the background map used in Fig. 1. The authors acknowledge the use of OpenStreetMaps for the background maps in Fig. 2, 4, 5, 6, 8 and S2-14. The authors also acknowledge funding from the NASA Atmospheric Composition Modelling and Analysis Program (ACMAP) (80NSSC23K1002). The preparation of this manuscript was funded by a grant from the Texas Air Quality Research Program (AQRP) at The University of Texas at Austin through the Texas Emission Reduction Program (TERP) and the Texas Commission on Environmental Quality (TCEQ). The findings, opinions and conclusions are the work of the authors and do not necessarily represent findings, opinions, or conclusions of the AQRP or the TCEQ. This work contains modified Copernicus Sentinel-5 Precursor data processed by KNMI and post-processed by George Washington University.

**Code Availability**

The scripts used to process these data for the intercomparison are available by correspondence with the authors upon reasonable request.

**Data Availability**

Ground level EPA AQS observations are available from EPA Air Data: [https://www.epa.gov/outdoor-air-quality-data](https://www.epa.gov/outdoor-air-quality-data). GEOS-CF data are available from the GrADS data server: https://opendap.nccs.nasa.gov/dods/. TROPOMI $NO_2$ column data are publicly available from the Copernicus Data Space Ecosystem: https://dataspace.copernicus.eu. GCAS and CAMx data are publicly available via NASA's Atmospheric Sciences Data Center:

**Author Contributions**

D.G., developed the project design, J.J. and G.Y. set-up and conducted the WRF-CAMx simulations, L.J. and the TRACER-AQ science team measured and processed the GCAS and Pandora Data, D.G downloaded and processed the TROPOMI Data and regridded all data to the WRF-CAMx grid, M.O.N further processed the data to match temporally and spatially, conducted the intercomparison, wrote the manuscript, and generated the figures except Fig. 1 and Fig. S1 that were generated by J.J. and D.G., respectively, B.D. gave feedback on the methodology, all authors edited the manuscript and gave feedback on the figures.

**Competing Interests**

The authors declare that they have no conflict of interest.

**References**

Achakulwisut, P., Brauer, M., Hystad, P., & Anenberg, S. C. (2019). Global, national, and urban burdens of paediatric asthma incidence attributable to ambient NO2 pollution: Estimates from global datasets. *The Lancet Planetary Health*, *3*(4), e166–e178. https://doi.org/10.1016/S2542-5196(19)30046-4

Anenberg, S. C., Mohegh, A., Goldberg, D. L., Kerr, G. H., Brauer, M., Burkart, K., Hystad, P., Larkin, A., Wozniak, S., & Lamsal, L. (2022). Long-term trends in urban NO2 concentrations and associated paediatric asthma incidence: Estimates from global datasets. *The Lancet Planetary Health*, *6*(1), e49–e58. https://doi.org/10.1016/S2542-5196(21)00255-2

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., & Compernolle, S. C. (2018). Improving algorithms and uncertainty estimates for satellite $NO_2$ retrievals: Results from the quality assurance for the essential climate variables (QA4ECV) project. *Atmospheric Measurement Techniques*, *11*(12), 6651–6678. https://doi.org/10.5194/amt-11-6651-2018

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., & Pickering, K. E. (2013). A new stratospheric and tropospheric $NO_2$ retrieval algorithm for nadir-viewing satellite instruments: Applications to OMI. *Atmospheric Measurement Techniques*, *6*(10), 2607–2626. https://doi.org/10.5194/amt-6-2607-2013

Cede, A. (2021). *Manual for Blick Software Suite 1.8*.

Collaboration, T. A., 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., … Contributors, A. P. (2022). The Astropy Project: Sustaining and Growing a Community-oriented Open-source Project and the Latest Major Release (v5.0) of the Core Package*. *The Astrophysical Journal*, *935*(2), 167. https://doi.org/10.3847/1538-4357/ac7c74

Danckaert, T., Fayt, C., Roozendael, M. V., Smedt, I. D., Letocart, V., Merlaud, A., & Pinardi, G. (2017). *QDOAS Software user manual*.

de Foy, B., Lei, W., Zavala, M., Volkamer, R., Samuelsson, J., Mellqvist, J., Galle, B., Martínez, A.-P., Grutter, M., Retama, A., & Molina, L. T. (2007). Modelling constraints on the emission inventory and on vertical dispersion for CO and $SO_2$ in the Mexico City Metropolitan Area using Solar FTIR and zenith sky UV spectroscopy. *Atmospheric Chemistry and Physics*, *7*(3), 781–801. https://doi.org/10.5194/acp-7-781-2007

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

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., & Pusede, S. E. (2020). Observing Nitrogen Dioxide Air Pollution Inequality Using High-Spatial-Resolution Remote Sensing Measurements in Houston, Texas. *Environmental Science & Technology*, *54*(16), 9882–9895. https://doi.org/10.1021/acs.est.0c01864

DOC/NOAA/NWS/NCEP/EMC. (2023). *Global Data Assimilation System (GDAS)* [dataset]. https://www.ncei.noaa.gov/access/metadata/landing-page.

Dunlea, E. J., Herndon, S. C., Nelson, D. D., Volkamer, R. M., Martini, F. S., 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.,

800     Marquez, C., Blanco, S., Cardenas, B., Retama, A., Villegas, C. R. R., … Molina, M. J. (2007). Evaluation

of nitrogen dioxide chemiluminescence monitors in a polluted urban environment. *Atmos. Chem. Phys.*, 14.

EPA. (2023). *United States Environmental Protection Agency (EPA). "Clean Air Markets Program Data."*

*Washington, DC: Office of Atmospheric Protection, Clean Air Markets Division. Available from EPA's Air*

*Markets Program Data web site: Https://campd.epa.gov/.* [dataset]. https://campd.epa.gov/.

805     Eskes, H. J., Eichmann, K.-U., Lambert, J.-C., Loyola, D., Stein-Zweers, D., Dehn, A., & Zehner, C. (2023). *S5P*

*MPC Product Readme Nitrogen Dioxide*.

European Space Agency. (2021). *Copernicus Sentinel-5P TROPOMI* [dataset]. European Space Agency.

https://doi.org/10.5270/S5P-9bnp8q8

Ge, S., Wang, S., Xu, Q., & Ho, T. (2021). Source apportionment simulations of ground-level ozone in Southeast

810     Texas employing OSAT/APCA in CAMx. *Atmospheric Environment*, *253*, 118370.

https://doi.org/10.1016/j.atmosenv.2021.118370

Geddes, J. A., Martin, R. V., Bucsela, E. J., McLinden, C. A., & Cunningham, D. J. M. (2018). Stratosphere–

troposphere separation of nitrogen dioxide columns from the TEMPO geostationary satellite instrument.

*Atmospheric Measurement Techniques*, *11*(11), 6271–6287. https://doi.org/10.5194/amt-11-6271-2018

815     Goldberg, D. L., Harkey, M., de Foy, B., Judd, L., Johnson, J., Yarwood, G., & Holloway, T. (2022). Evaluating $NO_x$

emissions and their effect on $O_3$ production in Texas using TROPOMI $NO_2$ and HCHO. *Atmospheric*

*Chemistry and Physics*, *22*(16), 10875–10900. https://doi.org/10.5194/acp-22-10875-2022

Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., & Wang, X. (2012). The

Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): An extended and

820     updated framework for modeling biogenic emissions. *Geoscientific Model Development*, *5*(6), 1471–1492.

https://doi.org/10.5194/gmd-5-1471-2012

Herman, J., Cede, A., Spinei, E., Mount, G., Tzortziou, M., & Abuhassan, N. (2009). NO2 column amounts from

ground-based Pandora and MFDOAS spectrometers using the direct-sun DOAS technique:

Intercomparisons and application to OMI validation. *Journal of Geophysical Research: Atmospheres*,

825     *114*(D13). https://doi.org/10.1029/2009JD011848

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

Hudman, R. C., Moore, N. E., Mebust, A. K., Martin, R. V., Russell, A. R., Valin, L. C., & Cohen, R. C. (2012). Steps towards a mechanistic model of global soil nitric oxide emissions: Implementation and space based-constraints. *Atmospheric Chemistry and Physics*, *12*(16), 7779–7795. https://doi.org/10.5194/acp-12-7779-2012

Jia, J., Cheng, S., Liu, L., Lang, J., Wang, G., Chen, G., & Liu, X. (2017). An Integrated WRF-CAMx Modeling Approach for Impact Analysis of Implementing the Emergency PM2.5 Control Measures during Red Alerts in Beijing in December 2015. *Aerosol and Air Quality Research*, *17*(10), 2491–2508. https://doi.org/10.4209/aaqr.2017.01.0009

Jin, X., Zhu, Q., & Cohen, R. C. (2021). Direct estimates of biomass burning $NO_x$ emissions and lifetimes using daily observations from TROPOMI. *Atmospheric Chemistry and Physics*, *21*(20), 15569–15587. https://doi.org/10.5194/acp-21-15569-2021

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., & Williams, D. (2020). Evaluating Sentinel-5P TROPOMI tropospheric $NO_2$ column densities with airborne and Pandora spectrometers near New York City and Long Island Sound. *Atmospheric Measurement Techniques*, *13*(11), 6113–6140. https://doi.org/10.5194/amt-13-6113-2020

Judd, L. M., Sullivan, J. T., Lefer, B., Haynes, J., Jensen, M. P., & Nadkarni, R. (2021). *TRACER-AQ Science Plan*.

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., & Pawson, S. (2021). Description of the NASA GEOS Composition Forecast Modeling System GEOS-CF v1.0. *Journal of Advances in Modeling Earth Systems*, *13*(4), e2020MS002413. https://doi.org/10.1029/2020MS002413

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., … Williams, E. (2011). 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. *Atmospheric Chemistry and Physics*, *11*(22), 11361–11386. https://doi.org/10.5194/acp-11-11361-2011

Kowalewski, M. G., & Janz, S. J. (2014). Remote sensing capabilities of the GeoCAPE Airborne Simulator. *Earth Observing Systems XIX*, *9218*, 496–507. https://doi.org/10.1117/12.2062058

Lambert, J.-C., Claas, J., Stein-Zweers, D., Ludewig, A., Loyola, D., Sneep, M., & Dehn, A. (2023). *Quarterly Validation Report of the Copernicus Sentinel-5 Precursor Operational Data Products #19*.

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., & Haffner, D. (2022). *TROPOMI/S5P NO2 Tropospheric, Stratospheric and Total Columns MINDS 1-Orbit L2 Swath 5.5 km x 3.5 km, NASA Goddard Space Flight Center, Goddard Earth Sciences Data and Information Services Center*.

Lamsal, L. N., Martin, R. V., van Donkelaar, A., Steinbacher, M., Celarier, E. A., Bucsela, E., Dunlea, E. J., & Pinto, J. P. (2008). Ground-level nitrogen dioxide concentrations inferred from the satellite-borne Ozone Monitoring Instrument. *Journal of Geophysical Research: Atmospheres*, *113*(D16). https://doi.org/10.1029/2007JD009235

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., & Wang, J. (2014). The GeoTASO airborne spectrometer project. *Earth Observing Systems XIX*, *9218*, 487–495. https://doi.org/10.1117/12.2063763

Li, K.-F., Khoury, R., Pongetti, T. J., Sander, S. P., Mills, F. P., & Yung, Y. L. (2021). Diurnal variability of stratospheric column $NO_2$ measured using direct solar and lunar spectra over Table Mountain, California (34.38° N). *Atmospheric Measurement Techniques*, *14*(12), 7495–7510. https://doi.org/10.5194/amt-14-7495-2021

Li, W., Wang, Y., Liu, X., Soleimanian, E., Griggs, T., Flynn, J., & Walter, P. (2023). Understanding offshore high-ozone events during TRACER-AQ 2021 in Houston: Insights from WRF-CAMx photochemical modeling. *EGUsphere*, 1–21. https://doi.org/10.5194/egusphere-2023-1117

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., & Judd, L. (2023). Evaluating WRF-GC v2.0 predictions of boundary layer height and vertical ozone

profile during the 2021 TRACER-AQ campaign in Houston, Texas. *Geoscientific Model Development*, *16*(18), 5493–5514. https://doi.org/10.5194/gmd-16-5493-2023

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., & Krol, M. C. (2019). Quantification of nitrogen oxides emissions from build-up of pollution over Paris with TROPOMI. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-56428-5

Luke, W. T., Kelley, P., Lefer, B. L., Flynn, J., Rappenglück, B., Leuchner, M., Dibb, J. E., Ziemba, L. D., Anderson, C. H., & Buhr, M. (2010). Measurements of primary trace gases and NOY composition in Houston, Texas. *Atmospheric Environment*, *44*(33), 4068–4080. https://doi.org/10.1016/j.atmosenv.2009.08.014

Mazzuca, G. M., Ren, X., Loughner, C. P., Estes, M., Crawford, J. H., Pickering, K. E., Weinheimer, A. J., & Dickerson, R. R. (2016). Ozone production and its sensitivity to $NO_x$ and VOCs: Results from the DISCOVER-AQ field experiment, Houston 2013. *Atmospheric Chemistry and Physics*, *16*(22), 14463–14474. https://doi.org/10.5194/acp-16-14463-2016

McDuffie, E. E., Smith, S. J., O'Rourke, P., Tibrewal, K., Venkataraman, C., Marais, E. A., Zheng, B., Crippa, M., Brauer, M., & Martin, R. V. (2020). 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 System Science Data*, *12*(4), 3413–3442. https://doi.org/10.5194/essd-12-3413-2020

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., & Alvarez, R. A. (2020). Characterizing Elevated Urban Air Pollutant Spatial Patterns with Mobile Monitoring in Houston, Texas. *Environmental Science & Technology*, *54*(4), 2133–2142. https://doi.org/10.1021/acs.est.9b05523

Murray, L. T. (2016). Lightning NOxand Impacts on Air Quality. *Current Pollution Reports*, *2*(2), 115–133. https://doi.org/10.1007/s40726-016-0031-7

NASA. (2023). *EOSDIS Worldview*. https://worldview.earthdata.nasa.gov/

NASA/LARC/SD/ASDC. (2022). *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.

885

890

895

900

905

https://doi.org/10.5067/ASDC/SUBORBITAL/TRACERAQ/DATA001/GV/AircraftRemoteSensing/GCAS_1

910 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., & Weinheimer, A. J. (2018). Nitrogen dioxide and formaldehyde measurements from the GEOstationary Coastal and Air Pollution Events (GEO-CAPE) Airborne Simulator over Houston, Texas. *Atmospheric Measurement Techniques*, *11*(11), 5941–5964. https://doi.org/10.5194/amt-11-5941-
915 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., & Li, Q. (2001). Air mass factor formulation for spectroscopic measurements from satellites: Application to formaldehyde retrievals from the Global Ozone Monitoring Experiment. *Journal of Geophysical Research: Atmospheres*, *106*(D13), 14539–14550. https://doi.org/10.1029/2000JD900772

920 Ramboll. (2022a). Enhance MARINER Tool for Commercial Marine Emission Inventories Final Report. *Final Report*.

Ramboll. (2022b). *User's Guide COMPREHENSIVE AIR QUALITY MODEL WITH EXTENSIONS Version 7.20*. https://camx-wp.azurewebsites.net/Files/CAMxUsersGuide_v7.20.pdf

Riess, T. C. V. W., Boersma, K. F., Van Roy, W., de Laat, J., Dammers, E., & van Vliet, J. (2023). To new heights by
925 flying low: Comparison of aircraft vertical $NO_2$ profiles to model simulations and implications for TROPOMI $NO_2$ retrievals. *Atmospheric Measurement Techniques*, *16*(21), 5287–5304. https://doi.org/10.5194/amt-16-5287-2023

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

TCEQ. (2021). *SIP Revision: Regional Haze*. Texas Commission on Environmental Quality. https://www.tceq.texas.gov/airquality/sip/bart/haze_sip.html

United States Census Bureau. (2022, June 13). *Metropolitan and Micropolitan Statistical Areas Population Totals:*

935   *2020-2022*. Census.Gov. https://www.census.gov/data/tables/time-series/demo/popest/2020s-total-metro-

and-micro-statistical-areas.html

van Geffen, J., Eskes, H., Compernolle, S., Pinardi, G., Verhoelst, T., Lambert, J.-C., Sneep, M., ter Linden, M.,

Ludewig, A., Boersma, K. F., & Veefkind, J. P. (2022). Sentinel-5P TROPOMI $NO_2$ retrieval: Impact of

version v2.2 improvements and comparisons with OMI and ground-based data. *Atmospheric Measurement*

940   *Techniques*, *15*(7), 2037–2060. https://doi.org/10.5194/amt-15-2037-2022

van Geffen, J., Eskes, H. J., Compernolle, S., Pinardi, G., Verhoelst, T., & Lambert, J.-C. (2021). Sentinel-5P

TROPOMI NO2 retrieval: Impact of version v2.2 improvements and comparisons with OMI and ground-

based data. *Atmospheric Measurement Techniques*, *15*(7), 2037–2060. https://doi.org/10.5194/AMT-15-

2037-2022

945   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., … Zehner, C. (2021). Ground-based validation of the Copernicus

Sentinel-5P TROPOMI NO<sub>2</sub> measurements with the NDACC ZSL-DOAS, MAX-

DOAS and Pandonia global networks. *Atmospheric Measurement Techniques*, *14*(1), 481–510.

950   https://doi.org/10.5194/amt-14-481-2021

Wagner, T., Warnach, S., Beirle, S., Bobrowski, N., Jost, A., Puķīte, J., & Theys, N. (2023). Investigation of three-

dimensional radiative transfer effects for UV–Vis satellite and ground-based observations of volcanic

plumes. *Atmospheric Measurement Techniques*, *16*(6), 1609–1662. https://doi.org/10.5194/amt-16-1609-

2023

955   Weather Underground. (2023). *Houston, TX Weather History*.

https://www.wunderground.com/history/daily/us/tx/houston/KHOU

Wiedinmyer, C., Akagi, S. K., Yokelson, R. J., Emmons, L. K., Al-Saadi, J. A., Orlando, J. J., & Soja, A. J. (2011).

The Fire INventory from NCAR (FINN): A high resolution global model to estimate the emissions from

open burning. *Geoscientific Model Development*, *4*(3), 625–641. https://doi.org/10.5194/gmd-4-625-2011

960