## AQRP Project 20-020

# New Satellite Tools to Evaluate Emission Inventories: Is a 3-D Model Necessary?

## Final Report

Prepared for:

Elena McDonald-Buller
Project Manager
Texas Air Quality Research Program

Mark Muldoon
Project Liaison
Texas Commission on Environmental Quality

Prepared by:

Tracey Holloway, Monica Harkey, Eliot Kim and Jennifer Bratburd
University of Wisconsin – Madison

Jeremiah Johnson, Greg Yarwood
Ramboll, Novato, California

Daniel Goldberg
George Washington University

August 31, 2021

# Contents

**Tables**



**Figures**

## ACKNOWLEDGEMENT

The preparation of this report (Project No. 20-020) 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.

## LIST OF ACRONYMS AND ABBREVIATIONS

| | |
|---|---|
| AMF | Air Mass Factor |
| AMPD | Air Markets Program Data |
| AQRP | Air Quality Research Program |
| BCs | Boundary Conditions |
| BOEM | Bureau of Ocean Energy Management |
| CAMD | Clean Air Markets Division |
| CAMS | Continuous Air Monitoring Station |
| CAMx | Comprehensive Air quality Model with extensions |
| CEMs | Continuous Emissions Monitoring systems |
| CMAQ | Community Multiscale Air Quality Model |
| CO | Carbon Monoxide |
| CST | Central Standard Time |
| DFW | Dallas-Fort Worth Area |
| EGU | Electrical Generating Unit |
| EMG | Exponentially Modified Gaussian |
| EPA | Environmental Protection Agency |
| ESA | European Space Agency |
| FINN | Fire INventory from NCAR |
| GDAS | GFS Data Assimilation System |
| GFS | Global Forecasting System |
| HGB | Houston-Galveston-Brazoria Area |
| HPMS | Highway Performance Monitoring System |
| km | kilometer |
| LI | Local Increment |
| LSM | Land Surface Model |
| m | meter |
| mb | millibars |
| MB | Mean Bias |
| MDA8 | daily maximum 8-hour average |
| ME | Mean Error |
| MEGAN | Model of Emissions of Gases and Aerosols from Nature |
| MODIS | Moderate Resolution Imaging Spectroradiometer |
| MOVES | EPA MOtor Vehicle Emissions Simulator |
| NASA | National Aeronautics and Space Administration |
| NCAR | National Center for Atmospheric Research |
| NMB | Normalized Mean Bias |
| NME | Normalized Mean Error |
| NO | Nitric Oxide |
| NOAA | National Oceanic and Atmospheric Administration |
| NOx | Oxides of Nitrogen |
| NRT | Near Real-Time |
| NRTEEM | Near-Real Time Exceptional Event Modeling |

| | |
|---|---|
| OMI | Ozone Monitoring Instrument |
| PBL | Planetary Boundary Layer |
| PM | Particulate Matter |
| $PM_{2.5}$ | Particulate Matter < 2.5 ug $m^{-3}$ |
| ppb | parts per billion |
| Q-Q | Quantile-Quantile |
| RMS | Root Mean Square |
| RMSE | Root Mean Squared Error |
| SIP | State Implementation Plan (for the ozone NAAQS) |
| TexAER | Texas Air Emissions Repository |
| TexN | Texas NONROAD |
| TCEQ | Texas Commission on Environmental Quality |
| tpd | tons per day |
| TROPOMI | Tropospheric Monitoring Instrument |
| US | United States |
| VOC | Volatile Organic Compound |
| VCD | Vertical Column Density |
| VMT | Vehicle Miles Traveled |
| WHIPS | Wisconsin Horizontal Interpolation Program for Satellites |
| WRF | Weather Research and Forecast model |

## EXECUTIVE SUMMARY

This study developed best-practice recommendations for the utilization of satellite data for emissions evaluation. Because of their radiative properties, nitrogen dioxide ($NO_2$) is among of a small group of gas-phase air pollutants that may be reliably detected from space. Because $NO_2$ has a short atmospheric lifetime, satellite-based observations of atmospheric abundance may be used as an indicator of fuel combustion and associated emissions. Although the characterization of gas-phase emissions has emerged as one of the leading areas for air quality utilization of satellite data, multiple atmospheric processes affect the relationship between satellite-derived column abundance and near surface. We evaluated two different methods to compare satellite $NO_2$ with emission inventories developed by the Texas Commission on Environmental Quality (TCEQ).

Our proposal directly responded to two Priority Research Areas for the Air Quality Research Program (AQRP): the use of remote sensing for (1) point source and (2) county-level emissions. We developed methods to leverage remote sensing capabilities to improve emission inventories, without undermining the process-based nature of the inventories, essential for their use in air quality management.

Originally, our specific objectives included:

1) Comparison of satellite-derived $NO_2$ and $SO_2$ from TROPOMI for summer 2019 with model simulations from a WRF and CAMx modeling system developed for the TCEQ;

2) Simpler approaches to comparing $NO_X$ emissions and TROPOMI data that do not require a photochemical grid model, especially the Exponentially Modified Gaussian (EMG) approach. These simpler methods would have been extended to $SO_2$ as resources and data integrity allowed.

The work on $SO_2$ was not advanced due to the limited ability TROPOMI to detect power plant emissions, even with the more-robust $NO_2$ product. A limited analysis of satellite and model $SO_2$ results are presented in Appendix B.

Both objectives were accomplished for $NO_2$, as summarized by the following:

**Objective 1**: Compared to ground-based and satellite observations over the April – September 2019 study period, CAMx shows reasonable skill in simulating near-surface NOx and ozone, as well as in simulating column $NO_2$ amounts. CAMx column $NO_2$ shows improved agreement with TROPOMI when emissions of lightning $NO_X$ are included. We find CAMx has a high bias in column $NO_2$ relative to TROPOMI, particularly over power plants and urban areas; this may be partially explained by a low-bias in TROPOMI [e.g. *Griffin et al*., 2019; *Judd et al.,* 2020; *Verhoelst et al.* 2021; *Zhao et al*., 2020]. Compared to CAMx, TROPOMI underestimates $NO_2$ column amounts near power plants when emissions from power plants could be meaningfully

differentiated from background amounts. TROPOMI and CAMx have general agreement for urban counties, while a detailed analysis suggests some fine-scale differences in the core downtown area of Dallas and city perimeters of Austin and San Antonio.

Different approaches for comparing TROPOMI and CAMx were employed, including a "user-friendly" approach based on the publicly available Wisconsin Horizontal Interpolation Program for Satellites (WHIPS). WHIPS was updated for TROPOMI under this grant, allowing users to grid data to a custom model grid. Vertical processing may then be conducted with standard averaging kernels associated with TROPOMI. We compared this approach with a "research-grade" approach where the air mass factor is recalculated from CAMx.

**Objective 2:** We were surprised at the low fidelity of satellite retrievals of power plants – independent of processing and analysis approach. Even aside from satellite accuracy, multiple atmospheric processes decouple the relationship between emissions and column. To determine theoretical upper bound values, we calculated CAMx Correlations in daily modeled emissions vs. modeled column, which range r = 0.1 – 0.4 for the five study cities, and 0.0 – 0.5 for the five study power plants. As expected, actual TROPOMI vs. emissions correlations are worse for these theoretical upper bound values with r-values ~0 – 0.2. On longer time periods, correlations improve. Satellite data, on their own, offer qualitative information on emission patterns and trends, but are not suitable for quantitative emission inventory evaluation.

The EMG method is useful for determining aggregated emissions from an urban area. The urban area must be large enough to suit this method. We find that Dallas is well suited to EMG, but smaller cities including San Antonio, Austin, and College Station are not. The EMG analysis does not discern mobile vs. industrial vs. power plant emissions within an urban area. The EMG method did not work as well as expected for power plants, because TROPOMI has difficulty observing the full magnitude of Texas power plant plumes.

## Conclusions

- We sought to answer the question of whether a 3-dimensional model is necessary to evaluate emissions with satellite data. Indeed, we found that a 3-D model is the only tool for evaluating emission inventories on regional and daily scales, supporting comparison with ground-based and satellite data.
- Gaussian plume methods succeed for well-detected sources over longer timescales. This approach offers a lower-cost strategy to account for meteorology and chemistry.
- TROPOMI is useful for assessing overall spatial patterns in NOx emissions and in modeled $NO_2$, including differences among cities, while a detailed analysis suggests difficulties in capturing patterns at short timescales or within individual cities.
- We have updated the Wisconsin Horizontal Interpolation Program for Satellites (WHIPS) to process TROPOMI, and provided additional guidance and comparison on "user-friendly" and "research-grade" model-satellite comparison approaches.

- Although this work was motivated by the idea that satellite data provides a "check" on models and emission inventories, there were instances where disagreement suggests errors in the satellite's detection ability. For the time period of study, TROPOMI is surprisingly unable to capture power plant plumes, as simulated by CAMx based on CEMS.

## Recommendations for Future Work

- A substantial portion of the $NO_2$ column resides above the PBL and this $NO_2$ can strongly influence comparisons between satellite and modeled $NO_2$ columns. CAMx showed reasonable performance in simulating $NO_2$ in the free troposphere. Nevertheless, we recommend continued efforts to improve the CAMx simulation of $NO_2$ in the mid- to upper-troposphere by focusing on: (1) emission estimates for lightning and aircraft NOx (2) representation of NOx influx from the lower-stratosphere to the upper-troposphere (3) ventilation of NOy from the PBL to the mid-troposphere (4) photochemical production of $NO_2$ from NOy (especially organic nitrates) in the mid- to upper-troposphere.
- We found TROPOMI $NO_2$ was generally lower than CAMx column amounts, which is consistent with prior work comparing TROPOMI with ground-based tropospheric column measurements. In fact, at the end of this study (July 2021), the TROPOMI retrieval was updated by ESA to address the documented low bias. Given that the new TROPOMI $NO_2$ algorithm generates larger values, it is likely that this new algorithm will have improved agreement with CAMx column $NO_2$. It may be worthwhile to re-quantify resulting changes in TROPOMI relationships with emissions over urban areas, and especially over power plants once a new product is released for 2019.
- We found poor agreement between TROPOMI and CAMx $NO_2$ columns near major power plants in Texas. Additional study is needed with consideration of the both the TROPOMI data products (particularly the horizontal resolution of the a priori $NO_2$ column used in calculating the averaging kernel) and the model simulations (e.g. the vertical distribution of $NO_2$, photochemical lifetime of $NO_2$ in power plant plumes)
- Future work could address challenges with $SO_2$, but may require careful processing to discern a clear emissions signal from TROPOMI. The TEMPO satellite planned for launch in 2022 will provide higher resolution, hourly $SO_2$ data. These may be needed for meaningful $SO_2$ analysis for Texas power plants.

# 1.0 INTRODUCTION

This document provides the final report for the Texas Air Quality Research Program (AQRP) Project 20-020, "New Satellite Tools to Evaluate Emission Inventories: Is a 3-D Model Necessary?" The goal of Project 20-020 is to explore satellite-based methods of evaluating NOx emissions from power plants and cities.

The project Principal Investigator is Dr. Tracey Holloway (UW Madison) with co-Principal Investigator Mr. Jeremiah Johnson (Ramboll). The AQRP project manager is Dr. Elena McDonald-Buller at the University of Texas, Austin. The project liaison for the Texas Commission on Environmental Quality (TCEQ) is Mr. Mark Muldoon.

## 1.1 Background

Air quality management in Texas, like other states, is closely linked with accurate emission inventories. These inventories quantify chemical releases into the atmosphere from identified sources, and they are used as the basis of regulatory decision-making, atmospheric modelling, and assessment of trends. Although continuous emissions monitoring systems (CEMs) measure power plant emissions directly, most other emission sources are calculated based on ancillary data and assumptions, such as vehicle activity and land use. As such, there is some uncertainty in the accuracy of existing emission inventories, and potential for satellite remote sensing to evaluate and improve inventory development.

Because of their radiative properties, nitrogen dioxide ($NO_2$) and sulfur dioxide ($SO_2$) are among of a small group of gas-phase air pollutants that may be reliably detected from satellite instruments [*Richter and Burrows*, 2002; *Richter et al.*, 2005; *Martin*, 2008; *Duncan et al.*, 2014]. With relatively short atmospheric lifetimes, satellite-based observations of $NO_2$ have been found useful as an indicator of fossil fuel combustion and associated emissions. Spatial and temporal patterns in atmospheric $NO_2$ have been shown to reflect spatial variation, day-to-day and year-to-year changes in emissions of NOx [e.g. *Martin*, 2003; *Lamsal et al.*, 2011; *Mijling and Van Der A*, 2012; *Tong et al.*, 2016].

Over the past few years, major advances have occurred in the integration of satellite data with air quality planning [*Holloway et al.*, 2018]. Characterization of gas-phase emissions has emerged as one of the leading areas for air quality utilization of satellite data from the National Aeronautics and Space Administration (NASA), the National Oceanic and Atmospheric Administration (NOAA), and satellites launched by international space agencies.

Although satellite data bear relevance to emissions characterization, atmospheric processes including boundary layer mixing, dispersion by winds, and photochemistry affect the relationship between satellite-derived column abundance and near surface emissions [e.g. *Harkey, Holloway et al.*, 2015]. Additionally, the once-a-day snapshot provided by current-generation polar-orbiting satellites cannot account for the temporal variability in emissions [*Denier van der Gon et al.,* 2011], and processes that may affect the column with a delay from the surface (e.g. timing of NOx emissions, as discussed in *Fishman et al.,* [2008]).

Satellites measure the column abundance of $NO_2$, known as the vertical column density (VCD). This column is affected by surface and elevated emissions, as well as meteorology and chemistry in the atmosphere. Of particular note, vertical transport of pollutants in the column can affect the satellite's sensing of $NO_2$ abundance; horizontal transport of pollutants weakens the relationship between column $NO_2$ and proximate emissions; and atmospheric chemistry can create or destroy $NO_2$, also weakening the relationship between column $NO_2$ and proximate emissions.

A three-dimensional photochemical grid model (PGM) is the only method that can fully account for all of these processes, and there is a growing body of work of using satellite data for air quality model evaluation [*Canty et al.*, 2015; *Harkey et al.*, 2015; *Kemball-Cook et al.*, 2015; *Karambelas et al.*, 2018]. Although using a PGM represents the "gold standard" for satellite evaluation of emissions, this approach is costly and limiting, given the computational and personnel resources required for high-quality PGM simulations. To leverage satellite capabilities even without an atmospheric model, satellite observations of $NO_2$ have also been directly compared with emissions [e.g. *Jin and Holloway*, 2015; *Montgomery and Holloway*, 2018]. For assessment of specific point sources and urban areas, this approach typically requires meteorological correction factors [*de Foy et al.*, 2015; *Goldberg et al.*, 2019b].

All analysis in this study uses data from the TROPOspheric Monitoring Instrument (TROPOMI). TROPOMI is polar-orbiting with daily global coverage at a nadir resolution of 7 km × 3.5 km (5.5 km × 3.5 km starting August 2019), launched in 2017. The spatial resolution offered by TROPOMI is over 10x higher than any previous gas-monitoring satellite, with the Ozone Monitoring Instrument (OMI; nadir resolution of 13 km × 24 km) offering the next-highest capability.

Our analysis evaluates the TCEQ 2020 modeling emissions inventory in two ways, with: 1) a high-resolution (12 and 4 km) PGM simulation; and 2) meteorological corrections on emissions sources, without a model. Our study goals include the validation of the TCEQ 2020 modeling inventory, as well as recommendations and software to support future TCEQ utilization of satellite data for emission evaluation.

## 1.2 Overview of Approach

The overarching objective of this work is to support regional evaluation of emissions inventories with satellite data. However, emissions are not directly comparable with the column abundance detected by satellites, as outlined above.

To support the appropriate utilization of satellite data for emissions evaluation, we advanced and compared existing methods for emissions evaluation:

- Compare satellite data for $NO_2$ columns with model simulations from the high- resolution Comprehensive Air Quality Model with Extensions (CAMx), with input meteorology from the Weather Research and Forecasting (WRF) model, including seasonal and monthly mean difference plots across the 12 km and 4 km modeling domains and in-depth

difference analyses for select areas. These plots will include a “zoom in” difference plot over West Texas and New Mexico.

- Evaluate the utility of satellite data for NOx emissions inventory evaluation, without the use of a high-resolution model
- Evaluate how model-based emissions assessment compares to emissions assessment in the absence of model, finalizing recommendations, software, and algorithms
- Develop best-practice recommendations and software to support future TCEQ utilization of satellite data for emission evaluation

## 1.3 Overview of Report

In Section 2, we compare simulated $NO_2$ amounts with the high-resolution CAMx model with observations at our focus sites. In Section 3, we evaluate emissions from power plants and urban areas using the EMG method. In Section 4, we describe a direct comparison between satellite data and emissions from power plants and urban areas. Finally in Section 5, we present conclusions and recommendations for future work. Appendix A includes monthly plots for column $NO_2$, and “zoom in” plots over western Texas and New Mexico. Appendix B includes satellite and model data for column $SO_2$ to the limited degree it was included in this study.

# 2.0 WHAT CAN WE LEARN ABOUT EMISSIONS FROM SATELLITES + 3-D MODEL?

A three-dimensional photochemical grid model (PGM) has long been considered the industry standard for emission inventory evaluation. Only with a PGM can one account for key chemical and meteorological processes linking air emissions with ambient observations. The necessity of accounting for chemistry and transport is true for ground-based monitors and other *in situ* data, as well as satellite and other remotely sensed data sources. This gold-standard approach formed the backbone of our team's analysis, and provided a meaningful comparison for simpler methods discussed in future chapters.

We ran the WRF and CAMx models for the 2019 ozone season, March 15 – October 15. The study year 2019 was selected as the first full year for which TROPOMI data are available. Ramboll conducted the model simulations using an existing high-resolution WRF and CAMx modeling platform for 2019 developed for TCEQ (Near Real-Time Exceptional Event Model; NRTEEM) described in *Johnson et al.* [2019] and summarized below. The WRF model configuration is unchanged from the 2019 TCEQ NRTEEM study, while the CAMx model is updated to incorporate emissions changes. We provide a model performance evaluation against surface monitors at the end of this section.

## 2.1 Model Configuration

### 2.1.1 WRF Model

The WRF model is a mesoscale numerical weather prediction system designed to serve both operational forecasting and atmospheric research needs [*Skamarock*, 2004; 2006; *Skamarock et al.,* 2005; 2008; 2019]. We used version 4.0.3 of the Advanced Research WRF (ARW) WRF in this study.

We define the WRF 2014 36/12/4-km modeling domains as slightly larger than the CAMx 36/12/4-km domains (Figure 2-1) because WRF uses the outermost boundary cells to interpolate from the boundary conditions outside the domain to the model solution. The 36 km CAMx modeling domain (red) includes all of the continental US, Mexico and large areas of Central America and Canada. The 12 km (blue) and East Texas 4 km (green) domains are the TCEQ State Implementation Plan (SIP) modeling domains. Table 2-1 presents the vertical layer mapping table from 43 WRF layers to 28 CAMx layers. As with the modeling domains, this layer mapping is from the TCEQ SIP modeling.

We provide the WRF physics options and data sources in Table 2-2. This configuration is similar to that used for the TCEQ SIP modeling. We used 0.25° Global Forecasting System (GFS) data assimilation system (GDAS) analysis data[1] as initial conditions for the WRF meteorological model. This GDAS data is also used for boundary conditions and data assimilation.

[1] https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-data-assimilation-system-gdas

As part of the 2019 TCEQ NRTEEM project, we evaluated WRF meteorological performance at surface weather stations in Texas. This evaluation revealed reasonable performance for photochemical modeling applications and is documented in *Johnson et al.* [2019].


*Figure 2-1. CAMx 36/12/4 km modeling domains.*

*Table 2-1. Vertical layer mapping from 43 WRF layers to 28 CAMx layers.*

| WRF Layer No. | WRF Eta Level | WRF Layer Pressure (mb) | WRF Layer Top (m) | CAMx Layer No. | CAMx Layer Top (m) | CAMx Layer Thickness (m) |
|---|---|---|---|---|---|---|
| 43 | 0.0000 | 50.00 | 20576 | | | |
| 42 | 0.0100 | 59.63 | 19458 | | | |
| 41 | 0.0250 | 74.08 | 18082 | | | |
| 40 | 0.0450 | 93.35 | 16616 | | | |
| 39 | 0.0650 | 112.61 | 15427 | | | |
| 38 | 0.0900 | 136.69 | 14198 | 28 | 14198 | 2077 |
| 37 | 0.1150 | 160.77 | 13169 | | | |
| 36 | 0.1450 | 189.67 | 12120 | 27 | 12120 | 3586 |
| 35 | 0.1750 | 218.57 | 11221 | | | |
| 34 | 0.2100 | 252.28 | 10304 | | | |
| 33 | 0.2500 | 290.81 | 9372 | | | |
| 32 | 0.2900 | 329.34 | 8534 | 26 | 8534 | 2030 |
| 31 | 0.3300 | 367.87 | 7773 | | | |
| 30 | 0.3700 | 406.40 | 7073 | | | |
| 29 | 0.4050 | 440.12 | 6504 | 25 | 6504 | 1040 |
| 28 | 0.4400 | 473.83 | 5969 | | | |
| 27 | 0.4750 | 507.54 | 5464 | 24 | 5464 | 870 |
| 26 | 0.5100 | 541.26 | 4985 | | | |
| 25 | 0.5400 | 570.16 | 4594 | 23 | 4594 | 737 |
| 24 | 0.5700 | 599.05 | 4219 | | | |
| 23 | 0.6000 | 627.95 | 3857 | 22 | 3857 | 684 |
| 22 | 0.6300 | 656.85 | 3509 | | | |
| 21 | 0.6600 | 685.75 | 3174 | 21 | 3174 | 324 |
| 20 | 0.6900 | 714.64 | 2849 | 20 | 2849 | 314 |
| 19 | 0.7200 | 743.54 | 2535 | 19 | 2535 | 304 |
| 18 | 0.7500 | 772.44 | 2231 | 18 | 2231 | 247 |
| 17 | 0.7750 | 796.52 | 1984 | 17 | 1984 | 241 |
| 16 | 0.8000 | 820.60 | 1743 | 16 | 1743 | 235 |
| 15 | 0.8250 | 844.68 | 1508 | 15 | 1508 | 230 |
| 14 | 0.8500 | 868.76 | 1279 | 14 | 1279 | 135 |
| 13 | 0.8650 | 883.21 | 1144 | 13 | 1144 | 134 |
| 12 | 0.8800 | 897.66 | 1010 | 12 | 1010 | 132 |
| 11 | 0.8950 | 912.11 | 878 | 11 | 878 | 130 |
| 10 | 0.9100 | 926.56 | 748 | 10 | 748 | 86 |
| 9 | 0.9200 | 936.19 | 662 | 9 | 662 | 85 |
| 8 | 0.9300 | 945.82 | 577 | 8 | 577 | 84 |
| 7 | 0.9400 | 955.46 | 493 | 7 | 493 | 84 |
| 6 | 0.9500 | 965.09 | 409 | 6 | 409 | 83 |
| 5 | 0.9600 | 974.72 | 326 | 5 | 326 | 82 |
| 4 | 0.9700 | 984.35 | 243 | 4 | 243 | 82 |
| 3 | 0.9800 | 993.99 | 162 | 3 | 162 | 81 |
| 2 | 0.9900 | 1003.62 | 80 | 2 | 80 | 48 |
| 1 | 0.9960 | 1009.40 | 32 | 1 | 32 | 32 |
| 0 | 1.0000 | 1013.25 | 0 | 0 | 0 | |

*Table 2-2. WRF v4.0.3 physics options and data sources used in this study.*

| WRF Option | Option Selected |
|---|---|
| Analysis Data | 0.25° GDAS (IC/BCs and analysis nudging on the 36 and 12 km domains) |
| Microphysics | WRF Single-Moment 6-class (WSM6) |
| Longwave Radiation | Rapid Radiative Transfer Model (RRTMG) |
| Shortwave Radiation | RRTMG |
| Surface Layer Physics | MM5 similarity |
| LSM | Noah |
| PBL scheme | Yonsei University (YSU) |
| Cumulus scheme | Multi-Scale Kain-Fritsch (MSKF) |

### 2.1.2 CAMx Model

Table 2-3 provides the science options used for the CAMx simulation. We used CAMx v7.00 with the CB6r4 chemical mechanism. The model configuration is identical to that used for NRTEEM and is consistent with TCEQ SIP modeling.

*Table 2-3. Science options used for CAMx modeling.*

| Science Options | CAMx Configuration |
|---|---|
| Version | Version 7.00 |
| Time Zone | CST |
| Vertical Grid Mesh | 28 Layers with 32 m deep surface layer and 15 layers in the lowest 1.5 km |
| Horizontal Grids | 2-way nested grids with spacings of 36, 12 and 4 km |
| Meteorology | 2019 TCEQ NRTEEM WRF meteorology |
| Chemistry Mechanism | CB6r4 gas-phase mechanism |
| Chemistry Solver | EBI |
| Photolysis Rates | TUV version 4.8 with TOMS ozone column adjustment and in-line adjustment for clouds |
| Advection Scheme | Piecewise Parabolic Method (PPM) |
| Planetary Boundary Layer (PBL) mixing | K-theory with KV100 patch to enhance vertical mixing over urban areas within the lowest 100 m |
| In-line Ix Emissions On | Inorganic iodine (Ix) emissions from saltwater areas |
| Parallelization | MPI (12 threads) and OMP (2 threads) |

#### 2.1.2.1 Modeling Emissions Inventory

We updated the CAMx modeling emissions inventory from 2019 NRTEEM to incorporate anthropogenic emissions from the projected 2020 TCEQ modeling inventory (closest to 2019 available; the 2019 NRTEEM project had used a 2017 modeling inventory), and 2019 hourly CEMS data for electrical generating units (EGUs) that are a focus of our analysis.

The 2020 modeling emissions inventory did not include impacts of the social and economic response to COVID-19, which was advantageous for this application since we modeled the 2019 ozone season. Table 2-4 provides NOx and VOC emission summaries for an average summer weekday across the entire CAMx 4 km domain. As described later in this report, we decided not

to adjust mobile NOx emissions after finding good agreement between mobile NOx emissions and satellite $NO_2$ columns.

*Table 2-4. CAMx 4 km domain-wide summary of average summer weekday emissions (tons per day) by sector.*

| Emission Sector | NOx (tpd) | VOC (tpd) |
|---|---|---|
| On-road mobile | 412.8 | 192.0 |
| Off-road mobile* | 472.3 | 178.9 |
| Area | 279.2 | 1,372.4 |
| Oil and Gas | 428.1 | 1,283.1 |
| Non-EGU Point Sources | 531.4 | 443.7 |
| EGU Point Sources | 321.6 | 7.2 |
| MEGAN biogenic | 2,947.2 | 46,548.0 |
| Wildfires | 40.8 | 136.8 |

* Includes non-road, locomotives, airport and commercial marine vessel emissions

#### 2.1.2.2 Anthropogenic Emissions Inventory

TCEQ developed the 2020 modeling emissions inventory for the Dallas-Fort Worth (DFW) and Houston-Galveston-Brazoria (HGB) Reasonable Further Progress (RFP) State Implementation Plan (SIP) revision. We provide summaries of data sources used for the point sources in Table 2-5 and on-road mobile sources in Table 2-6. Table 2-7 provides data sources for the non-road mobile, area and oil and gas emissions sources. These tables are adapted from TCEQ's HGB and DFW RFP SIP revision and the only change that we implemented for this study is the update from 2018 to 2019 EPA Clean Air Markets Division (CAMD) Air Markets Program Data (AMPD) hourly data for EGU point sources.

In general, TCEQ estimated 2020 emissions by projecting forward in time from a base year inventory (see Table 2-5 and Table 2-7) and accounting for growth in activity as well as existing federal, state, and local emission controls. The 2020 modeling emissions benefit from the Federal Tier 3 Vehicle Emission and Fuel Standards Program, the Midlothian Cement Kiln caps and the EPA's final Cross-State Air Pollution Rule (CSAPR) update[2].

[2] https://www.tceq.texas.gov/assets/public/implementation/air/sip/dfw/dfw_ad_sip_2019/DFWAD_19078SIP_Appendix_B_pro.pdf

*Table 2-5. Data sources for point source emissions inventory. Adapted from Table 1-4 Appendix B of TCEQ's HGB and DFW RFP SIP Revision.*

| Region | Data Source |
|---|---|
| Texas | 2016 State of Texas Air Reporting System (STARS) for non-EGUs |
| Other States | 2011 EPA Modeling Platform interpolated for 2020 using 2017 and 2023, for non-EGUs |
| All States | 2019 EPA CAMD AMPD hourly data for EGUs* |
| Offshore | 2011 Bureau of Ocean Energy Management (BOEM) Gulf-wide Emission Inventory (GWEI) platforms of western Gulf of Mexico |
| Mexico | 2017 and 2023 Mexican EI from 2011 EPA Modeling Platform, interpolated for 2020 |
| Canada | 2023 Canadian EI from 2011 EPA Modeling Platform |

* TCEQ's 2020 modeling emission inventory used 2018 EPA CAMD AMPD hourly data for EGUs

*Table 2-6. Data sources for on-road mobile source emissions inventory. Adapted from Table 1-5 Appendix B of TCEQ's HGB and DFW RFP SIP Revision.*

| Region | Data Source |
|---|---|
| Texas | 2020 based on MOVES2014a[3] emission rates and Highway Performance Monitoring System (HPMS) for Vehicle Miles Traveled (VMT) activity estimates. |
| Outside Texas | 2020 based on MOVES2014 July default runs. |

*Table 2-7. Data sources for non-road mobile, area and oil and gas emissions inventory. Adapted from Table 1-6 of Appendix B of TCEQ's HGB and DFW RFP SIP Revision.*

| Region | Non-Road Mobile Sources | Area Sources | Oil and Gas Sources |
|---|---|---|---|
| Texas | 2020 run of Texas NONROAD (TexN) model, version 2 | 2017 Texas Air Emissions Repository (TexAER) projected to 2020 | Texas Railroad Commission data and equipment-specific emission rates. |
| Non-Texas U.S. | 2020 run MOVES2014b[4] | 2014 EPA National Emissions Inventory (NEI) | 2014 EPA NEI |
| Canada | 2011 EPA Modeling Platform | 2011 EPA Modeling Platform | 2011 EPA Modeling Platform |
| Mexico | 2011 EPA Modeling Platform | 2011 EPA Modeling Platform | 2011 EPA Modeling Platform |

#### *2.1.2.2.1 2019 EGU Emissions Based on CEMS*

We developed hourly-specific EGU emissions using hourly measurements from the EPA's CAMD. Most EGUs use CEMS to report emissions under the Clean Air Act, including emissions of sulfur dioxide ($SO_2$), NOx, and $CO_2$, along with other parameters such as heat input. The EPA's CAMD quality controls the reported raw hourly measurements which they provide on the AMPD

[3] https://19january2017snapshot.epa.gov/moves/moves2014a-latest-version-motor-vehicle-emission-simulator-moves_.html

[4] https://www.epa.gov/moves/information-running-moves2014b

website[5]. We downloaded hourly data from EPA's AMPD website for the continental US for March through October 2019. Stack parameters were based on EPA's 2017 NEI data. The 2017 NEI data with matching CEM facilities in Texas were then adjusted to their 2019 annual totals. Table 2-8 provides 2019 annual CEMS NOx emissions for the five power plants examined in detail in this study.

*Table 2-8. 2019 annual $NO_x$ emissions for the five power plants examined in this study.*

| Station | NOx (tons/yr) |
|---|---|
| Martin Lake Electrical Station | 9,489 |
| Limestone Electric Generation Station | 7,470 |
| Sam Seymour (Fayette) Power Project | 6,210 |
| Oklaunion Power Station | 5,214 |
| Forney Energy Center | 1,249 |

#### 2.1.2.3 Natural Emissions

We estimated biogenic emissions for 2019 from the Model of Emissions of Gases and Aerosols from Nature v3.1 developed by Ramboll in AQRP project 18-005; (MEGAN; [*Guenther et al.*, 2006]), and fire emissions from Fire INventory of NCAR (FINN) version 1. Finally, we developed lightning NOx emissions with the CAMx LNOx processor[6] using the 2019 WRF meteorological data.

## 2.2 Near-Surface Model Performance Evaluation

We evaluated CAMx NOx and ozone surface concentrations using data collected at TCEQ Continuous Air Monitoring Stations (CAMS). We evaluated performance by geographical subregions by grouping monitors as shown in Table 2-9, which also reports the number of NOx and ozone monitors in each region. College Station does not have ozone or NOx monitoring, so we chose Waco (one TCEQ CAMS with NOx monitoring; three with ozone monitoring) instead. Similarly, Shreveport does not have NOx monitoring, so we chose Tyler (three TCEQ CAMS with NOx and ozone monitoring) instead.

*Table 2-9. Number of TCEQ CAMS NOx and ozone monitors for the five study areas.*

| City | # of NOx monitors | # of Ozone monitors |
|---|---|---|
| Dallas/Fort Worth | 15 | 20 |
| San Antonio | 5 | 12 |
| Austin | 2 | 11 |
| College Station (Waco) | 0 (1) | 0 (3) |
| Shreveport (Tyler) | 0 (3) | 0 (3) |

[5] https://ampd.epa.gov/ampd/

[6] Available at https://www.camx.com/download/support-software/

$NO_x$ monitors deployed for routine monitoring, e.g., at TCEQ CAMS, have limitations for $NO_2$. These monitors measure NO and consequently $NO_2$ is chemically converted to NO for measurement. However, the converter also captures other compounds including peroxyacyl nitrate (PAN) and a portion of $HNO_3$. These NOx monitors have a detection limit of around 1 ppb but differentiation between NO and $NO_2$ is less accurate near the detection limit. Therefore, we compare both CAMx NOx (i.e., NO + $NO_2$) and NOy (i.e., NO + $NO_2$ + PAN compounds + $HNO_3$) to monitored NOx in Figure 2-2. These scatter plots show hourly midday (12-3 PM CST to correspond with TROPOMI overpass times) measurements and model pollutant concentrations at all of the TCEQ CAMs located in each region: Austin (top left), San Antonio (top right), Tyler (middle left), Waco (middle right) and Dallas-Fort Worth (bottom).

High observed NOx detected by ground monitors in urban areas (e.g. > 10 ppb) are not resolved at the CAMx model's 4 km horizontal grid resolution. For example, Dallas Hinton St (CAMS 0401) is located 0.9 km from a major freeway interchange and 200 m from a busy road (Mockingbird Lane). In contrast, Tyler Airport (CAMS 0082) is in a rural location removed from busy roads and the nearby airport is regional and not highly trafficked.

When compared with monitored NOx in less polluted areas (i.e. < 10 ppb), CAMx NOx tends to be lower than measured NOx whereas CAMx NOy tends to be higher than measured NOx. We therefore conclude that CAMx is consistent with the ambient NOx measurements within limitations of the monitoring equipment capabilities and siting.

We present similar scatter plots for maximum daily 8-hour average (MDA8) ozone in Figure 2-3. CAMx shows skill in identifying low and high ozone days, with $R^2$ values from 0.56 (Austin) to 0.61 (Tyler). CAMx displays a positive ozone bias across all five regions, with mean bias (MB) ranging from 4.8 ppb (Waco) to 10.1 ppb (San Antonio). *Emery et al.* [2017] defines the criteria standards for MDA8 ozone as ± 15% for normalized mean bias (NMB) and < 25% for normalized mean error (NME). Only Waco and Dallas-Fort Worth meet the criteria standard for NMB, while all regions except San Antonio meet the criteria standard for NME.

To determine how well CAMx estimates ozone production at the scale of a single metropolitan area, we calculate the local increment (LI) to MDA8 ozone as the difference between the downwind and upwind ozone. For metropolitan areas with an extensive monitoring network, such as Dallas-Fort Worth (DFW), the LI can be computed from observations and compared to the model simulation. The LI of ozone is sensitive to local ozone precursor emissions and the conduciveness of the atmosphere to ozone production on each day. Figure 2-4 displays a map of all DFW CAMS. We classify the monitors with green pushpins as potential background sites (meaning that when they are upwind of the urban area, they are indicative of background) and calculate the median MDA8 ozone concentration across these monitors for each day. Then we find the difference between this background value and the maximum MDA8 ozone concentration across all monitors in the same region. We refer to this difference as the MDA8 ozone LI. We perform the corresponding calculation on the model simulation.

In Figure 2-5, we present quantile-quantile (Q-Q) plots for Dallas MDA8 ozone LI for April 15-October 15, 2019. Q-Q plots are useful to understand whether the model is capable of simulating the full range of observed MDA8 ozone LI. Q-Q plots can be considered a less stringent measure of model performance because the model and observations are not paired in time or space. However, the Q-Q plots reveal more about how well the model simulates days when background ozone dominates (low LI) as well as days when local ozone production is high (high LI). The x-axis shows the observed LI and the y-axis shows CAMx LI. The plot shows very good agreement, with a slight underestimation that occurs throughout the full range of observed LI.

In summary, we find that CAMx shows good skill in identifying low and high ozone days for cities in Texas with $R^2$ values from 0.56 (Austin) to 0.61 (Tyler). CAMx has a positive ozone bias that is regional across eastern Texas and sometimes exceeds model performance criteria, although CAMx meets model error criteria for all cities evaluated. Using the extensive ozone monitoring network for DFW we determined that CAMx performs well in simulating the amount of ozone produced by emissions in the DFW area. We attempted to evaluate CAMx $NO_2$ performance using data from standard NOx instruments, focusing on the mid-day time period when satellite $NO_2$ measurements are made, but the evaluation was hampered by instrument detection limits and locations.

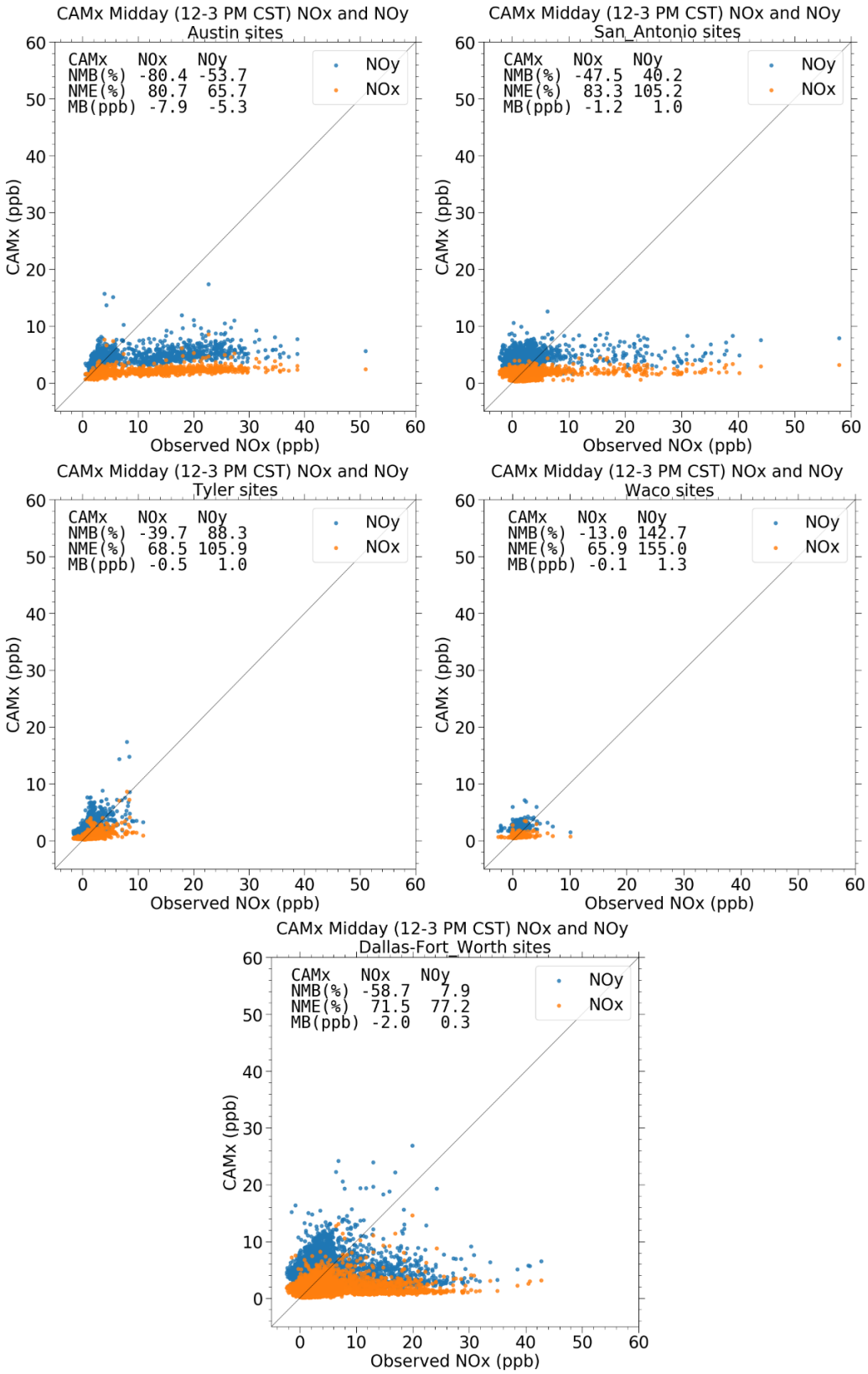



*Figure 2-2. Hourly CAMx midday (12-3 PM CST) NOx (orange) and NOy (blue) plotted against observed $NO_x$ across all TCEQ CAMS sites within Austin (top left), San Antonio (top right), Tyler (middle left), Waco (middle right) and Dallas-Fort Worth (bottom) for the March 15 – October 15, 2019 modeling period.*

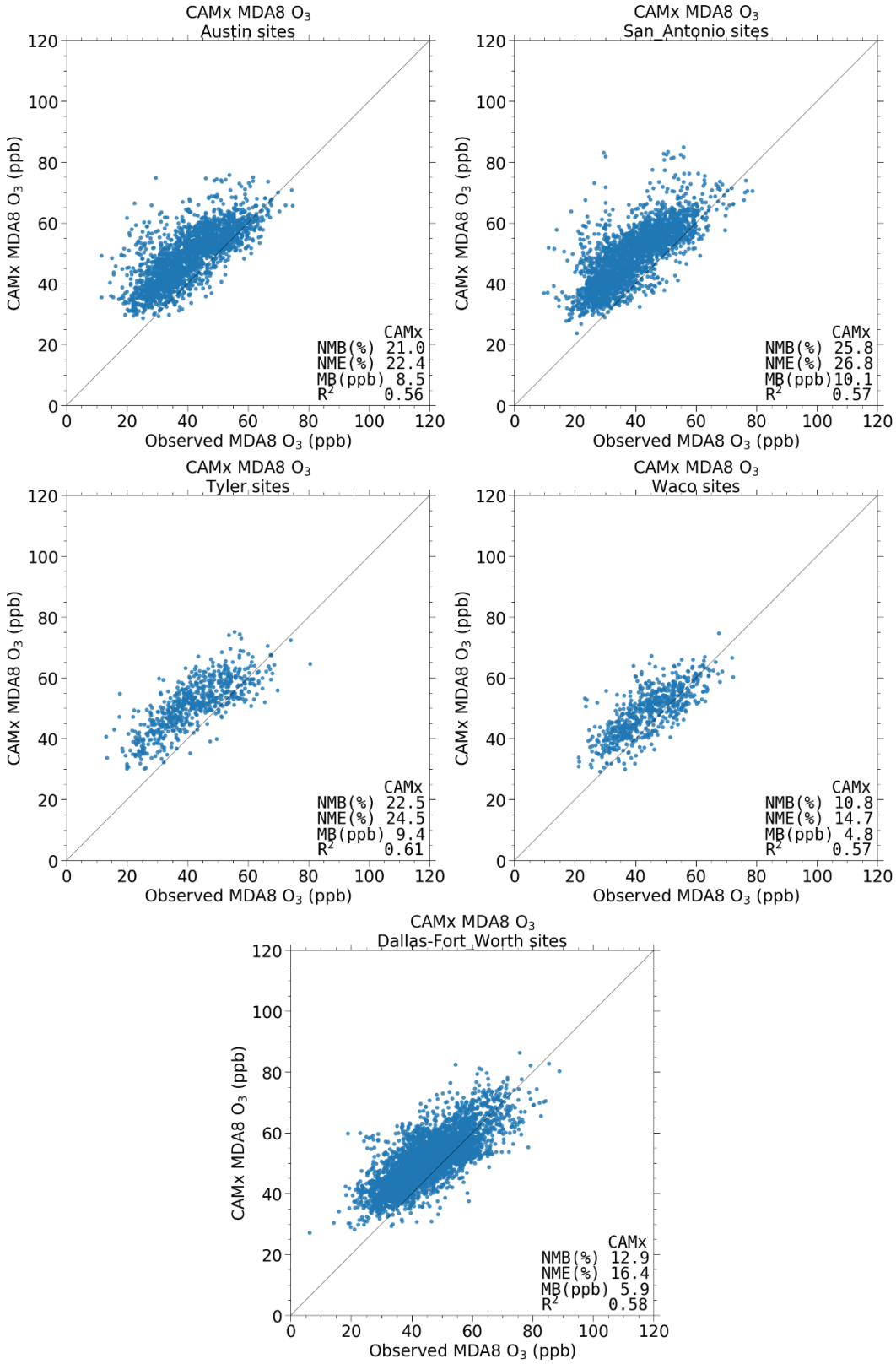



*Figure 2-3. CAMx and observed MDA8 ozone across all TCEQ CAMS sites within Austin (top left), San Antonio (top right), Tyler (middle left), Waco (middle right) and Dallas-Fort Worth (bottom) for the March 15 – October 15, 2019 modeling period.*

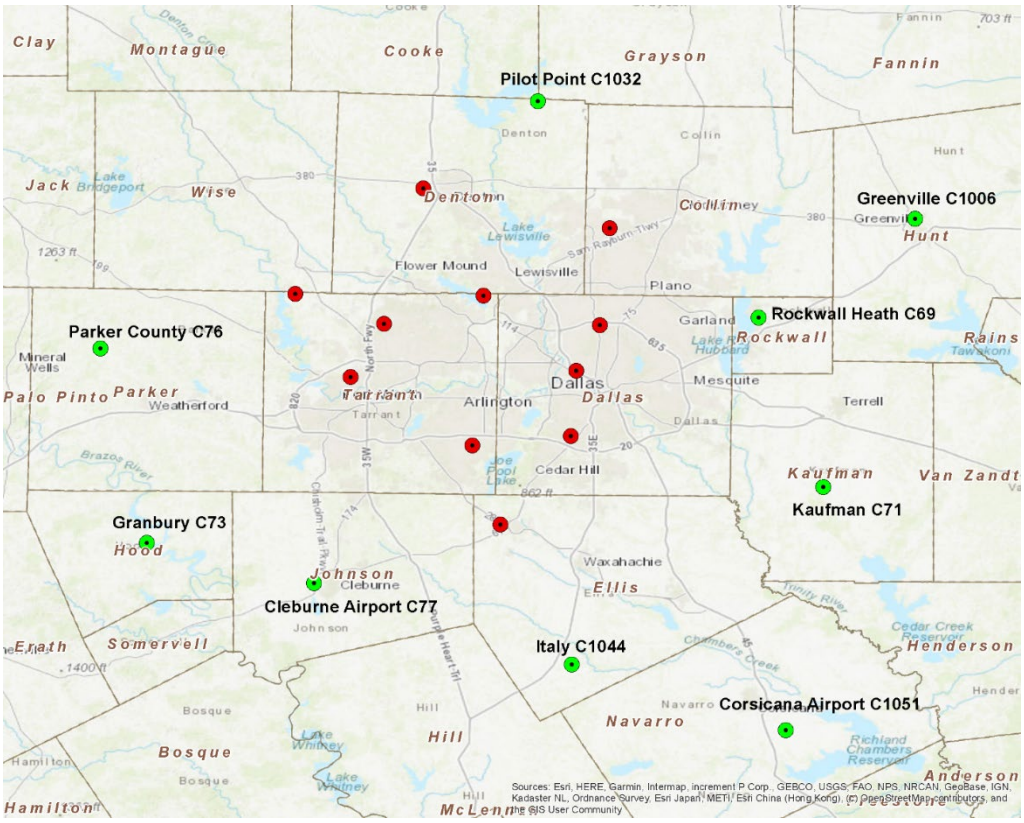



*Figure 2-4. Map of Dallas-Fort Worth CAMS monitoring locations. The nine potential background sites have green markers and are labelled.*

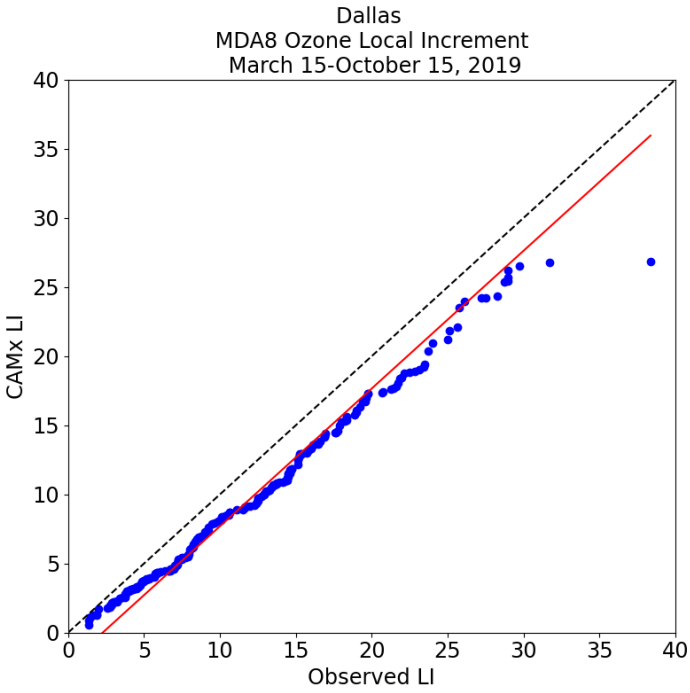



*Figure 2-5. Quantile-quantile plot for Dallas MDA8 ozone local increment for March 15-October 15, 2019.*

## 2.3 Model Performance Evaluation with TROPOMI

As a polar-orbiting satellite with an afternoon overpass, care must be taken in the interpretation of TROPOMI column retrievals as an indicator of near-surface emissions [*Streets et al.*, 2013; *Goldberg et al.*, 2019b; *Penn and Holloway*, 2020]. TROPOMI provides "snapshots" at the same time each day, except as limited by cloud cover, surface albedo, or instrument errors.

The TROPOMI data are provided as irregular quadrilaterals, a typical "level-2" data format. Each polygon value reflects the column total of $NO_2$. However, the instrument is more sensitive to some layers of the atmosphere, so this column total should be considered the weighted vertical integral of $NO_2$, with some layers weighted more than others.

For comparison with CAMx, we gridded TROPOMI data to the model grid using the publicly available Wisconsin Horizontal Interpolation Program for Satellites (WHIPS)[7]. WHIPS regrids the irregular polygons provided by "level-2" data and allocates the data on a specified grid using a variety of available gridding algorithms. In this sense, WHIPS allows users to create custom "level-3" data products for comparison with each other or model data on a common grid.

WHIPS was first updated to support TROPOMI data, then used to grid TROPOMI $NO_2$ data to both the 12 km and 4 km CAMx modeling domains. Though WHIPS calculates a common horizontal grid for the model and satellite observations, treatment of vertical sensitivity requires additional model processing to calculate a satellite equivalent VCD from vertical model layers. The most user-friendly approach involves an "averaging kernel" specific to a satellite data product may be as the weights in a weighted vertical integral (we refer to this as AK).

We also calculate the satellite air mass factor (AMF) using the model column and used to rescale the satellite data (we refer to this as the "research-grade" approach below, as significant resources are required to create the AMF-recalculated version of TROPOMI).

Finally, in the interest of comparing less computationally intensive methods, we also consider the results of summing layers without applying the averaging kernel, which gives more weight to near-surface concentrations, and thus larger $NO_2$ column amounts over urban areas (we refer to this as the "Summation" method). It should be noted that the summation approach does not capture the known sensitivity of the satellite to the upper troposphere, so these results should be considered illustrative only.

For the domain wide analyses, the AK and Summation methods are used; for city- and power-plant specific analyses, including EMG, the AMF and Summation methods are used. In both cases, the weighted approach to calculating the vertical column (AK or AMF) is considered the more scientifically defensible approach; Summation is given to highlight the sensitivity of results to the vertical processing approach.

[7] https://nelson.wisc.edu/sage/data-and-models/software.php

For the domain-wide comparisons, we screened TROPOMI $NO_2$ for a cloud fraction greater than 0.5, a solar zenith angle greater than or equal to 82 degrees, and a quality assurance flag value greater than 0.5, which is appropriate for work where the averaging kernel is applied to model columns [*Eskes et al.*, 2021]. Unless otherwise noted, we applied the gridded TROPOMI $NO_2$ averaging kernel to the model output following *Deeter* [2002] and as used in prior model evaluation work by *Harkey et al*. [2015] and *Harkey et al*. [2021]. We also present model column $NO_2$ calculated as a vertical integration without the averaging kernel for reference. For our comparisons with an updated AMF, TROPOMI data were screened for a quality assurance flag greater than 0.75.

### 2.3.1 Domain-Wide Evaluation with TROPOMI

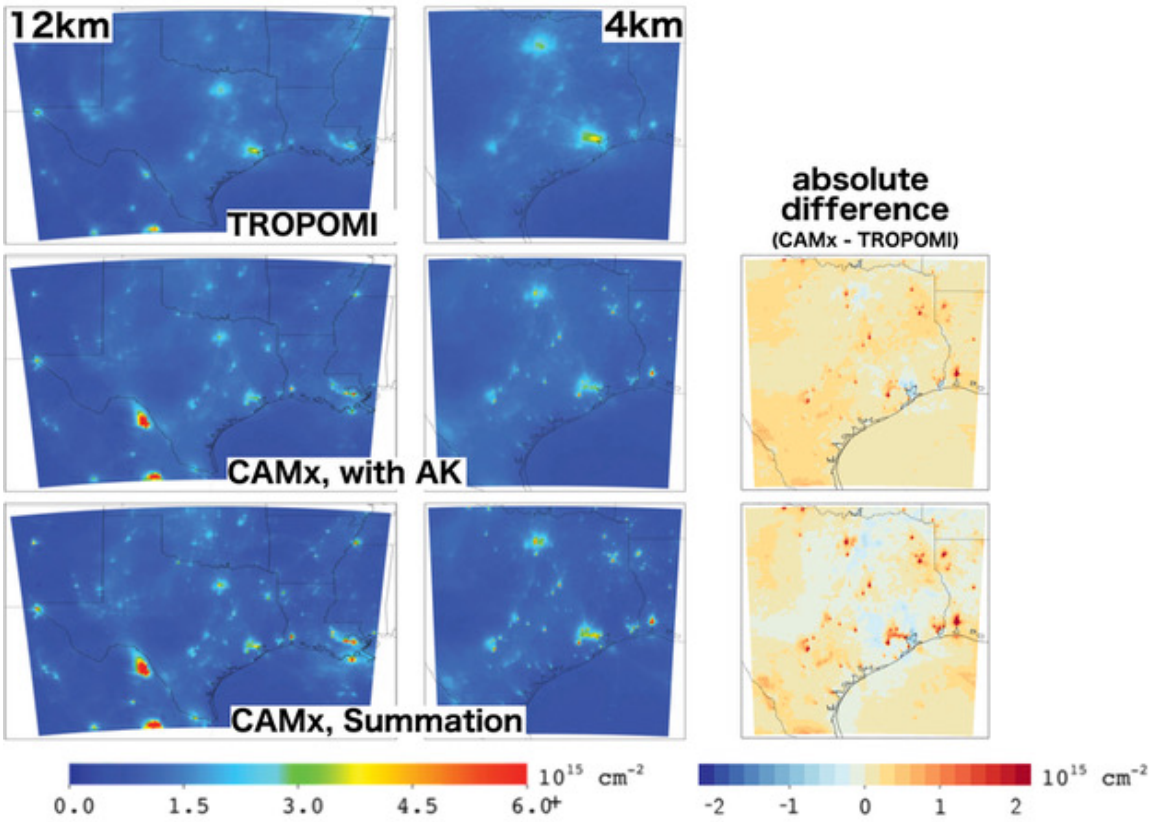



*Figure 2-6. April – September 2019 average $NO_2$ column amounts from TROPOMI as gridded with WHIPS (top) and CAMx with the averaging kernel applied (middle row), and CAMx without the averaging kernel (bottom row) for the 12 km domain (left column) and 4 km domain (middle column). Absolute difference between CAMx and TROPOMI shown in right column.*

Over both the 12 km and 4 km model domains, gridded TROPOMI captures elevated column $NO_2$ amounts over large urban areas such as Dallas – Fort Worth, Houston, and San Antonio, as well as the I-35 corridor linking San Antonio, Austin, and Dallas – Fort Worth (Figure 2-6, top row). Modeled column $NO_2$ is less diffuse than observed, with higher column amounts that are more spatially concentrated. For example, for the 4 km domain, observed column amounts over Dallas and Fort Worth are close to 3 x $10^{15}$ molecules/$cm^2$, while the largest column amounts modeled

for the area are found over Fort Worth, and are closer to 4 x $10^{15}$ molecules/$cm^2$; observed column amounts over the Martin Lake power plant in northeastern Texas are elevated relative to the background, but below 3 x $10^{15}$ molecules/$cm^2$, while simulated amounts are over 6 x $10^{15}$ molecules/$cm^2$ (Figure 2-6) for a single grid cell. The overall agreement between TROPOMI and CAMx column $NO_2$, with model overestimates relative to TROPOMI at power plants and most urban areas, is also evident on a monthly basis (Appendix A). However, we note that while a positive bias exists across both domains for the majority of the study period, in April there is an overall negative bias, and the largest positive bias occurs in August, which suggests seasonal differences between modeled and observed column $NO_2$.

The CAMx and TROPOMI results were compared using standard performance metrics, given below in Table 2-10. We find the model simulations compare favorably to TROPOMI for the April – September 2019 study period, with small overestimates of $NO_2$ column amounts compared to TROPOMI. As expected, the 4 km simulation performs somewhat better than the 12 km simulation, and application of the averaging kernel slightly improves agreement as well, such that the 4 km with AK exhibits the lowest RMSE (0.79) and highest r value (0.43) (Table 2-10).

*Table 2-10. TROPOMI and CAMx model mean column $NO_2$ for the period April – September 2019 and associated error metrics for both model domains, with and without the application of the averaging kernel. All units in $10^{15}$ molecules/$cm^2$ unless otherwise noted.*

| Domain | Model Mean | TROPOMI Mean | RMSE | Mean Error | Mean Fractional Error (%) | Mean Bias | Mean Fractional Bias (%) | r |
|---|---|---|---|---|---|---|---|---|
| 12km (with AK) | 1.19 | 1.02 | 0.91 | 0.45 | 47.94 | 0.16 | 24.55 | 0.36 |
| 12km (no AK) | 1.17 | 1.02 | 1.09 | 0.47 | 42.83 | 0.15 | 15.37 | 0.30 |
| 4km (with AK) | 1.29 | 1.12 | 0.79 | 0.49 | 43.50 | 0.17 | 21.65 | 0.43 |
| 4km (no AK) | 1.19 | 1.12 | 0.85 | 0.49 | 67.88 | 0.07 | 38.95 | 0.34 |

### 2.3.2 Effects of a priori profiles & Lightning NOx

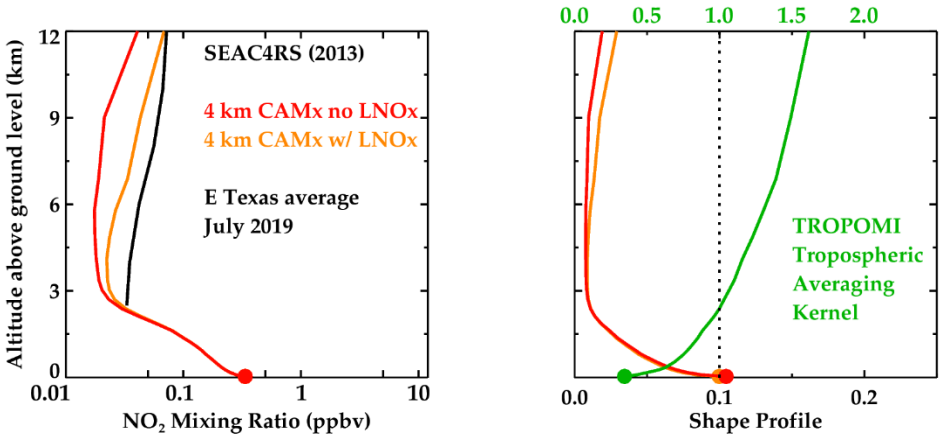



*Figure 2-7. (Left) $NO_2$ concentration vertical profiles between the surface and 12 km altitude from the CAMx regional model with and without lightning NOx emissions, and 2013 NASA SEAC4RS field campaign for the E Texas model domain. (Right) $NO_2$ shape profiles from the same three model simulations and the TROPOMI tropospheric averaging kernel*

We investigated the effects of lightning NOx emissions on the vertical distribution of $NO_2$ in CAMx by conducting a sensitivity simulation that excluded lightning NOx. In the right-hand panel of Figure 2-7, the vertical sensitivity of TROPOMI is denoted by the averaging kernel; the satellite is nearly three times more sensitive to $NO_2$ at an altitude of 12 km than at the surface. In the CAMx simulation with lightning NOx emissions, $NO_2$ between 6 km and 12 km nearly doubled and better matched observations from a SEAC4RS flight campaign in the Southeast United States during July 2013. Between 3 and 6 km it appears that there was still a model underestimate although the comparison is uncertain because model simulations and SEAC4RS measurements are for different years (2019 and 2013, respectively). An accurate simulation of mid- to upper-tropospheric $NO_2$ is an important consideration when evaluating NOx emissions using satellite data and a PGM.

The tropospheric $NO_2$ columns observed by TROPOMI vary considerably across Texas as seen in Figure 2-6 with larger column amounts observed over major cities. Combining the CAMx $NO_2$ height profiles and TROPOMI averaging kernel from Figure 2-7 explains how the contribution of ground-level NOx emissions to column $NO_2$ also varies across Texas, as shown in Figure 2-8. Over major cities such as Dallas (DFW) and Houston (HGB) where the column $NO_2$ reaches 3—4 x $10^{15}$ molecules/$cm^2$ about 80% of the total $NO_2$ column (as observed by TROPOMI) resides in the lowest 2 km of the atmosphere where surface NOx emissions are certainly the dominant influence on $NO_2$. In contrast, over the Gulf of Mexico (Gulf) and West Texas (WTx) where the column $NO_2$ is less than 1 x $10^{15}$ molecules/$cm^2$ only about 20% (or less) of the total $NO_2$ column resides in the lowest 2 km of the atmosphere with the majority of $NO_2$ observed by TROPOMI attributable to lightning NOx (LNOx) and other background sources. This explains why satellites such as TROPOMI can be much more successful at characterizing surface NOx emissions over major cities than rural areas. For intermediate size cities, such as San Antonio (SAT) and Austin (AUS), LNOx and other sources contribute about one third to half of the $NO_2$ column and thus uncertainties in characterizing the contributions of LNOx and other sources will be a limiting factor in using satellite data to characterize surface NOx emissions.

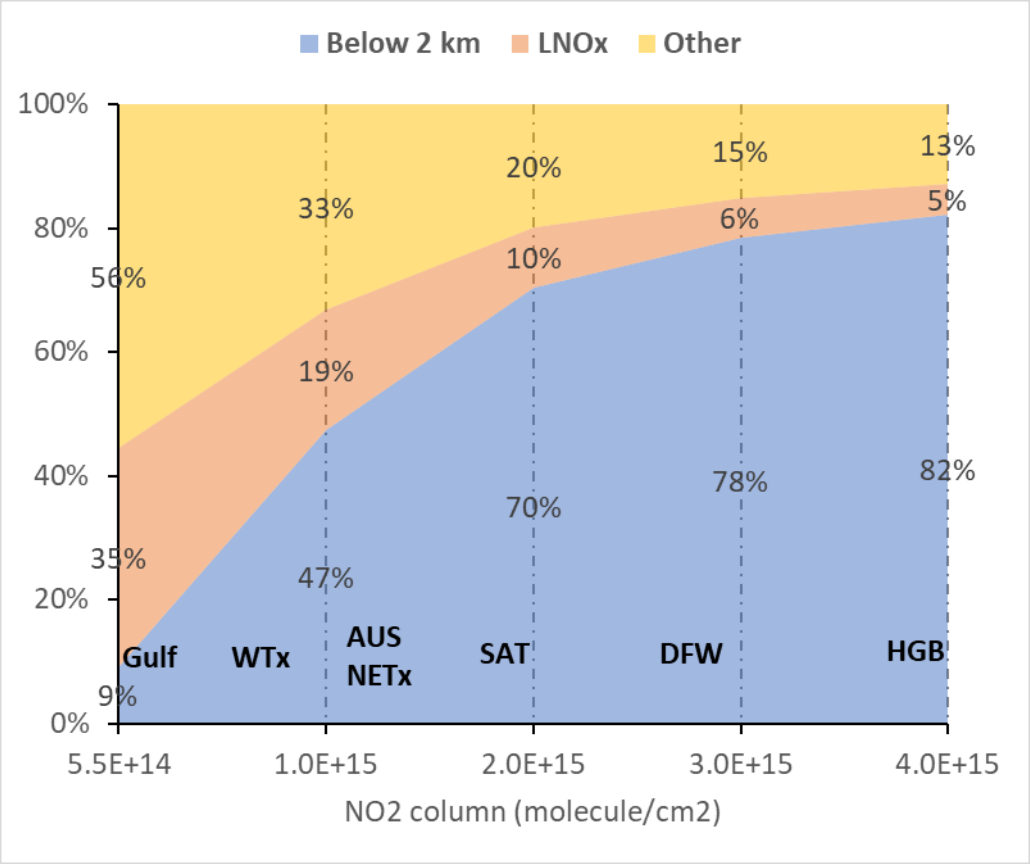



*Figure 2-8. Apportionment of the total $NO_2$ column observed by TROPOMI to the lowest 2 km of the troposphere, lightning NOx emissions (LNOx) and other background sources. Regions of Texas are marked according to their approximate total $NO_2$ column observed by TROPOMI.*

### 2.3.3 Evaluation with TROPOMI Over Urban Areas and Power Plants

To support our evaluation of the TCEQ emissions inventory, we examined CAMx column $NO_2$ over five urban areas and five power plants. Power plants were chosen because the accuracy of the CEMS data allowed us to evaluate methods of controlling for meteorology to link emissions and satellite-derived column densities. Vehicle emissions in urban areas were compared among cities with emission inventories developed in greater detail (link-based travel demand model) and lesser detail (MOVES defaults) to investigate whether inventory methodology influences the comparison. We considered the following study areas:

a) Five power plants in Texas with significant $NO_2$ (sources are in Table 2-11). These large and relatively isolated point sources have well-constrained emissions measured by CEMS. They were evaluated by comparing the PGM and TROPOMI results, as well as to test the methodology of *Goldberg et al.* [2019b] to capture emissions correctly without the use of an atmospheric model. We studied sources with differing spatial isolation, surrounding emissions, and fuel.

b) Five cities in Texas with significant on-road vehicle contribution to $NO_X$ emissions are selected for evaluation with both the model and no-model approach (see Table 2-11). These cities were selected to represent three different methodologies for developing mobile source emission inventories: i) emissions factors from the EPA MOtor Vehicle Emissions Simulator (MOVES) combined with activity from a link-based travel demand model (Dallas/Fort Worth and San Antonio); ii) emissions factors from MOVES combined with Texas-specific but non link-based activity (Austin and College Station); iii) default activity and emission factors from MOVES (Shreveport). We hypothesized that cities using more advanced mobile source activity data would show improved agreement with satellite $NO_2$, both in the CAMx framework and using the *Goldberg et al.* [2019a] adjustments. We did not select Houston for evaluation due to its complex meteorology (bay and gulf breezes that can recirculate emissions; *Banta et al.,* [2005]) and diverse emissions inventory with many important source sectors.

*Table 2-11. Power plants and cities proposed for evaluating the TCEQ emissions inventory. Power plant emission rates are from EPA Acid Rain Data for the 2018 ozone season*

| **Power Plant** | **Comments** |
|---|---|
| Martin Lake | Largest 2018 NOx point source in Texas; lignite and coal; rural high biogenic area; major sources ~15 mile distant; (NOx 7241 ton/y in 2018) |
| Limestone | High NOx; lignite and coal; rural high biogenic area; gas production nearby; (NOx 5676 ton/y in 2018) |
| Oklaunion | High NOx; coal; rural moderate/low biogenic area; (NOx 4495 ton/y in 2018). Decommissioned in 2020 but active in 2019. |
| Sam Seymour (Fayette) | High NOx; coal; rural high biogenic area; between Austin and Houston; aka Fayette; (NOx 4730 ton/y in 2018) |
| Forney Energy Center | Moderate NOx; gas; rural outskirts of Dallas; (NOx 782 ton/y in 2018) |
| **City** | |
| Dallas/Fort Worth | Combined population 2.2 million; Mobile source emissions link-based with city-specific MOVES inputs |
| San Antonio | Population 1.5 million; Mobile source emissions link-based with city-specific MOVES inputs |
| Austin | Population 1.0 million; Mobile source emissions non-link-based with city-specific MOVES inputs |
| College Station | Population 0.12 million; Mobile source emissions non-link-based with city-specific MOVES inputs |
| Shreveport | Population 0.19 million; Mobile source emissions non-link-based with default MOVES inputs |

### 2.3.3.1 Comparison of column $NO_2$ over cities

#### *2.3.3.1.1 Comparison with model columns calculated with the TROPOMI averaging kernel*

For our five study cities, model column $NO_2$ amounts are greater than those observed by TROPOMI, with model mean biases ranging from 0.01 - 0.41 x $10^{15}$ molecules/$cm^2$ for the study period (Table 2-12). The high model bias may be partially related to a TROPOMI algorithm low bias. We found the greatest correlation between TROPOMI and modeled column $NO_2$ over the Dallas – Fort Worth area (r = 0.40, Table 2-12). The comparisons in Figure 2-9 also show that when TROPOMI $NO_2$ column amounts are higher (> 5 x $10^{15}$ molecules/$cm^2$), modeled amounts tend to be low, most notably for Austin, College Station, and Shreveport, and to a lesser extent, Dallas – Fort Worth. This suggests that while the model has a small overall high bias relative to TROPOMI, there are challenges in simulating days with elevated emissions, or in capturing mixing and transport to the vertical levels where TROPOMI is most sensitive.

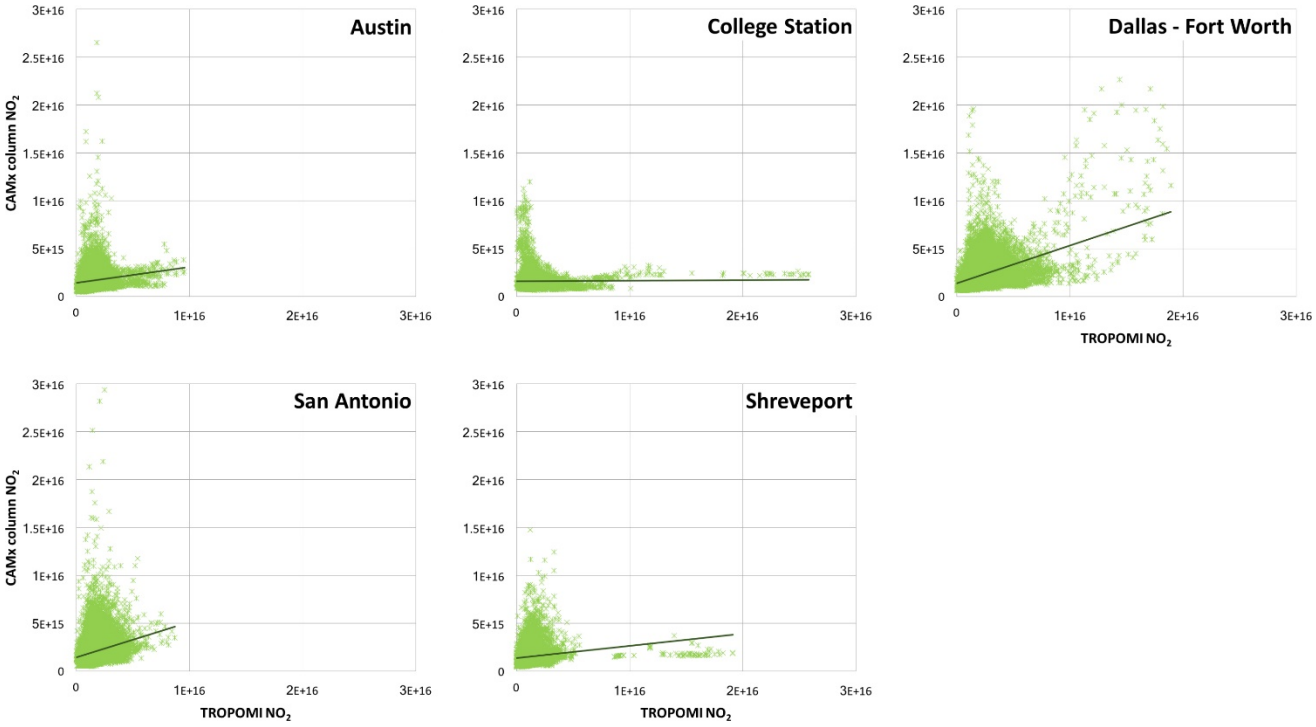



*Figure 2-9. TROPOMI and CAMx column $NO_2$ for the period April – September 2019 for five cities, as defined by the counties including each city: Austin (Travis County), College Station (Brazos County), Dallas and Fort Worth (Dallas and Tarrant Counties, respectively), San Antonio (Bexar County), and Shreveport (Caddo County).*

*Table 2-12. TROPOMI and CAMx model mean column $NO_2$ for the period April – September 2019 and associated error metrics for five cities. All values except r have units of $10^{15}$ molecules/$cm^2$ unless otherwise noted.*

| City (county) | Model Mean | TROPOMI mean | RMSE | Mean Error | Mean Fractional Error (%) | Mean Bias | Mean Fractional Bias (%) | r |
|---|---|---|---|---|---|---|---|---|
| **Austin (Travis)** | 1.65 | 1.47 | 1.03 | 0.66 | 41.47 | 0.19 | 14.60 | 0.16 |
| **College Station (Brazos)** | 1.59 | 1.49 | 1.56 | 0.78 | 41.61 | 0.10 | 12.16 | 0.01 |
| **Dallas-Fort Worth (Dallas and Tarrant)** | 2.26 | 2.18 | 1.37 | 0.90 | 31.16 | 0.01 | -5.42 | 0.40 |
| **San Antonio (Bexar)** | 2.02 | 1.61 | 1.22 | 0.80 | 46.12 | 0.41 | 25.48 | 0.26 |
| **Shreveport (Caddo)** | 1.53 | 1.27 | 1.16 | 0.68 | 48.19 | 0.25 | 24.07 | 0.16 |

#### *2.3.3.1.2 Comparison with model columns calculated with "research-grade" versions of TROPOMI*

We developed three "research-grade" TROPOMI $NO_2$ products to aid in the comparison. In the first "research-grade" TROPOMI $NO_2$ product ("TROPOMI $NO_2$ bias-corrected (b-c)"), the TROPOMI values are multiplied by a factor 1.25 to remove low bias in the algorithm [*Judd et al.,* 2020; *Verhoelst et al.,* 2021]. In a second "research-grade" TROPOMI $NO_2$ product ("TROPOMI $NO_2$ b-c CAMx AMFs"), the air mass factor and tropospheric vertical column are re-calculated using the PGM and applied to the first "research-grade" bias-corrected product. In a third "research-grade" TROPOMI $NO_2$ product ("TROPOMI $NO_2$ b-c CAMx AMFs + Downscaled"), we use the second "research-grade" product and then use spatial model variability to regrid the satellite measurements to the 4 km PGM grid [*Kim et al.,* 2016]. The most appropriate comparison for $NO_2$ is between the PGM and "TROPOMI $NO_2$ CAMx b-c AMFs + Downscaled" product.

When comparing CAMx to these research-grade TROPOMI products over a 6-month average (April – Sept), there is excellent model agreement (Figure 2-10). On an urban scale, CAMx has more $NO_2$ near the DFW airport and less $NO_2$ in downtown Dallas and east of the city. In San Antonio and Austin, both cities had good agreement in the downtown areas, but TROPOMI underestimates the $NO_2$ near large point sources on the perimeter of the city. The TROPOMI

signals in College Station and Shreveport were too close to the background concentration to perform a meaningful analysis.

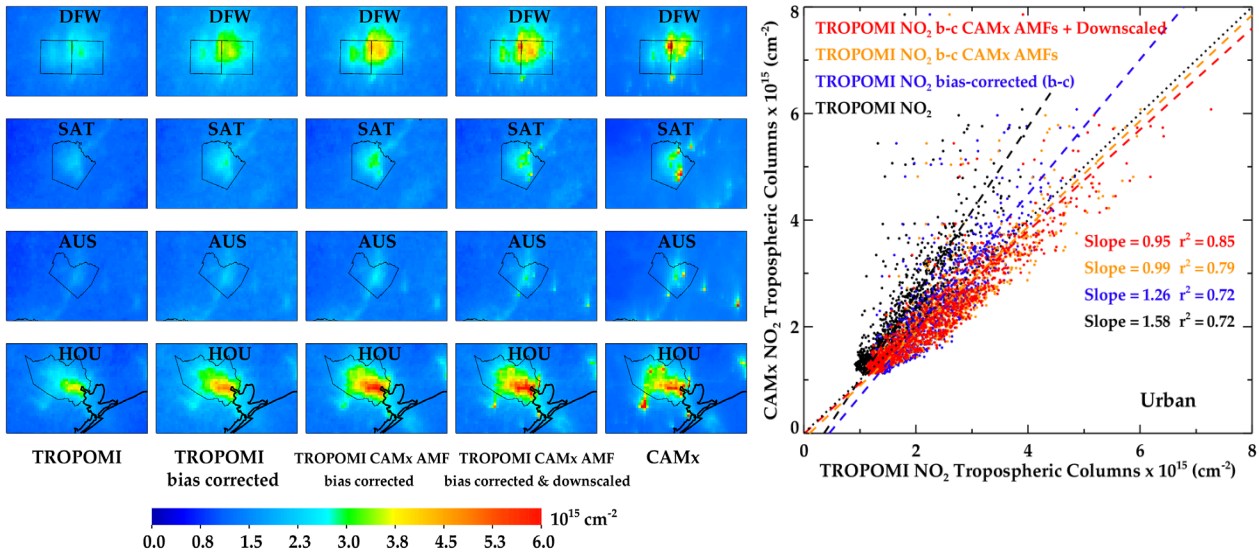



*Figure 2-10. $NO_2$ tropospheric vertical column amounts averaged across April – September 2019 from TROPOMI, TROPOMI bias-corrected, TROPOMI b-c CAMx AMFs, TROPOMI b-c CAMx AMFs + Downscaled, and CAMx for the largest four metropolitan areas (Dallas, San Antonio, Austin and Houston). (Right) Scatterplot showing slope and correlation of various TROPOMI configurations and CAMx.*

### 2.3.3.2 Comparison of column $NO_2$ at power plants

#### *2.3.3.2.1 Comparison with model columns calculated with the TROPOMI averaging kernel*

We find simulated $NO_2$ column amounts are greater than those observed over the five power plant locations of focus, as observed amounts are always < 10 x $10^{15}$ molecules/$cm^2$ while modeled column $NO_2$ ranges from three times greater (Forney, Oklaunion; Figure 2-11) to ten times greater (Martin Lake; Figure 2-11). As shown in Figure 3-1, TROPOMI observations do not capture the same finer-scale features as the model. This difference is likely related to the relatively small spatial footprint of an individual power plant $NO_2$ plume relative to the native TROPOMI resolution (7 km x 3.5 km) which may be further degraded by re-gridding to 4 km x 4 km resolution. However, even when TROPOMI is unable to spatially-resolve a relatively narrow $NO_2$ plume, the satellite measurement should be able to quantify the amount of $NO_2$ in the plume because each satellite pixel inherently represents the average $NO_2$ over the pixel area (i.e., mass of $NO_2$ should be conserved in the satellite detection, as well as in the subsequent data processing). The comparisons presented in Figure 2-11 suggest that, according to TROPOMI, the emission inventory for NOx from Martin Lake and Limestone is over-reported, but this is difficult to reconcile with emissions from these power plants being measured in real-time by CEMS installed in their stacks.

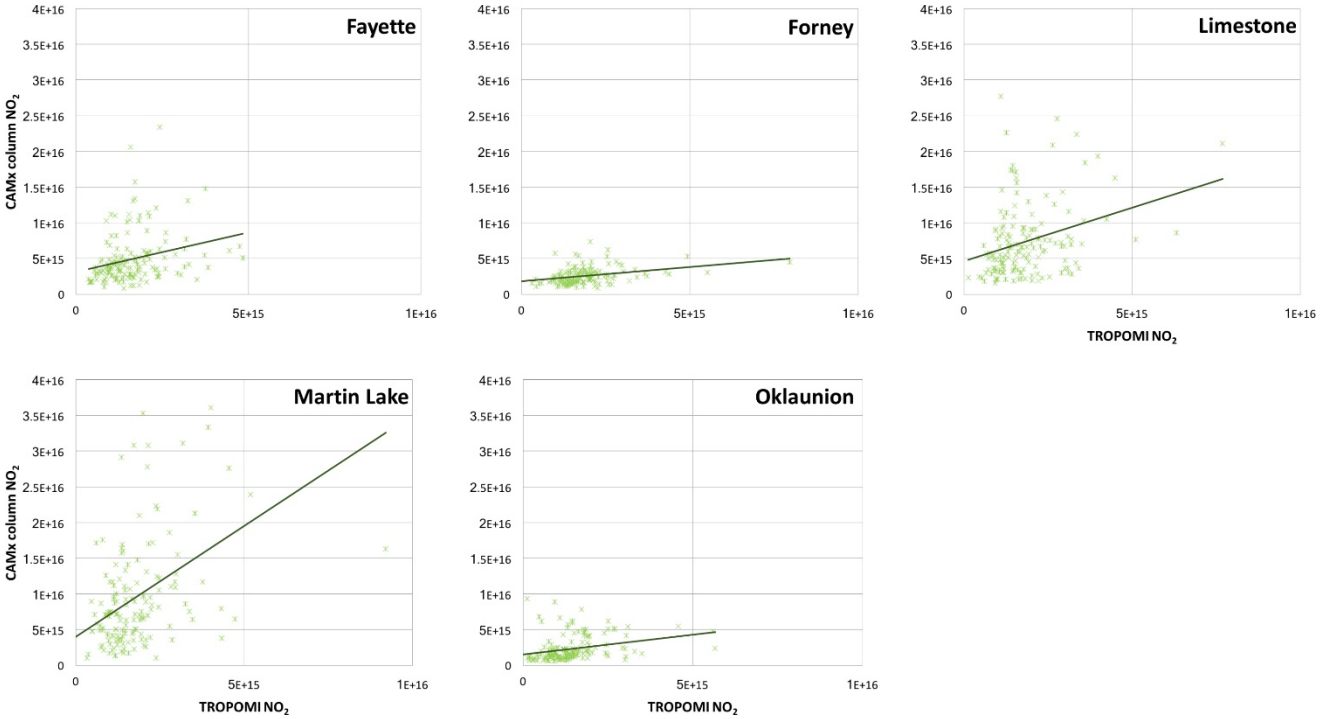



*Figure 2-11. TROPOMI and CAMx column $NO_2$ for the period April – September 2019 for five power plants.*

There are moderate positive correlations between observed and modeled column $NO_2$ over the five power plants of interest (Table 2-13). These correlations (r = 0.26 - 0.43) are overall greater than those over cities (r = 0.01 - 0.40; Table 2-12).

*Table 2-13. TROPOMI and CAMx model mean column $NO_2$ for the period April – September 2019 and associated error metrics for five power plant locations. All units in $10^{15}$ molecules/$cm^2$ unless otherwise noted.*

| Power plant | Model Mean | TROPOMI mean | RMSE | Mean Error | Mean Fractional Error (%) | Mean Bias | Mean Fractional Bias (%) | r |
|---|---|---|---|---|---|---|---|---|
| Fayette | 4.90 | 1.62 | 4.76 | 3.35 | 92.18 | 3.30 | 89.22 | 0.26 |
| Forney | 2.56 | 1.87 | 1.37 | 0.99 | 43.53 | 0.71 | 31.66 | 0.34 |
| Limestone | 7.40 | 1.85 | 7.39 | 5.57 | 107.83 | 5.56 | 107.28 | 0.31 |
| Martin Lake | 9.58 | 1.79 | 11.08 | 7.93 | 122.35 | 7.90 | 121.13 | 0.43 |
| Oklaunion | 2.28 | 1.38 | 1.93 | 1.21 | -48.84 | 0.91 | -65.96 | 0.29 |

#### *2.3.3.2.2 Comparison with model columns calculated with updated AMF*

The comparison between TROPOMI and CAMx at the locations of large power plants had worse agreement; CAMx always yielded larger amounts than observed by TROPOMI, usually by a factor of 3 – 5 (Figure 2-12). We investigated the NOx plumes from Martin Lake, Limestone, and Sam Seymor; the plumes from Oklaunion and Forney could not be reliably differentiated from the background $NO_2$ concentration. We think the combination of relatively smaller values (single power plant: ~10 Mg/hr vs. Dallas – Fort Worth: ~75 Mg/hr), narrow plume width, especially

short effective $NO_2$ lifetime (high sun angle and breezy) are leading to the poor disagreement at the locations of power plants. This type of disagreement was not seen for the Colstrip Power Plant in Montana [Goldberg et al., 2019b], which has larger NOx emissions in an environment that supports a longer effective $NO_2$ lifetime. While the correlation between CAMx and TROPOMI was smaller at the locations of power plants as compared to urban areas, the correlation when all three power plants were combined was $r^2$=0.49, and $r^2$=0.94 when the spatial variability of the model was imposed on the satellite observations. The moderate correlation without downscaling ($r^2$=0.49) suggests that TROPOMI can discern power plant plumes from background concentrations. The significant increase in correlation when model spatial variability was imposed ($r^2$=0.94) was expected and suggests that the native TROPOMI resolution is unable to capture the fine-scale features of power plant plumes, and that using a high-resolution PGM to further downscale the satellite measurements can be advantageous.

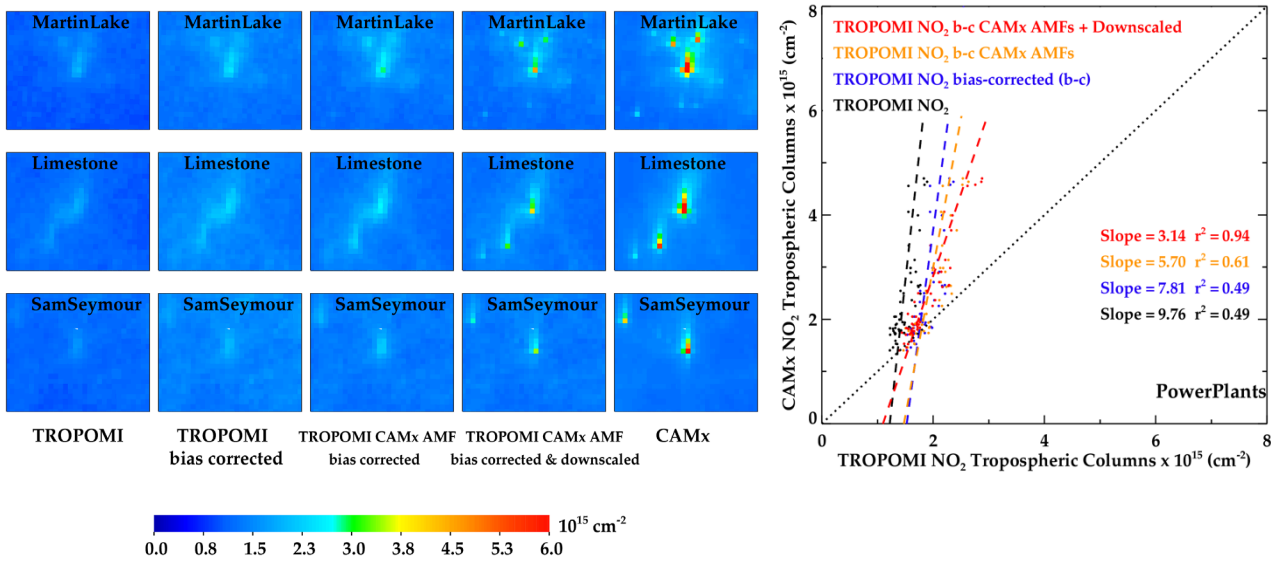



*Figure 2-12. $NO_2$ tropospheric vertical column amounts averaged across April – September 2019 from TROPOMI, TROPOMI bias-corrected, TROPOMI b-c CAMx AMFs, TROPOMI b-c CAMx AMFs + Downscaled, and CAMx for the largest three power plants (Martin Lake, Limestone and Sam Seymour (Fayette)). (Right) Scatterplot showing slope and correlation of various TROPOMI configurations and CAMx.*

## 3.0 WHAT CAN WE LEARN ABOUT EMISSIONS FROM THE EMG METHOD?

There are a number of methods that can be used to directly compare emissions and satellite data without a PGM. These range from direct comparison of temporal and spatial emissions patterns [e.g. *Montgomery and Holloway*, 2018], to more sophisticated methods that approximate the effects of meteorology and chemistry, even without the use of a three-dimensional model.

As a first step in our analysis, daily TROPOMI $NO_2$ and CAMx column $NO_2$ data—with the TROPOMI averaging kernel applied—were compared with $NO_X$ emissions, to assess agreement in the absence of meteorological corrections.

For our second step, we considered the statistical fit of the effective $NO_2$ plume decay over time. This approach, originally proposed by *Beirle et al.* [2011], involves the statistical fitting of satellite-observed $NO_2$ plumes to a Gaussian function that exponentially decays over time (exponentially modified Gaussian; EMG). Daily $NO_2$ plumes from TROPOMI were mapped onto an x-y grid and then rotated based on the daily wind-direction. As a result, all plumes were superimposed, increasing the signal-to-noise ratio and generating a more robust fit [*Valin et al.*, 2013; *Lu et al.*, 2015; *Goldberg et al.*, 2019a, 2019b, 2019c]. NOx emissions associated with plumes were calculated using the following equation:

$$NOx\ emissions = 1.33\ (\alpha\ /\ \tau_{effective}),\ \ where\ \ \tau_{effective} = x_o\ /w \qquad (1)$$

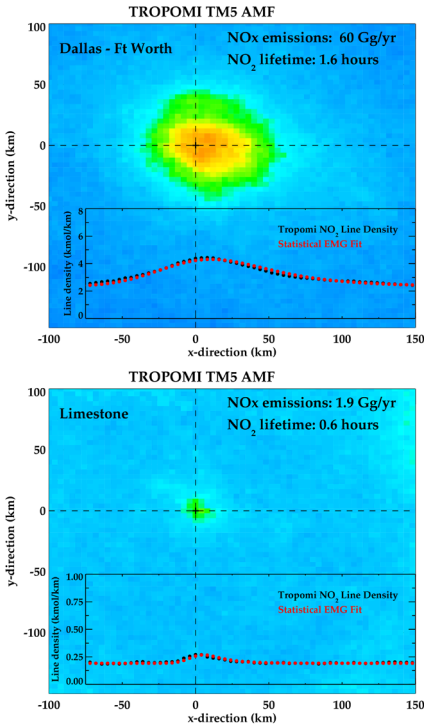
In this equation, $\tau_{effective}$ represents the mean effective $NO_2$ lifetime and is a combination of the photochemical lifetime and dispersion lifetime [*de Foy et al.,* 2014]; $x_o$ represents the fitted decay distance; $w$ represents wind speed; and $\alpha$ represents total burden obtained by the exponentially modified Gaussian fit. $NO_2$ is converted to $NO_X$ by multiplying by a factor of 1.33 which is typical of the mean column-averaged $NOx/NO_2$ ratio in an urban area during the mid-afternoon. This ratio varies in space and time, as discussed in *Goldberg et al.* [2019b]. The effective $NO_2$ ranged from 0.5 – 2.5 hours for this analysis; typically, the effective $NO_2$ lifetime is 2 – 5 hours at other US cities and power plants [*Goldberg et al.,* 2019b].

The wind speed and direction needed for these calculations was taken from the Ramboll WRF simulations at 12 and 4 km, and compared with more widely available re-analysis data (such as ERA-5). Mean near-surface wind speed over all days with valid satellite data was included.

We applied the EMG method for all five cities and five power plants. The statistical fit had realistic solutions (appropriate fit and $NO_2$ lifetime) for only one metropolitan area, Dallas Ft Worth, and one power plant, Limestone. The reason for the unrealistic fit for the other eight locations is probably due to a combination of especially short effective $NO_2$ lifetimes in Texas during summer, and lack of instrument sensitivity to the relatively small $NO_2$ plumes from the medium-sized cities on a day-to-day basis.

Results for Dallas Fort Worth and the Limestone Power plant are shown in Figure 3-1. On the left, is the standard TROPOMI $NO_2$ product, in the center is re-processed TROPOMI $NO_2$ product and on the right is the CAMx model simulation. In Dallas Fort Worth, the satellite-based $NO_X$ emissions were derived to be 71 Gg/yr in the early afternoon. The same method applied to the model simulation model yielded an estimate of 73 Gg/yr, which means the satellite and emissions inventory have excellent agreement. The value of 73 Gg/yr could have also been calculated using an effective radius of 70 km applied to the model-ready emissions, which represents three times the sigma (23 km) of the Gaussian plume. The effective $NO_2$ lifetime was ~1.5 hours for the satellite and ~0.9 hours for the model, and is derived using Equation 1. This suggests that the model simulation might have a $NO_2$ chemical lifetime that is too short.

For the Limestone Power Plant, the satellite-based method estimated the NOx emissions to be 2.7 Gg/yr, while reported emissions were 7.0 Gg/yr. This suggests that TROPOMI is having difficulty capturing the full magnitude of power plant plumes in Texas. This method has been applied to other power plants outside Texas, which are more isolated and have large emissions (e.g., Colstrip in Montana), and there is much better agreement at those locations. Further research will investigate this discrepancy at power plants in other US locations.




| Dallas, TX | NOx Emissions | | NO2 lifetime | |
|---|---|---|---|---|
| TROPOMI | 60 | Gg/ yr | 1.6 | hr |
| TROPOMI new AMF | 66 | Gg/ yr | 1.4 | hr |
| TROPOMI b-c | 75 | Gg/ yr | 1.6 | hr |
| TROPOMI b-c new AMF | 83 | Gg/ yr | 1.4 | hr |
| CAMx | 73 | Gg/ yr | 0.9 | hr |

| Limestone Power Plant | NOx Emissions | | NO2 lifetime | |
|---|---|---|---|---|
| TROPOMI | 1.9 | Gg/ yr | 0.6 | hr |
| TROPOMI new AMF | 2.1 | Gg/ yr | 0.5 | hr |
| TROPOMI b-c | 2.3 | Gg/ yr | 0.6 | hr |
| TROPOMI b-c new AMF | 2.6 | Gg/ yr | 0.5 | hr |
| CAMx | 7.0 | Gg/ yr | 0.9 | hr |

*Figure 3-1. EMG method to derive NOx emissions applied to (top) Dallas-Fort Worth and (bottom) Limestone Power Plant using TROPOMI. ERA5 100-m winds used to rotate daily TROPOMI $NO_2$ plumes, while average of WRF winds between 0 - 500 m are used to rotate CAMx $NO_2$ plumes. Tables on the right show the emissions for various "research-grade" TROPOMI products in addition to standard TROPOMI (TM5 AMF) product and CAMx.*

# 4.0 WHAT CAN WE LEARN ABOUT EMISSIONS FROM SATELLITES WITHOUT PROCESSING?

Despite the atmospheric processes that obscure the relationship between emissions and $NO_2$ column, there have been a number of applications where satellite data are used to directly infer spatial and temporal patterns in surface NOx emissions. Here we quantify that relationship relative to other methods discussed above. To benchmark the satellite-emissions relationship, we compare with the CAMx-calculated column $NO_2$ (calculated with the averaging kernel) and its relationship to emissions. By examining the CAMx column, we have insight into the physical and chemical processes decoupling emissions from column abundance, independent of satellite accuracy.

## 4.1 Direct Comparison of column $NO_2$ and emissions from cities

We find a weak relationship between TROPOMI and area NOx emissions, with correlations less than 0.2 (Figures 4-1 – 4-5). For example, in the Dallas – Fort Worth area r = 0.18 between TROPOMI $NO_2$ and emissions (Figure 4-4).

Even aside from satellite accuracy, multiple atmospheric processes decouple the relationship between emissions and column. To determine theoretical upper bound values, we calculated CAMx Correlations in daily modeled emissions vs. modeled column, which range r = 0.1 – 0.4 for the five study cities, and 0.0 – 0.5 for the five study power plants. As expected, actual TROPOMI vs. emissions correlations are worse these theoretical upper bound values with r-values ~0 – 0.2. On longer time periods, correlations improve. Figure 4-6 considers these comparisons for the case of DFW to highlight the different levels of agreement between emissions, surface $NO_2$ (from model), column $NO_2$ (from model), and column $NO_2$ (from satellite).

In general, we expect that NOx emissions have higher correlations with near-surface $NO_2$ than with column $NO_2$, given shorter pathways from emissions to ground-based monitors or simulated surface-layer amounts. With self-contained inputs, including meteorology, modeled emissions have higher correlations with modeled column amounts than with observed column amounts—for the five cities discussed here, we find the relationships between emissions and TROPOMI $NO_2$ are about half as strong as the relationships between emissions and CAMx column $NO_2$, as illustrated in Figure 4-6.

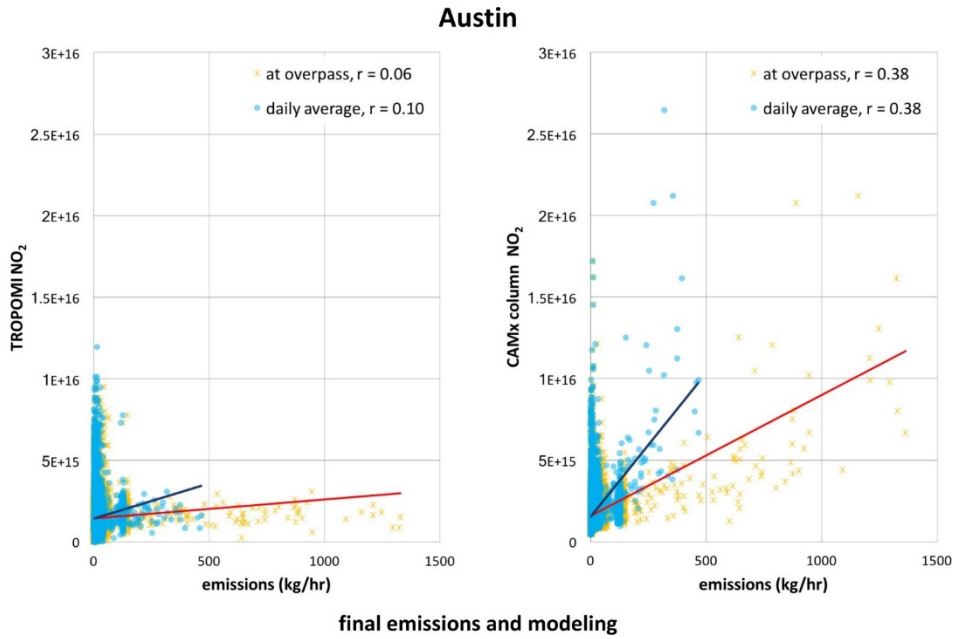



*Figure 4-1. TROPOMI (left) and CAMx column $NO_2$ (right) for Austin (Travis County) compared to aggregated anthropogenic area NOx emissions at the TROPOMI overpass (orange stars) and the daily average emissions (blue circles), for the period April – September 2019.*

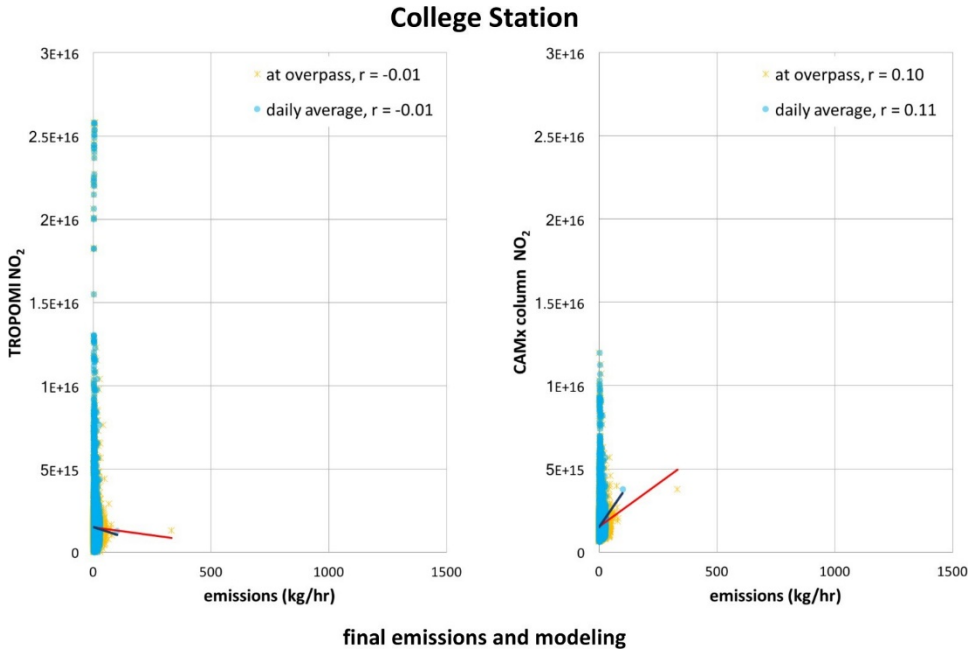



*Figure 4-2. TROPOMI (left) and CAMx column $NO_2$ (right) for College Station (Brazos County) compared to aggregated anthropogenic area NOx emissions at the TROPOMI overpass (orange stars) and the daily average emissions (blue circles), for the period April – September 2019.*

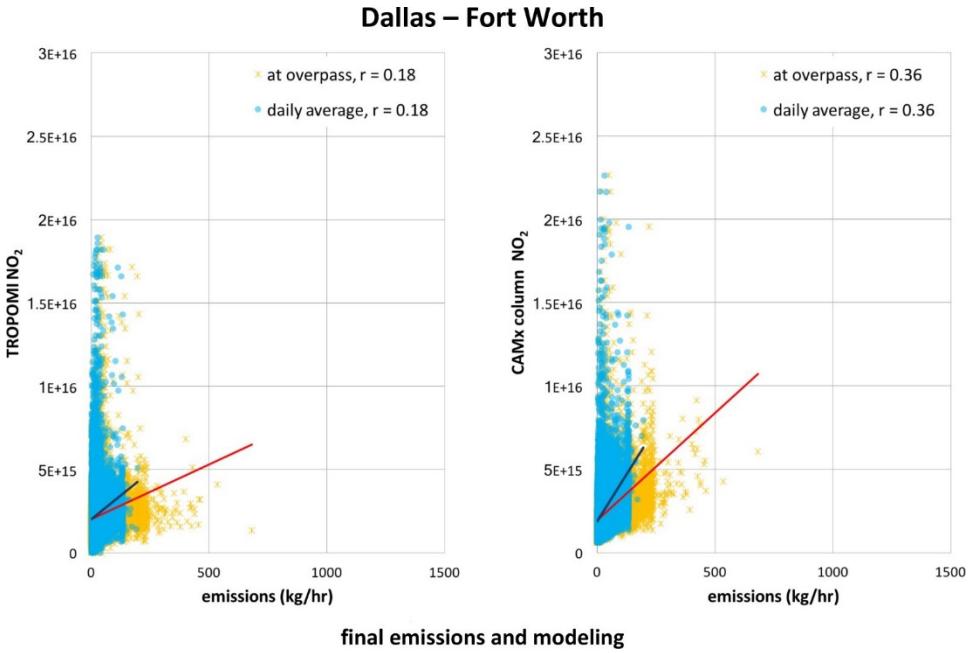



*Figure 4-3. TROPOMI (left) and CAMx column $NO_2$ (right) for Dallas and Fort Worth (Dallas and Tarrant counties, respectively) compared to aggregated anthropogenic area NOx emissions at the TROPOMI overpass (orange stars) and the daily average emissions (blue circles), for the period April – September 2019.*

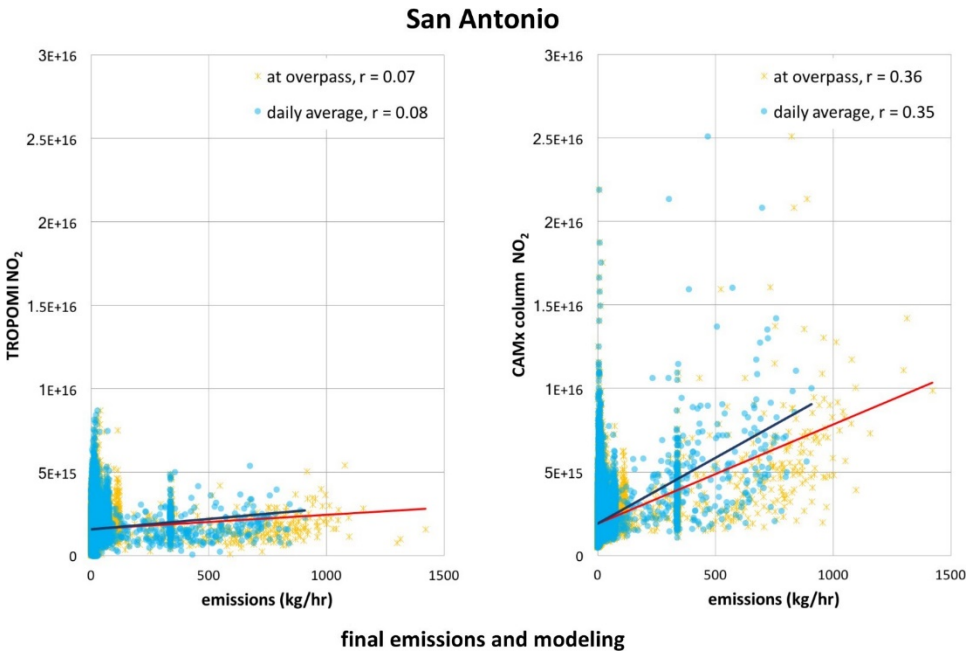



*Figure 4-4. TROPOMI (left) and CAMx column $NO_2$ (right) for San Antonio (Bexar County) compared to aggregated anthropogenic area NOx emissions at the TROPOMI overpass (orange stars) and the daily average emissions (blue circles), for the period April – September 2019.*

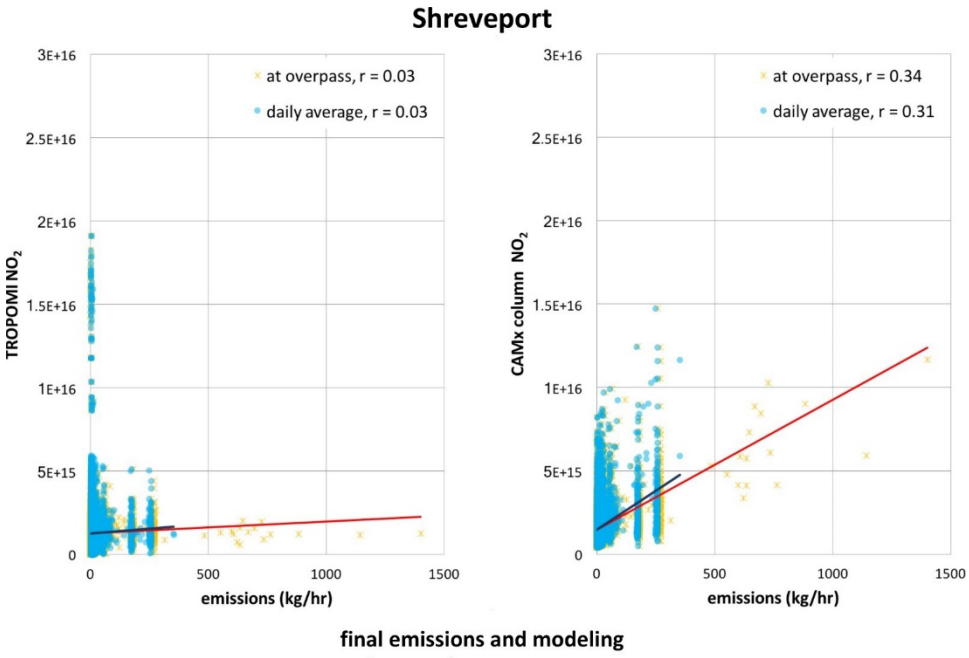



*Figure 4-5. TROPOMI (left) and CAMx column $NO_2$ (right) for Shreveport (Caddo County) compared to aggregated anthropogenic area NOx emissions at the TROPOMI overpass (orange stars) and the daily average emissions (blue circles), for the period April – September 2019.*

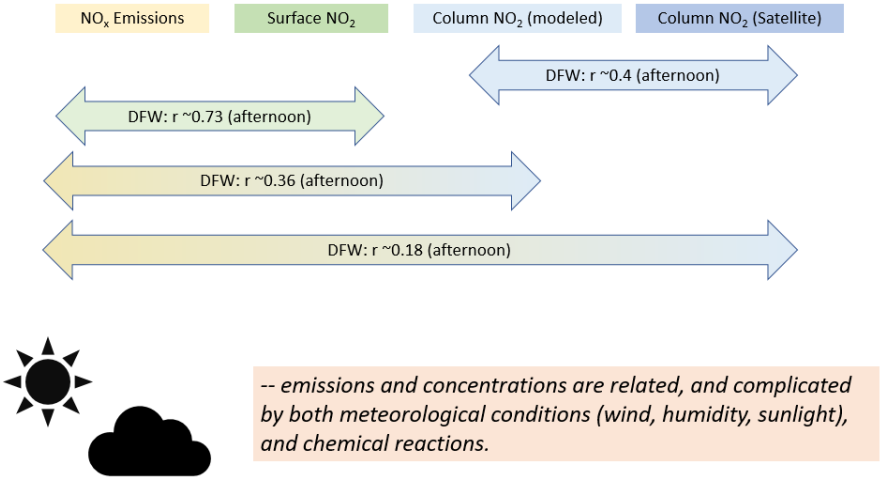



*Figure 4-6. A summary of correlations among emissions and $NO_2$ amounts as simulated by CAMx and observed by TROPOMI for Dallas-Fort Worth.*

## 4.2 Direct Comparison of column $NO_2$ and emissions from power plants

Power plants may have similar NOx emissions as urban areas. NOx emissions for the five cities discussed above range up to ~1400 kg/hour, or ~1.5 tons/hour; NOx emissions for the five power plants discussed in this section range up to ~1.8 tons/hour. However, power plants have a smaller spatial footprint than urban areas, and as such, the emissions signature may be more affected by meteorological conditions and more challenging to detect on a daily basis than for urban areas. This challenge is illustrated in Figures 4-7 – 4-11, which present each TROPOMI overpass around each power plant over the course of July 2019. The emissions from Fayette (Sam Seymour) are evident in ~19% of overpasses (Figure 4-7); emissions from Forney are not clearly evident, and only slightly evident in 16% of overpasses (Figure 4-8); emissions from Limestone are evident in 13.5% of overpasses (Figure 4-9); emissions from Martin Lake are evident in 40.5% of overpasses (Figure 4-10); emissions from Oklaunion are evident in ~43% of overpasses (Figure 4-11).

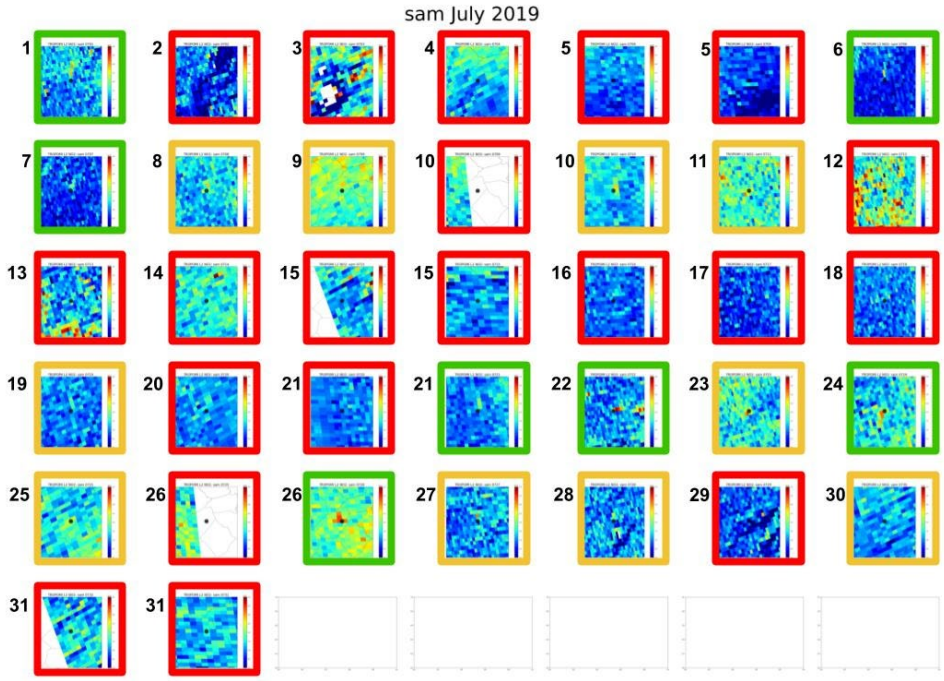



*Figure 4-7. Level-2 TROPOMI $NO_2$ from each overpass over Fayette (a.k.a. Sam Seymour; location given by black dot in the center) during July 2019. Where emissions are clearly evident the overpass image is highlighted with green; where emissions are slightly evident the overpass image is highlighted with yellow; where emissions are not evident the overpass image is highlighted in red.*

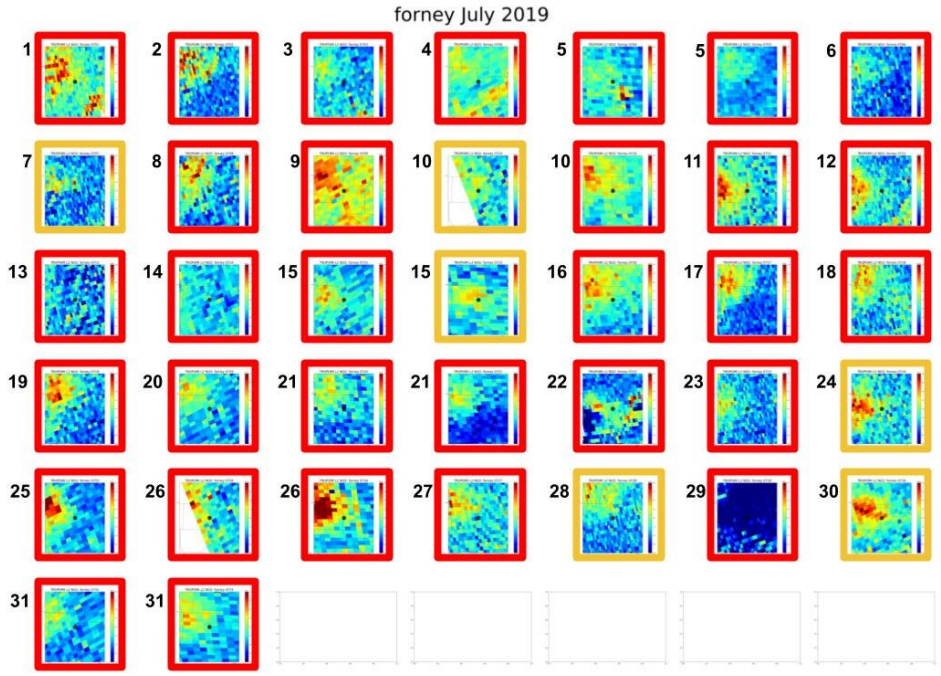



*Figure 4-8. Level-2 TROPOMI $NO_2$ from each overpass over Forney (location given by black dot in the center) during July 2019. Where emissions are clearly evident the overpass image is highlighted with green; where emissions are slightly evident the overpass image is highlighted with yellow; where emissions are not evident the overpass image is highlighted in red.*

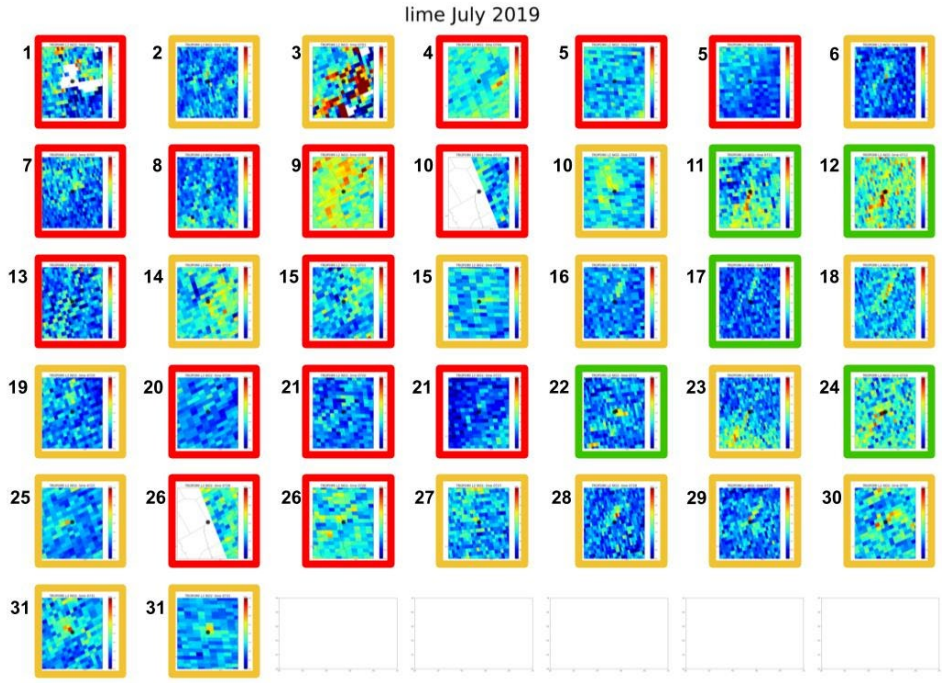



*Figure 4-9. Level-2 TROPOMI $NO_2$ from each overpass over Limestone (location given by black dot in the center) during July 2019. Where emissions are clearly evident the overpass image is highlighted with green; where emissions are slightly evident the overpass image is highlighted with yellow; where emissions are not evident the overpass image is highlighted in red.*

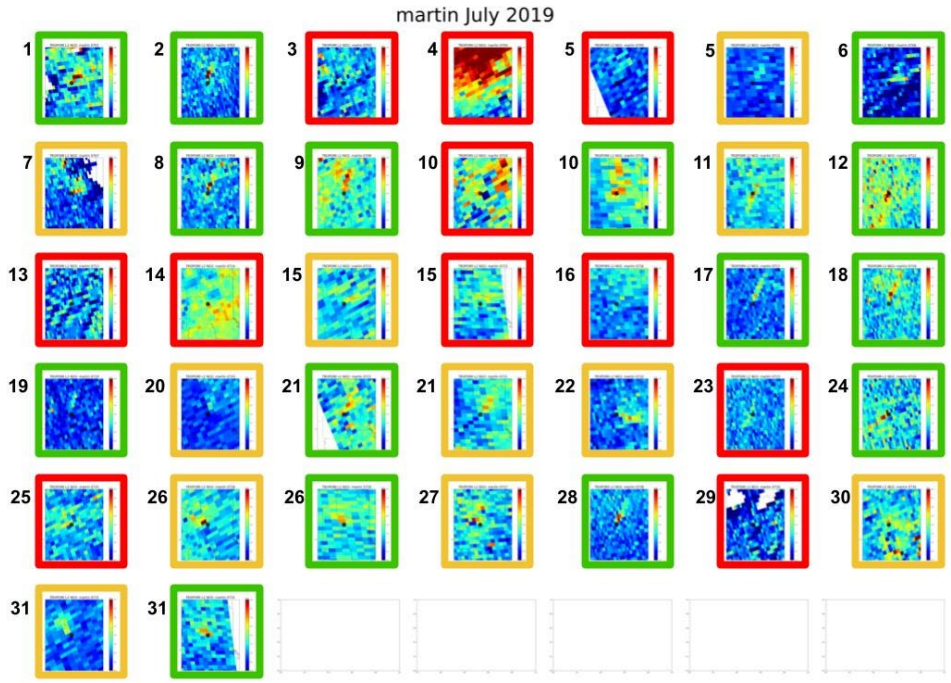



*Figure 4-10. Level-2 TROPOMI $NO_2$ from each overpass over Martin Lake (location given by black dot in the center) during July 2019. Where emissions are clearly evident the overpass image is highlighted with green; where emissions are slightly evident the overpass image is highlighted with yellow; where emissions are not evident the overpass image is highlighted in red.*

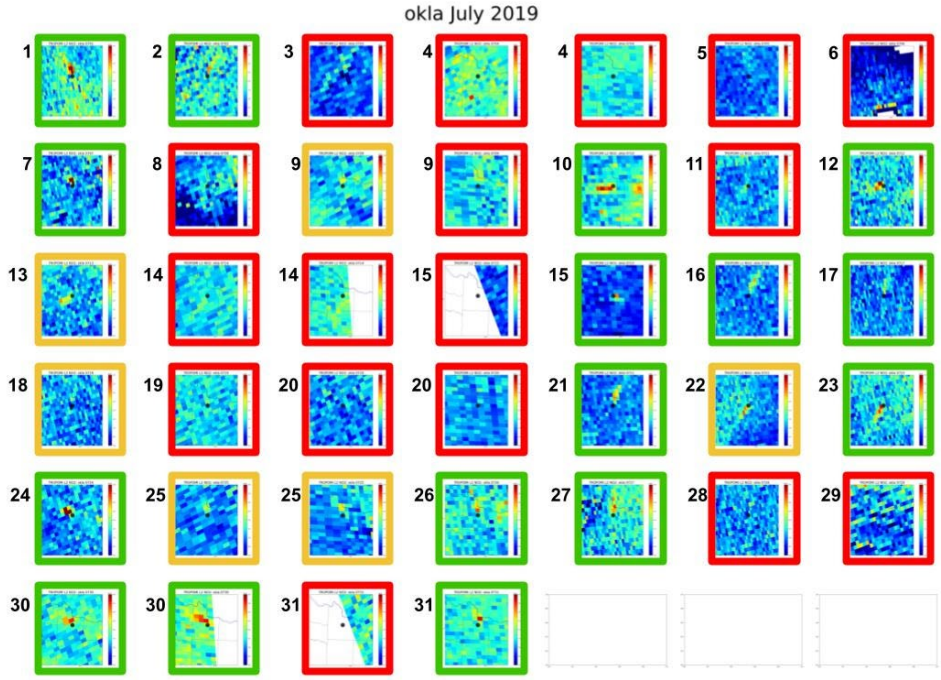



*Figure 4-11. Level-2 TROPOMI $NO_2$ from each overpass over Oklaunion (location given by black dot in the center) during July 2019. Where emissions are clearly evident the overpass image is highlighted with green; where emissions are slightly evident the overpass image is highlighted with yellow; where emissions are not evident the overpass image is highlighted in red.*

As seen in the relationships between city-area column $NO_2$ and NOx emissions, we find similar correlations between emissions at the mid-day TROPOMI overpass time and mid-day column $NO_2$, and between daily average power plant NOx emissions and mid-day column $NO_2$ (Figures 4-12 – 4-16). Similar to our evaluation over cities, too, we find greater correlations between model column $NO_2$ and power plant NOx (r = -0.10 - 0.25) than between TROPOMI $NO_2$ and power plant NOx (r = 0.02 - 0.49; Figures 4-12 – 4-16).

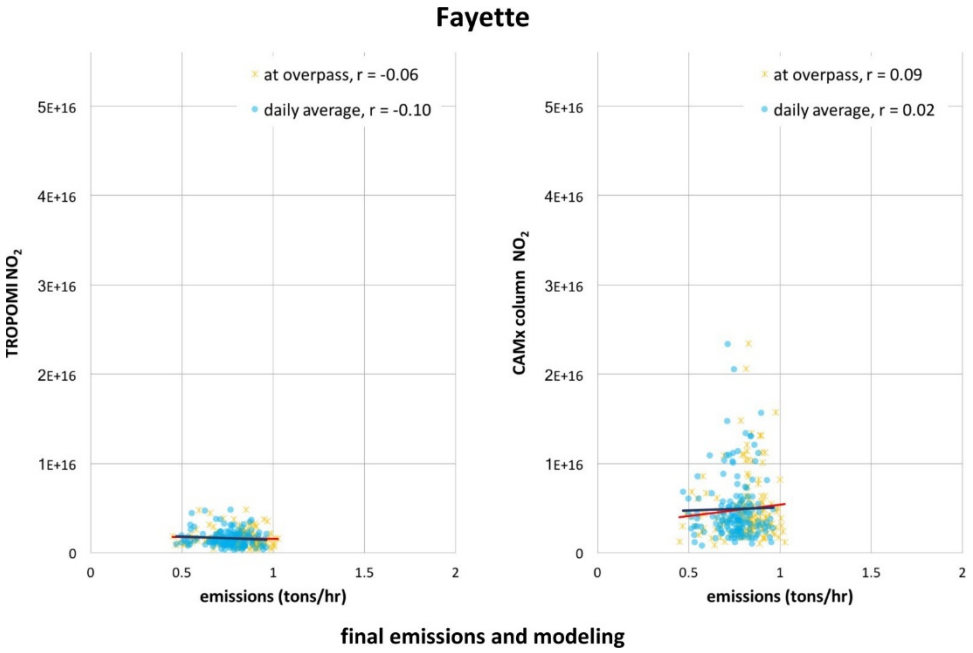



*Figure 4-12. TROPOMI (left) and CAMx column $NO_2$ (right) compared to NOx emissions at Fayette at the TROPOMI overpass (orange stars) and the daily average emissions (blue circles), for the period April – September 2019.*

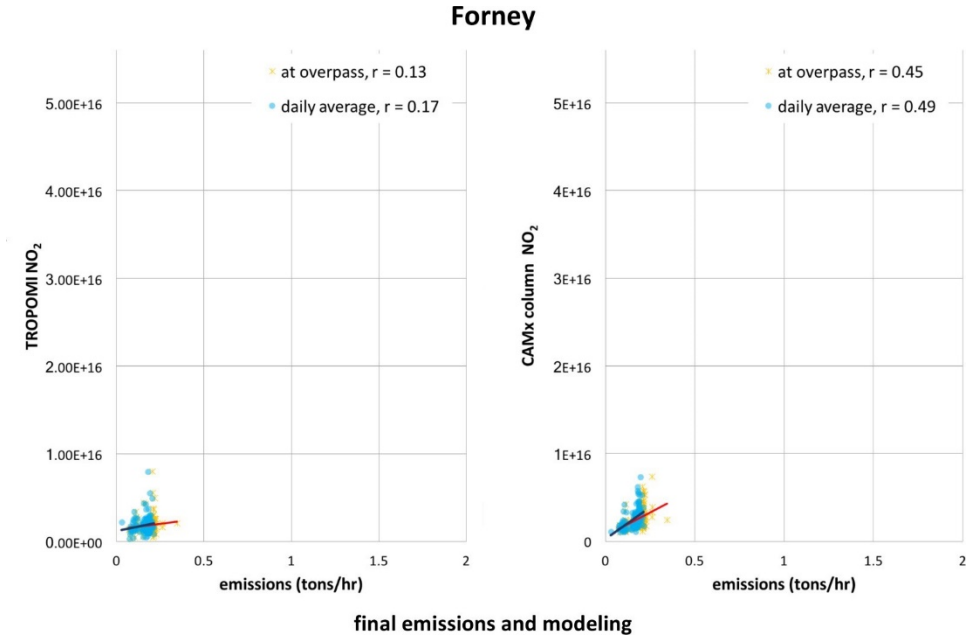



*Figure 4-13. TROPOMI (left) and CAMx column $NO_2$ (right) compared to NOx emissions at Forney at the TROPOMI overpass (orange stars) and the daily average emissions (blue circles), for the period April – September 2019.*

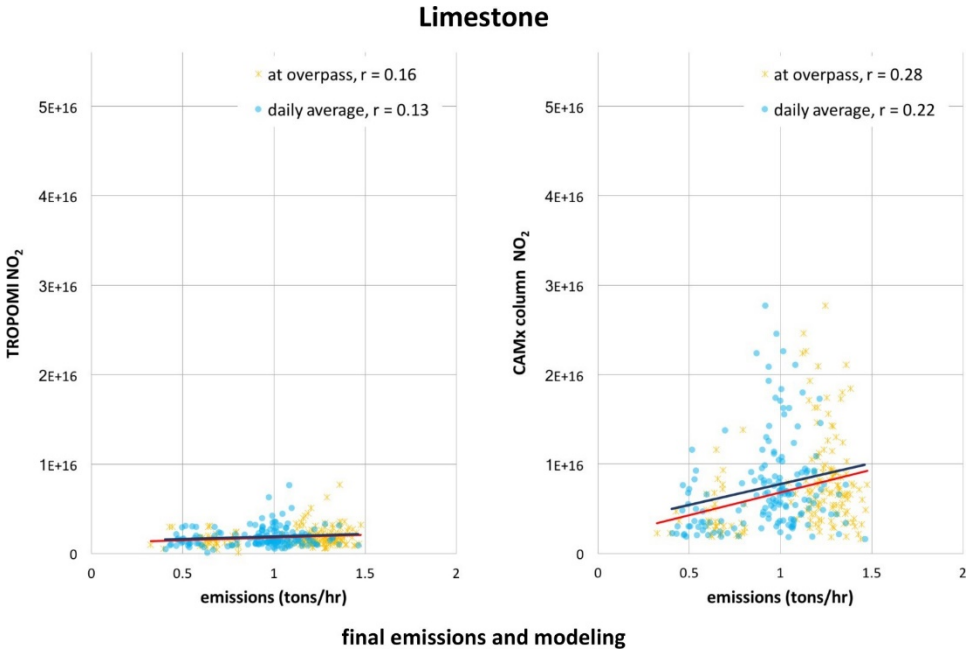



*Figure 4-14. TROPOMI (left) and CAMx column $NO_2$ (right) compared to NOx emissions at Limestone at the TROPOMI overpass (orange stars) and the daily average emissions (blue circles), for the period April – September 2019.*

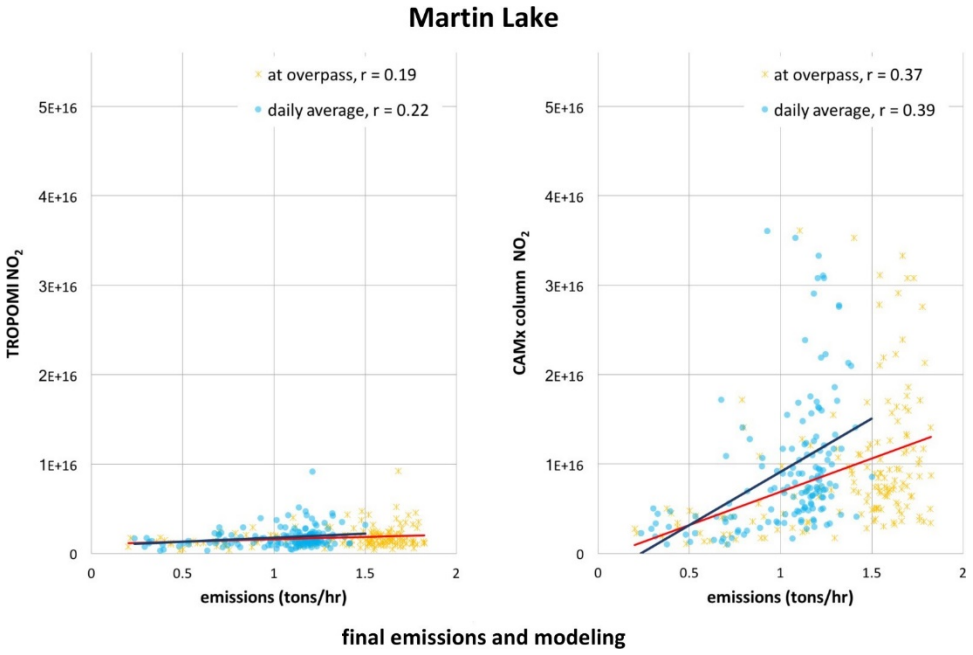



*Figure 4-15. TROPOMI (left) and CAMx column $NO_2$ (right) compared to NOx emissions at Marin Lake at the TROPOMI overpass (orange stars) and the daily average emissions (blue circles), for the period April – September 2019.*

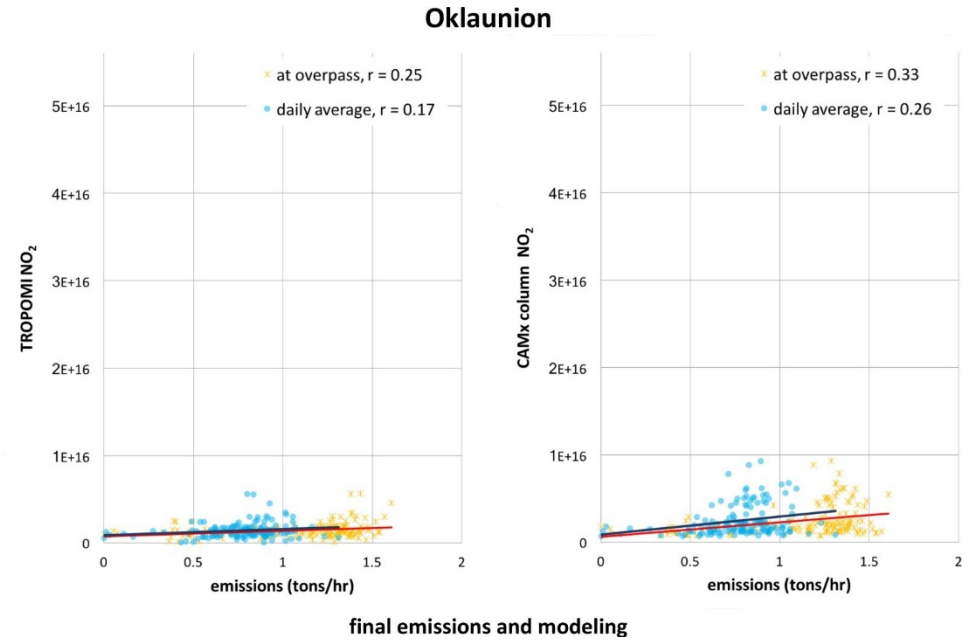



*Figure 4-16. TROPOMI (left) and CAMx column $NO_2$ (right) compared to NOx emissions at Oklaunion at the TROPOMI overpass (orange stars) and the daily average emissions (blue circles), for the period April – September 2019.*

The fraction of NOx that attributed to onroad mobile emissions for the Dallas-Ft Worth area are shown in Table 4-1. In Dallas, mobile source emissions represent approximately 1/3 of all NOx emissions. While this value seems appropriate for the full metropolitan area, we would have expected the fraction to have been larger in the downtown urban core. This suggests a potential spatial misallocation, but otherwise the magnitude of onroad mobile NOx emissions in the aggregated metropolitan area appears to be reasonable, as discussed in earlier sections.

*Table 4-1. Total and mobile NOx emissions within various radii of the Dallas city center*

| **Dallas Ft Worth** | | | |
|---|---|---|---|
| **Radius** | **All NOx (Gg/yr)** | **Fraction onroad mobile** | **Onroad mobile NOx (Gg/yr)** |
| 5 km | 1.8 | 0.33 | 0.6 |
| 10 km | 4.5 | 0.40 | 1.8 |
| 20 km | 18 | 0.32 | 5.6 |
| 30 km | 33 | 0.33 | 11 |
| 40 km | 48 | 0.33 | 16 |
| 50 km | 58 | 0.34 | 20 |
| 60 km | 65 | 0.34 | 22 |
| 70 km | 73 | 0.33 | 24 |
| 80 km | 81 | 0.32 | 26 |

# 5.0 CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

Through this analysis, we have concluded that a 3-D model is really the only tool for evaluating emission inventories on a regional and daily scales, supporting comparison with ground-based and satellite data. To support TCEQ in future model-satellite comparisons, we have updated the Wisconsin Horizontal Interpolation Program for Satellites (WHIPS) to process TROPOMI, and provided additional guidance and comparison on "user-friendly" and "research-grade" model-satellite comparison approaches. Gaussian plume methods succeed for well-detected sources over longer timescales, offering a lower-cost strategy to account for meteorology and chemistry. Although this work was motivated by the idea that satellite data provides a "check" on models and emission inventories, there were instances where disagreement suggests errors in the satellite's detection ability.

## 5.1 Summary of Findings

Overall, we find CAMx shows skill relative to ground- and satellite-based observations. However, the complicated pathways between emissions and column amounts render only moderate relationships between emissions and modelled column amounts; the relationships between emissions and TROPOMI $NO_2$ are complicated further by meteorology and ongoing work to improve satellite retrievals. The EMG methodology addresses the impact of meteorological differences between observations and modelling, yet we find the TROPOMI $NO_2$ retrieval used in this work has limited application to directly informing emissions inventories.

Compared to ground-based observations over the April – September 2019 study period, we find that CAMx shows good skill in identifying low and high ozone days for cities in Texas with $R^2$ values from 0.56 (Austin) to 0.61 (Tyler). CAMx has a positive ozone bias that is regional across eastern Texas but does not lead to excessive model error. We found that CAMx performs well in simulating the amount of ozone produced by emissions in the DFW area. We attempted to evaluate CAMx $NO_2$ performance focusing on the mid-day time period when satellite $NO_2$ measurements are made, but the evaluation was hampered by instrument detection limits and locations.

Compared to TROPOMI $NO_2$ columns, we find CAMx also shows skill, and has an overall positive bias of about 13%, depending on the domain. To the best of our knowledge, there is no prior work describing CTM performance over the U.S. with respect to TROPOMI $NO_2$ to compare with this study. However, past work shows both the operational retrieval and "research grade" TROPOMI $NO_2$ products have a low bias relative to ground-based tropospheric column measurements over North America [*Griffin et al*., 2019; *Judd et al.*, 2020; *Zhao et al*., 2020]. We demonstrate that a bias-corrected TROPOMI largely removes the aforementioned 13% difference, thus concluding that Texas urban NOx emissions inventory are within the uncertainty bounds of the accuracy of the satellite measurement.

We find disagreement between CAMx and TROPOMI near power plants, with the CAMx columns being larger. This positive bias is seen both in a direct comparison of TROPOMI $NO_2$ with simulated column amounts (with TROPOMI averaging kernel applied), and with TROPOMI $NO_2$ refactored using model-derived *a priori* profiles and variability. With differences in native resolution and the ratio of emissions relative to background amounts, some power plants and smaller urban areas could not be meaningfully differentiated in TROPOMI observations. The inability of TROPOMI to detect power plants was even greater for $SO_2$, as illustrated in Figure B-1. While the lower TROPOMI $NO_2$ columns might be interpreted to suggest that the CAMx NOx emission inventory for major power plants (i.e., Martin Lake and Limestone) is over-reported, such an assertion would be difficult to reconcile with the CAMx emissions for these power plants being traceable to real-time emission measurements (by CEMS) in the power plant stacks.

In our TROPOMI-CAMx column $NO_2$ comparison, we find CAMx overestimates column $NO_2$ by ~4% (Dallas) to 25% (San Antonio); when we correct simulated background $NO_2$, we find $NO_2$ matches observed amounts very well in urban areas (<5% difference in the investigated cities when aggregated). When meteorological conditions are accounted for via EMG analysis, we find differences in modeled and observed $NO_2$ are likely the result of spatial misallocation of emissions rather than an areawide magnitude disagreement.

We find poor agreement between TROPOMI and CAMx over power plants. In our direct comparison, we find an average correlation of 0.33, not including two power plants that could not be meaningfully differentiated from background observations (Forney and Oklaunion). Since power plant emissions are well constrained by in-stack measurements, this disagreement is likely the result of the difficulties in resolving narrow plume widths, and short effective $NO_2$ lifetimes. When we imposed model variability on TROPOMI $NO_2$, the correlation for Martin Lake, Limestone, and Sam Seymour (Fayette) improved from 0.47 to 0.91.

Although best-practice, the utilization of a PGM, such as CAMx, is time-consuming and requires specialized knowledge. We characterized the relationship between emissions and satellite-based observations both directly and with the EMG method. These simpler methods offer the potential for TCEQ to perform emissions evaluation with satellite data analysis over multiple years and/or considering multiple emission scenarios at a greatly reduced cost. We found the EMG most useful for determining aggregated emissions from relatively large urban areas, such as Dallas-Fort Worth, and less useful for smaller urban areas (e.g. Austin, College Station, and San Antonio) and power plants in Texas.

Without processing by the CAMx model or EMG, relationships between TROPOMI $NO_2$ column and emissions are poorly defined, highlighting the value of more advanced approaches. The low agreement between satellite and emissions may be partially explained by physical and chemical processes decoupling emissions from column abundance. These physical processes were isolated by comparing CAMx-calculated columns with emissions (in a manner parallel to the comparison of TROPOMI data and emissions). Whereas the correlation between CAMx $NO_2$ column and NOx emissions over urban counties varies from 0.10 – 0.38, the correlation between TROPOMI $NO_2$ and NOx emissions varies from -0.01 – 0.18. Where the correlation between CAMx $NO_2$ column

and NOx emissions from power plants varies from 0.02 – 0.49, the correlation between TROPOMI $NO_2$ and NOx emissions varies from -0.10 – 0.25.

Multiple atmospheric processes decouple the relationship between NOx emissions and column $NO_2$. Satellite data, on their own, offer qualitative information on emission patterns and trends, but data resolution and retrieval inputs may weaken the suitability of satellite data for quantitative emission inventory evaluation. The relationship between TROPOMI $NO_2$ and NOx emissions is strengthened when meteorological conditions are accounted for, as used in the EMG analysis. As such, we find the more streamlined and less computationally intensive EMG analysis is suitable for informing emissions where observed column $NO_2$ is greater than background amounts.

## 5.2 Recommendations for Future Work

- A substantial portion of the $NO_2$ column resides above the PBL and this $NO_2$ can strongly influence comparisons between satellite and modeled $NO_2$ columns. CAMx showed reasonable performance in simulating $NO_2$ in the free troposphere. Nevertheless, we recommend continued efforts to improve the CAMx simulation of $NO_2$ in the mid- to upper-troposphere by focusing on: (1) emission estimates for lightning and aircraft NOx (2) representation of NOx influx from the lower-stratosphere to the upper-troposphere (3) ventilation of NOy from the PBL to the mid-troposphere (4) photochemical production of $NO_2$ from NOy (especially organic nitrates) in the mid- to upper-troposphere.
- We found TROPOMI $NO_2$ was generally lower than CAMx column amounts, which is consistent with prior work comparing TROPOMI with ground-based tropospheric column measurements. In fact, at the end of this study (July 2021), the TROPOMI retrieval was updated to address the documented low bias. Given the likelihood that the new TROPOMI $NO_2$ retrieval shows an improved agreement with CAMx column $NO_2$, it may be worthwhile to quantify resulting changes in TROPOMI relationships with emissions over urban areas, and especially over power plants.
- We found poor agreement between TROPOMI and CAMx $NO_2$ columns near major power plants in Texas. Additional study is needed with consideration of the both the TROPOMI data products (particularly the horizontal resolution of the *a priori* $NO_2$ column used in calculating the averaging kernel) and the model simulations (e.g. the vertical distribution of $NO_2$, photochemical lifetime of $NO_2$ in power plant plumes). It would be helpful to have column measurements, such as GeoTASO or Pandora, to further validate TROPOMI measurements in the presence of power plant plumes.
- Future work could address challenges with $SO_2$, but may require careful processing to discern a clear emissions signal from TROPOMI. The TEMPO satellite planned for launch in 2022 will provide higher resolution, hourly $SO_2$ data. These may be needed for meaningful $SO_2$ analysis for Texas power plants.

# 6.0 AUDITS OF DATA QUALITY

We performed Quality Assurance/Quality Control (QA/QC) procedures in accordance with the Quality Assurance Project Plan (QAPP) completed at the beginning of this project. Per requirements for Category III projects, we performed data audits on at least 10% of the data sets. In this section, we report the results of our QA/QC.

## 6.1 CAMx simulations

The WRF-CAMx model was run by Ramboll. Validation of WRF-CAMx modeling with the NRTEEM platform is described by *Johnson et al.* [2019]. Ambient $NO_X$ concentrations simulated by the WRF-CAMx modeling were compared with all available ground-based observations of $NO_X$ during the modeling time period, as detailed in Section 3.1.

## 6.2 TROPOMI screening

We employed two methodologies for comparison of model column $NO_2$ with TROPOMI $NO_2$. The direct-comparison methodology used model column amounts calculated with the application of the TROPOMI averaging kernel. As appropriate for work where the averaging kernel is applied, this methodology used all TROPOMI data where the quality assurance flag was greater than 0.5 [*Eskes*, 2021]. Our second method entailed a recalculation of the air mass factor, which is a component of the averaging kernel, using CAMx column $NO_2$. Since this method did not employ the averaging kernel, for this methodology we used all TROPOMI data where the quality assurance flag was greater than 0.75. The latter TROPOMI VCD screening was also employed for the EMG analysis. We found the April – September 2019 TROPOMI $NO_2$ average calculated with the two methodologies was within 10%, with larger TROPOMI $NO_2$ amounts when data was screened using a quality assurance flag of 0.5.

## 6.3 EMG Analysis

The EMG technique presented by *Goldberg et al.* [2019a, 2019b, 2019c] was selected for its ability to produce emission estimates directly from TROPOMI data with minimal additional data. Calculation of parameters in the EMG approach are based on wind speed and direction, which were taken from the Ramboll WRF simulations at 4km. We also applied the EMG technique using widely available re-analysis data, the ECMWF ERA-5 [*Hersbach et al.,* 2020], for the purpose of anticipating how using re-analysis data (rather than WRF simulations) may influence the EMG analysis in the future if WRF simulations are not available. We focused on the meteorological parameter that EMG uses, wind speed. We found the $NO_X$ emissions that resulted from using the two different wind speed datasets were within 20%. Therefore, wind speed may contribute up to a third of the total uncertainty of the EMG method (~60%).

**References**


Banta, R.M., C.J. Senff, J. Nielsen-Gammon, L.S. Darby, T.B. Ryerson, R.J. Alvarez, S.P. Sandberg, E.J. Williams, and M. Trainer, 2005: A Bad Air Day in Houston. Bull. *Amer. Meteor. Soc.*, 86, 657–670, https://doi.org/10.1175/BAMS-86-5-657.

Beirle, S., K. F. Boersma, U. Platt, M. G. Lawrence, and T. Wagner (2011), Megacity emissions and lifetimes of nitrogen oxides probed from space, *Science*, 333(6050), 1737-1739, doi:10.1126/science.1207824.

Canty, T. P., L. Hembeck, T. P. Vinciguerra, D. C. Anderson, D. L. Goldberg, S. F. Carpenter, D. J. Allen, C. P. Loughner, R. J. Salawitch, and R. R. Dickerson (2015), Ozone and NOx chemistry in the eastern US: Evaluation of CMAQ/CB05 with satellite (OMI) data, *Atmos. Chem. Phys.*, 15(19), 10965–10982, doi:10.5194/acp-15-10965-2015.

Denier van der Gon, H., Hendriks, C., Kuenen, J., Segers, A., & Visschedijk, A. (2011). Description of current temporal emission patterns and sensitivity of predicted AQ for temporal emission patterns EU FP7 MACC deliverable report D_D-EMIS_1.3. Retrieved from https://atmosphere.copernicus.eu/sites/default/files/2019-07/MACC_TNO_del_1_3_v2.pdf

Duncan, B. N., A. I. Prados, L. N. Lamsal, Y. Liu, D. G. Streets, P. Gupta, E. Hilsenrath, R. a. Kahn, J. E. Nielsen, A. J. Beyersdorf, S. P. Burton, A. M. Fiore, J. Fishman, D. K. Henze, C. a. Hostetler, N. A. Krotkov, P. Lee, M. Lin, S. Pawson, et al. (2014), Satellite data of atmospheric pollution for U.S. air quality applications: Examples of applications, summary of data end-user resources, answers to FAQs, and common mistakes to avoid, *Atmos. Environ.*, 94, 647–662, doi:10.1016/j.atmosenv.2014.05.061.

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

de Foy, B., Z. Lu, D. G. Streets, L. N. Lamsal, and B. N. Duncan (2015), Estimates of power plant NOx emissions and lifetimes from OMI NO2 satellite retrievals, *Atmos. Environ*., 116(2), 1–11, doi:10.1016/j.atmosenv.2015.05.056.

Deeter, M. N. Calculation and Application of MOPITT Averaging Kernels. (2002) Natl. Cent. Atmospheric Res.

Ek, M., Bill Lapenta, Geoff DiMego, Hendrik Tolman, John Derber, Yuejian Zhu, Vijay Tallapragada, Mark Iredell, Shrinivas Moorthi, Suru Saha. 2014. The NOAA Operational Numerical Guidance System. 29th Session of the World Weather/Climate Research Programme (WWRP/WCRP) Working Group on Numerical Experimentation (WGNE-29) Bureau of Meteorology, Melbourne, Australia, 10-14 March 2014.

Emery, C., Z. Liu, A. G. Russell, M. T. Odman, G. Yarwood and N. Kumar. (2017) Recommendations on statistics and benchmarks to assess photochemical model performance, Journal of the Air & Waste Management Association, 67:5, 582-598, DOI: 10.1080/10962247.2016.1265027.

Eskes, H.; van Geffen, J.; Boersma, F.; Eichmann, K.-U.; Apituley, A.; Pedergnana, M.; Sneep, M.; Veefkind, J. P.; Loyola, D. (2021) Sentinel-5 Precursor/TROPOMI Level 2 Product User Manual Nitrogen Dioxide. Royal Netherlands Meteorological Institute.

Goldberg, D. L., P. E. Saide, L. N. Lamsal, B. De Foy, Z. Lu, J. H. Woo, Y. Kim, J. Kim, M. Gao, G. Carmichael, and D. G. Streets (2019a), A top-down assessment using OMI NO2 suggests an underestimate in the NOx emissions inventory in Seoul, South Korea, during KORUS-AQ, *Atmos. Chem. Phys*., 19(3), 1801–1818, doi:10.5194/acp-19-1801-2019.

Goldberg, D. L., Z. Lu, D. G. Streets, B. De Foy, D. Griffin, C. A. Mclinden, L. N. Lamsal, N. A. Krotkov, and H. Eskes (2019b), Enhanced Capabilities of TROPOMI NO2: Estimating NOX from North American Cities and Power Plants, *Environ. Sci. Technol.*, 53(21), 12594–12601, doi:10.1021/acs.est.9b04488.

Goldberg, D. L., Z. Lu, T. Oda, L. N. Lamsal, F. Liu, D. Griffin, C. A. McLinden, N. A. Krotkov, B. N. Duncan, and D. G. Streets (2019c), Exploiting OMI NO2 satellite observations to infer fossil-fuel CO2 emissions from U.S. megacities, *Sci. Total Environ*., 695(2), doi:10.1016/j.scitotenv.2019.133805.

Griffin, D., X. Zhao, C. A. McLinden, F. Boersma, A. Bourassa, E. Dammers, D. Degenstein, H. Estes, L. Fehr, V. Fioletov, K. Hayden, S. K. Kharol, S.-M. Li, P. Makar, R. V. Martin, C. Mihele, R. L. Mittermeier, N. Krotkov, M. Sneep, L. N. Lamsal, M. ter Linden, J. van Geffen, P. Veefkind, and M. Wolde (2019), High-resolution mapping of nitrogen dioxide with TROPOMI: First results and validation over the Canadian Oil Sands, Geophys. Res. Lett., 46(2), 1049-1060, doi:10.1029/2018GL081095.

Guenther, A., T. Karl, P. Harley, C. Wiedinmyer, P. I. Palmer, and C. Geron (2006), Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), *Atmos. Chem. Phys. Discuss*., 6(1), 107–173, doi:10.5194/acpd-6-107-2006.

Harkey, M., T. Holloway, E. J. Kim, K. R. Baker, and B. Henderson (2021) Satellite formaldehyde to support model evaluation and health assessments, J. Geophys. Res.-Atmos., 126(4), e2020JD032881, DOI: 10.1029/2020JD032881.

Harkey, M., T. Holloway, J. Oberman, and E. Scotty (2015), An evaluation of CMAQ NO2 using observed chemistry-meteorology correlations, *J. Geophys. Res*., 120(22), doi:10.1002/2015JD023316.

Hersbach, H., B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Randu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. De Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R. J. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume, J.-N. Thépaut (2020), The ERA5 global reanalysis, *Q. J. R. Meteorol. Soc.*, 146(730), 1999-2049, doi:10.1002/qj.3803.

Holloway, T., D. J. Jacob, and D. Miller (2018), Short history of NASA applied science teams for air quality and health, *J. Appl. Remote Sens*., doi:10.1117/1.jrs.12.042611.

Jin, X., and T. Holloway (2015), Spatial and temporal variability of ozone sensitivity over China observed from the Ozone Monitoring Instrument*, J. Geophs. Res.-Atmos*., 120(14), 7229–7246, doi:10.1002/2015JD023250

Johnson, J., G. Wilson, S. Kemball-Cook, K. Tanner, Y. Shi, J. Guo, R. Beardsley, and G. Yarwood. (2019), Near-Real Time Exceptional Event Modeling, Prepared for Mark Estes, TCEQ.

Judd, L. M., J. A. Al-Saadi, J. J. Szykman, L. C. Valin, S. J. Janz, M. G. Kowalewski, H. J. Eskes, J. P. Veefkind, A. Cede, M. Mueller, M. Gebetsberger, R. Swap, R. B. Pierce, C. R. Nowlan, G. González Abad, A. Nehrir, and D. Williams (2020), TROPOMI tropospheric NO2 column densities with airborne and Pandora spectrometers near New York City and Long Island Sound, *Atmos. Meas. Tech*., 13(11), 6113-6140, doi:10.5194/amt-13-6113-2020.

Karambelas, A., T. Holloway, G. Kiesewetter, and C. Heyes (2018), Constraining the uncertainty in emissions over India with a regional air quality model evaluation, *Atmos. Environ*., 174, doi:10.1016/j.atmosenv.2017.11.052.

Kemball-Cook, S., G. Yarwood, J. Johnson, B. Dornblaser, and M. Estes (2015), Evaluating NOx emission inventories for regulatory air quality modeling using satellite and air quality model data, *Atmos. Environ*., 117(March 2016), 1–8, doi:10.1016/j.atmosenv.2015.07.002.

Kim, H. C. P. Lee., L. Judd, L. Pan, B. Lefer (2016) OMI NO2 Column Densities over North American Urban Cities: The Effect of Satellite Footprint Resolution. *Geosci. Model Dev.* 2016, 9 (3), 1111–1123. https://doi.org/10.5194/gmd-9-1111-2016.

Lamsal, L. N., R. V. Martin, a. Padmanabhan, a. van Donkelaar, Q. Zhang, C. E. Sioris, K. Chance, T. P. Kurosu, and M. J. Newchurch (2011), Application of satellite observations for timely updates to global anthropogenic NO x emission inventories, *Geophys. Res. Lett*., 38(5), 1–5, doi:10.1029/2010GL046476.

Lu, Z., D. G. Streets, B. De Foy, L. N. Lamsal, B. N. Duncan, and J. Xing (2015), Emissions of nitrogen oxides from US urban areas: Estimation from Ozone Monitoring Instrument retrievals for 2005- 2014, *Atmos. Chem. Phys*., 15(18), 10367–10383, doi:10.5194/acp-15-10367-2015.

Martin, R. V. (2003), Global inventory of nitrogen oxide emissions constrained by space-based observations of NO2 columns, *J. Geophys. Res*., 108(D17), 1–12, doi:10.1029/2003JD003453.

Martin, R. V. (2008), Satellite remote sensing of surface air quality, *Atmos. Environ*., 42(34), 7823–7843, doi:10.1016/j.atmosenv.2008.07.018.

Mijling, B., and R. J. Van Der A (2012), Using daily satellite observations to estimate emissions of short- lived air pollutants on a mesoscopic scale, *J. Geophys. Res. Atmos*., 117(17), 1–20, doi:10.1029/2012JD017817.

Montgomery, A., and T. Holloway (2018), Assessing the relationship between satellite-derived NO2 and economic growth over the 100 most populous global cities, *J. Appl. Remote Sens.*, doi:10.1117/1.jrs.12.042607.

Penn, E., and T. Holloway (2020), Evaluating Current Satellite Capability to Observe Diurnal Change in Nitrogen Oxides in Preparation for Geostationary Satellite Missions, *Environ. Res. Lett*., https://doi.org/10.1088/1748-9326/ab6b36.

Richter, A., and J. P. Burrows (2002), Tropospheric NO2 from GOME measurements, *Adv. Sp. Res*., 29(11), 1673–1683, doi:10.1016/S0273-1177(02)00100-X.

Richter, A., J. P. Burrows, H. Nüß, C. Granier, and U. Niemeier (2005), Increase in tropospheric nitrogen dioxide over China observed from space, *Nature*, 437(7055), 129–132, doi:10.1038/nature04092.

Skamarock, W. C. 2004. Evaluating Mesoscale NWP Models Using Kinetic Energy Spectra. *Mon. Wea. Rev.*, Volume 132, pp. 3019-3032. December. (http://www.mmm.ucar.edu/individual/skamarock/spectra_mwr_2004.pdf).

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang and J. G. Powers. 2005. A Description of the Advanced Research WRF Version 2. National Center for Atmospheric Research (NCAR), Boulder, CO. June. (http://www.mmm.ucar.edu/wrf/users/docs/arw_v2.pdf)

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X-Y. Huang, W. Wang and J. G. Powers. 2008. A Description of the Advanced Research WRF Version 3. National Center for Atmospheric Research (NCAR), Boulder, CO. June. (https://opensky.ucar.edu/islandora/object/technotes%3A500/datastream/PDF/view)

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, Z. Liu, J. Berner, W. Wang, J. G. Powers, M. G. Duda, D. M. Barker and X-Y. Huang. 2019. A Description of the Advanced Research WRF Model Version 4 (No. NCAR/TN-556+STR). doi:10.5065/1dfh-6p97 (https://opensky.ucar.edu/islandora/object/technotes%3A576/datastream/PDF/download/citation.pdf)

Skamarock, W. C. 2006. Positive-Definite and Monotonic Limiters for Unrestricted-Time-Step Transport Schemes. *Mon. Wea. Rev.*, Volume 134, pp. 2241-2242. June. (http://www.mmm.ucar.edu/individual/skamarock/advect3d_mwr.pdf).

Streets, D. G., T. Canty, G. R. Carmichael, B. De Foy, R. R. Dickerson, B. N. Duncan, D. P. Edwards, J. A. Haynes, D. K. Henze, M. R. Houyoux, D. J. Jacob, N. A. Krotkov, L. N. Lamsal, Y. Liu, Z. Lu, R. V. Martin, G. G. Pfister, R. W. Pinder, R. J. Salawitch, et al. (2013), Emissions estimation from satellite retrievals: A review of current capability, *Atmos. Environ.,* 77, 1011–1042, doi:10.1016/j.atmosenv.2013.05.051.

Streets, D. G., B. de Foy, B. N. Duncan, L. N. Lamsal, C. Li, and Z. Lu (2014), Using satellite observations to measure power plant emissions and their trends., *Environ. Manag*., (February).

Tong, D., L. Pan, W. Chen, L. Lamsal, P. Lee, Y. Tang, H. Kim, S. Kondragunta, and I. Stajner (2016), Impact of the 2008 Global Recession on air quality over the United States: Implications for surface ozone levels from changes in NO x emissions, *Geophys. Res. Lett.*, (x), 9280–9288, doi:10.1002/2016GL069885.

Valin, L. C., A. R. Russell, and R. C. Cohen (2013), Variations of OH radical in an urban plume inferred from NO2 column measurements*, Geophys. Res. Lett*., 40(9), 1856–1860, doi:10.1002/grl.50267.

Verhoelst, T., Compernolle, S., Pinardi, G., Lambert, J.-C., Eskes, H. J., Eichmann, K.-U., et al. (2021). Ground-based validation of the Copernicus Sentinel-5P TROPOMI NO2 measurements with the NDACC ZSL-DOAS, MAX-DOAS and Pandonia global networks. *Atmos. Meas. Tech.*, *14*(1), 481–510. https://doi.org/10.5194/amt-14-481-2021

Zhao, X., D. Griffin, V. Fioletov, C. McLinden, A. Cede, M. Tiefengraber, M. Müller, K. Bognar, K. Strong, F. Boersma, H. Eskes, J. Davies, A. Ogyu, S. C. Lee (2020), Assessment of the quality of TROPOMI high-spatial-resolution $NO_2$ data products in the Greater Toronto Area, *Atmos. Meas. Tech.*, 13(4), 2131-2159, doi:10.5194/amt-13-2131-2020.

## APPENDIX A

Here we present monthly TROPOMI and CAMx column $NO_2$ (with the averaging kernel) for each month of the study period. Figures A-1 and A-2, along with Tables A-1 and A-2 show monthly column $NO_2$, differences, and comparison metrics for each domain. Figures A-3 and A-4 present a zoom-in of the larger 12 km resolution domain over western Texas.

*Table A-1 TROPOMI and CAMx model mean column $NO_2$ for the months April – September 2019 and associated error metrics for the 12 km domain. All units in $10^{15}$ molecules/$cm^2$ unless otherwise noted.*

| | Model Mean | TROPOMI mean | RMSE | Mean Error | Mean Fractional Error (%) | Mean Bias | Mean Fractional Bias (%) | r |
|---|---|---|---|---|---|---|---|---|
| **April** | 0.84 | 0.94 | 1 | 0.43 | 97.52 | -0.1 | 49.52 | 0.3 |
| **May** | 0.94 | 0.88 | 0.84 | 0.42 | 49.28 | 0.06 | 17.19 | 0.36 |
| **June** | 1.14 | 0.97 | 0.97 | 0.47 | 39.57 | 0.17 | 16.56 | 0.32 |
| **July** | 1.34 | 1.08 | 0.85 | 0.42 | 37.66 | 0.25 | 24.7 | 0.39 |
| **August** | 1.51 | 1.14 | 1 | 0.53 | 35.86 | 0.37 | 24.83 | 0.37 |
| **September** | 1.28 | 1.1 | 0.79 | 0.4 | 35.12 | 0.18 | 16.59 | 0.37 |

*Table A-2 TROPOMI and CAMx model mean column $NO_2$ for the months April – September 2019 and associated error metrics for the 4 km domain. All units in $10^{15}$ molecules/$cm^2$ unless otherwise noted.*

| | Model Mean | TROPOMI mean | RMSE | Mean Error | Mean Fractional Error (%) | Mean Bias | Mean Fractional Bias (%) | r |
|---|---|---|---|---|---|---|---|---|
| **April** | 0.94 | 1.12 | 0.94 | 0.47 | 46.9 | -0.19 | -2.25 | 0.44 |
| **May** | 1.05 | 0.99 | 0.74 | 0.46 | 49.43 | 0.06 | 19.31 | 0.44 |
| **June** | 1.27 | 1.04 | 0.9 | 0.56 | 50.39 | 0.23 | 30.96 | 0.4 |
| **July** | 1.4 | 1.15 | 0.67 | 0.45 | 37.67 | 0.25 | 24.96 | 0.52 |
| **August** | 1.59 | 1.19 | 0.87 | 0.58 | 45.35 | 0.4 | 35.69 | 0.42 |
| **September** | 1.36 | 1.19 | 0.62 | 0.42 | 34.15 | 0.17 | 16.01 | 0.47 |

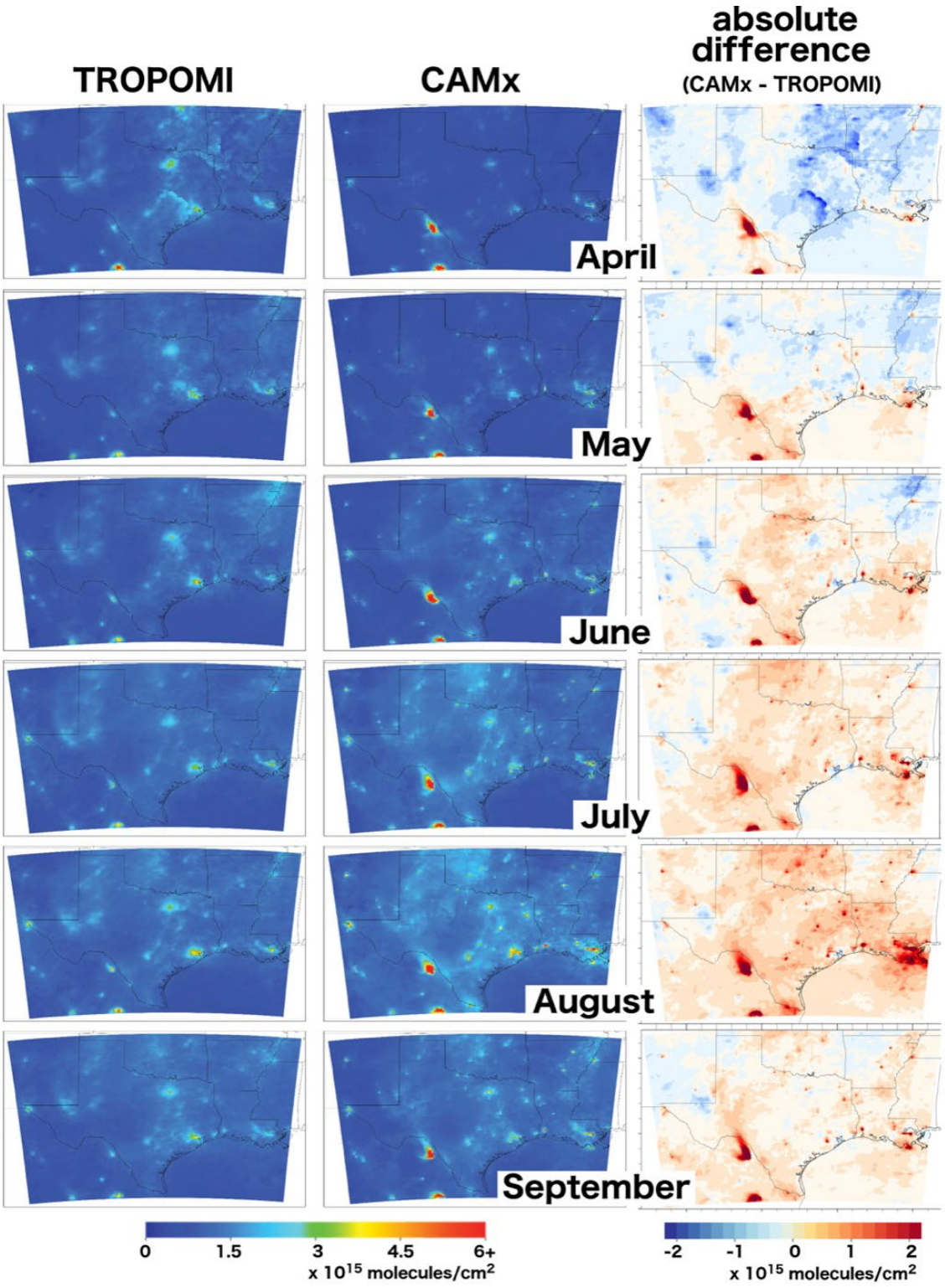



*Figure A-1 April – September monthly 2019 average $NO_2$ column amounts from TROPOMI as gridded with WHIPS (left) and CAMx with the averaging kernel applied (middle) for the 12 km domain. Absolute difference between CAMx and TROPOMI shown in right column.*

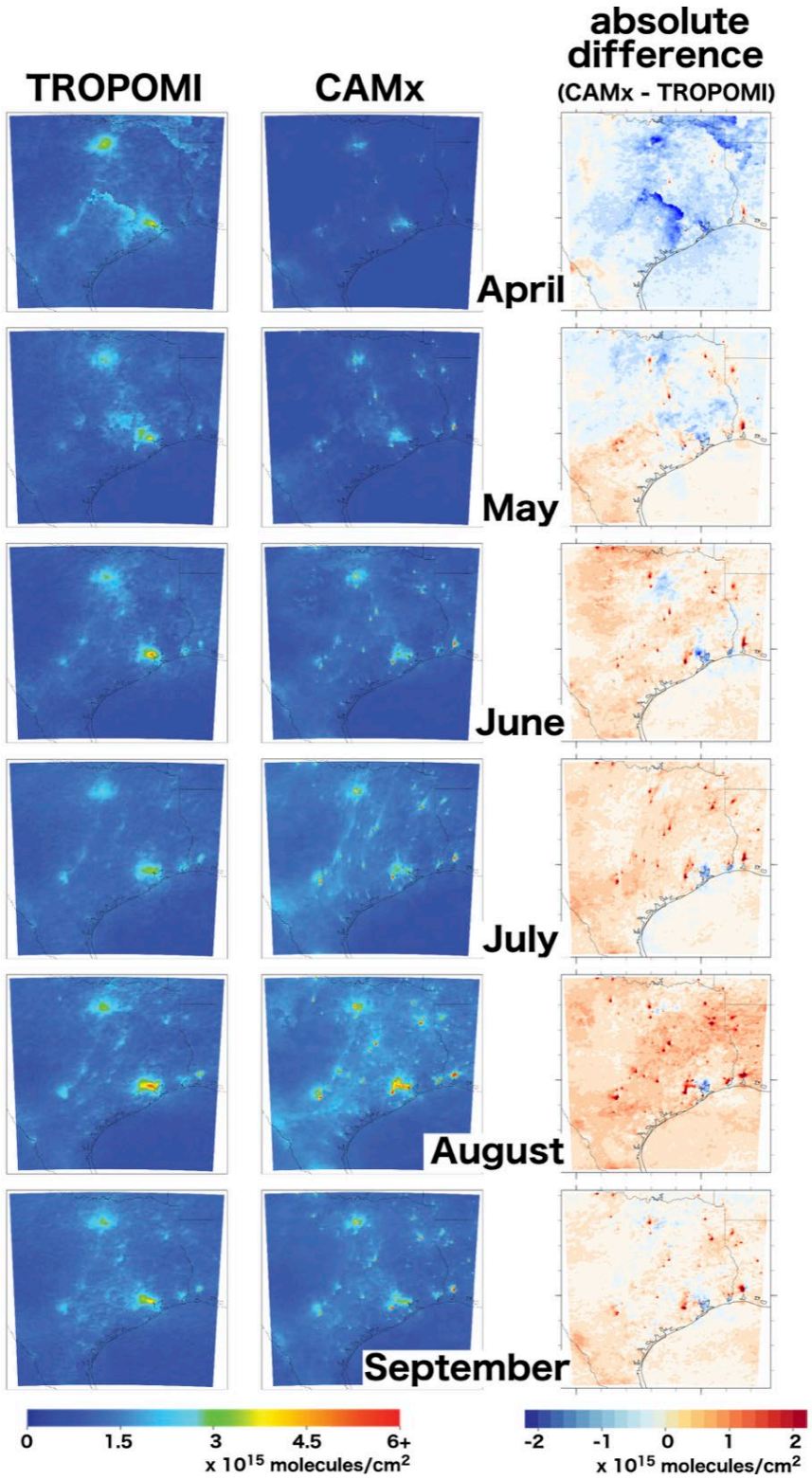



*Figure A-2 April – September monthly 2019 average $NO_2$ column amounts from TROPOMI as gridded with WHIPS (left) and CAMx with the averaging kernel applied (middle) for the 4 km domain. Absolute difference between CAMx and TROPOMI shown in right column.*

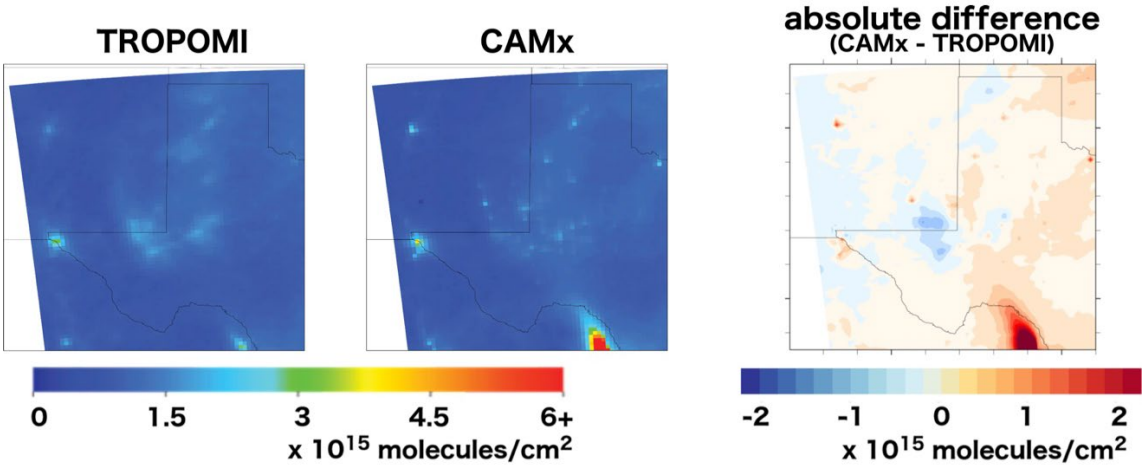



*Figure A-3 April – September average 2019 average $NO_2$ column amounts from TROPOMI as gridded with WHIPS (left) and CAMx with the averaging kernel applied (middle) for the 12 km domain over western Texas and New Mexico. Absolute difference between CAMx and TROPOMI shown in right column.*

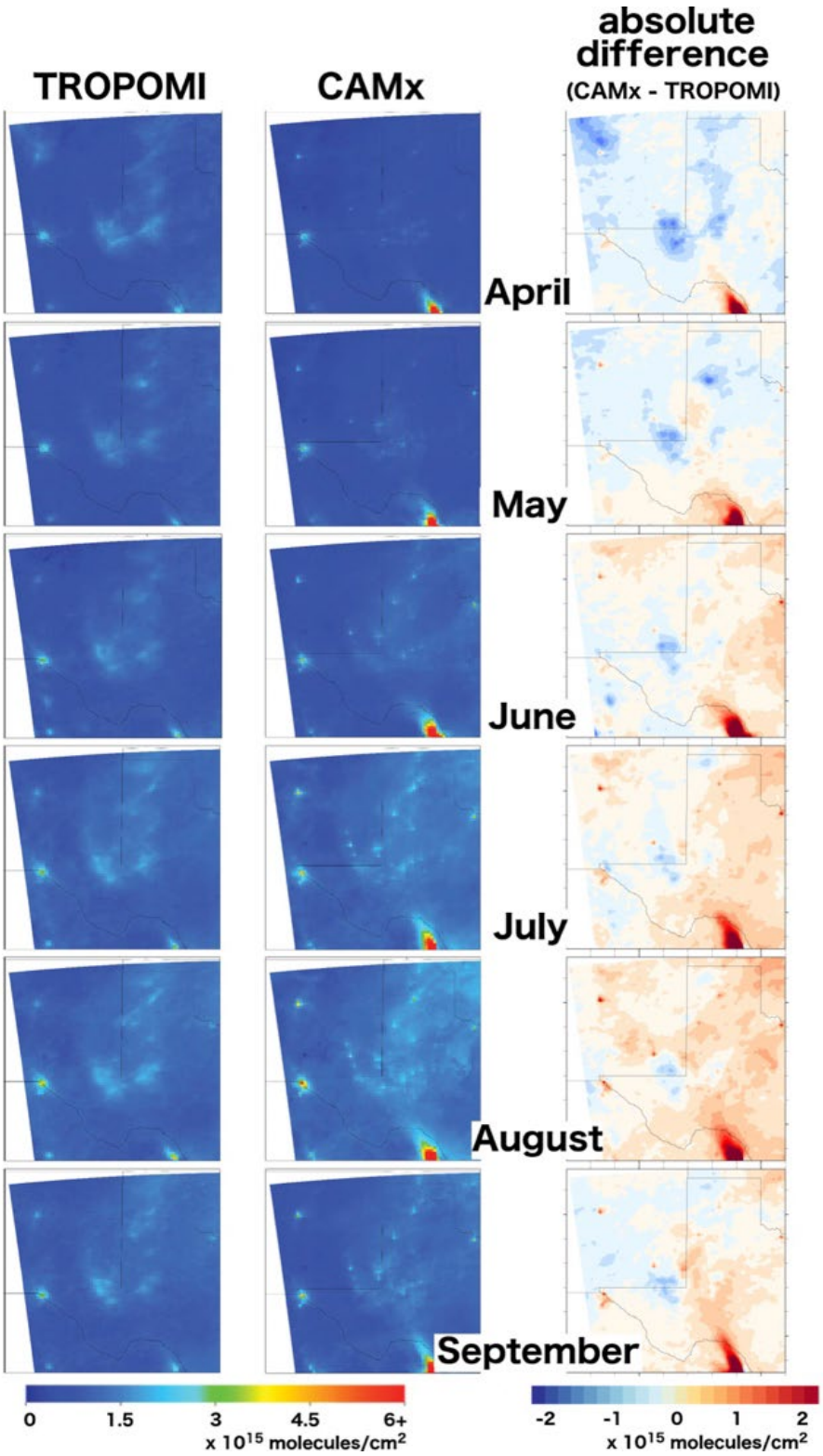



*Figure A-4 April – September monthly 2019 average $NO_2$ column amounts from TROPOMI as gridded with WHIPS (left) and CAMx with the averaging kernel applied (middle) for the 12 km domain over western Texas and New Mexico. Absolute difference between CAMx and TROPOMI shown in right column.*

## APPENDIX B

We had planned to expand our analysis to $SO_2$ to a limited extent as resource and data integrity allow. Our $NO_2$ analysis highlighted nuances in satellite retrieval algorithms and in processing model column estimates, and our team agreed that expanding our analysis to include $SO_2$ would not be simple, and would require resources and time beyond what is available. Furthermore, satellite capability to observe $SO_2$ is even more limited than $NO_2$ (and $NO_2$ was less robust than expected). Figure B-1 shows daily Level-2 TROPOMI $SO_2$ over the Oklaunion power plant, which may be compared with Figure 4-11 for the same area and time period. Where the power plant plume is evident in 43% of days for $NO_2$, but 0% of days for $SO_2$. However, CAMx modelling still included $SO_2$ amounts, and here we present in Figures B-2 and B-3 the seasonal and monthly average CAMx column $SO_2$ as calculated without the TROPOMI $SO_2$ averaging kernel.

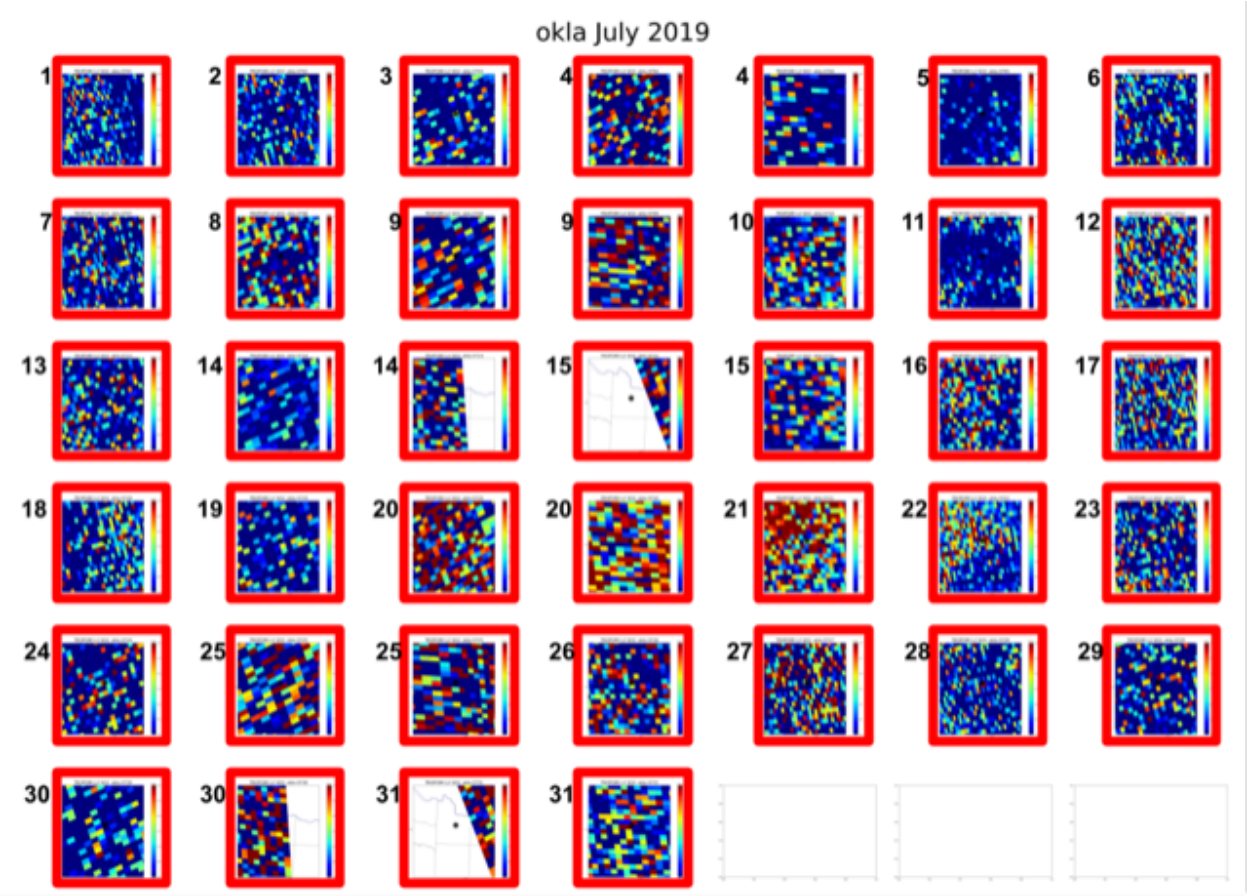



*Figure B-1 July 2019 daily Level-2 $SO_2$ column amounts from TROPOMI over the Oklaunion power plant. Compare with Figure 4-11 for $NO_2$.*

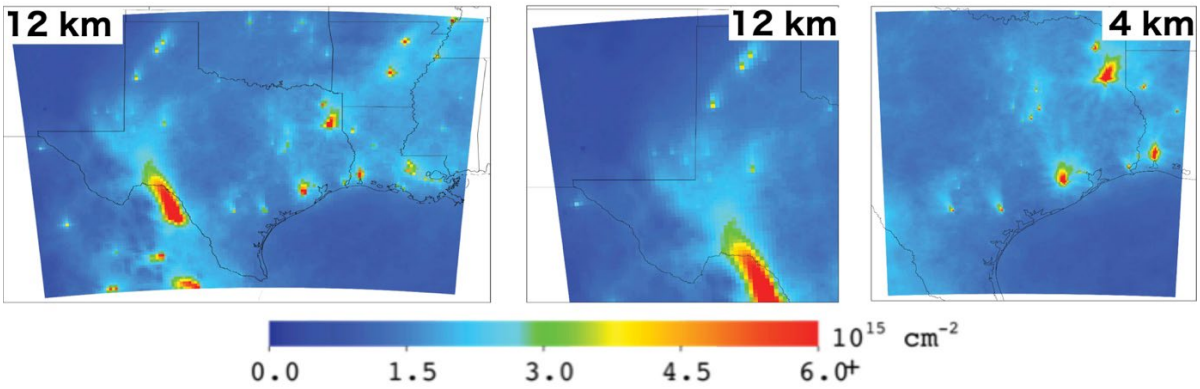



*Figure B-2 April – September average 2019 average $SO_2$ column amounts from CAMx for the 12km domain (left), the 12km domain zoomed in to western Texas (middle), and for the 4 km domain (right).*

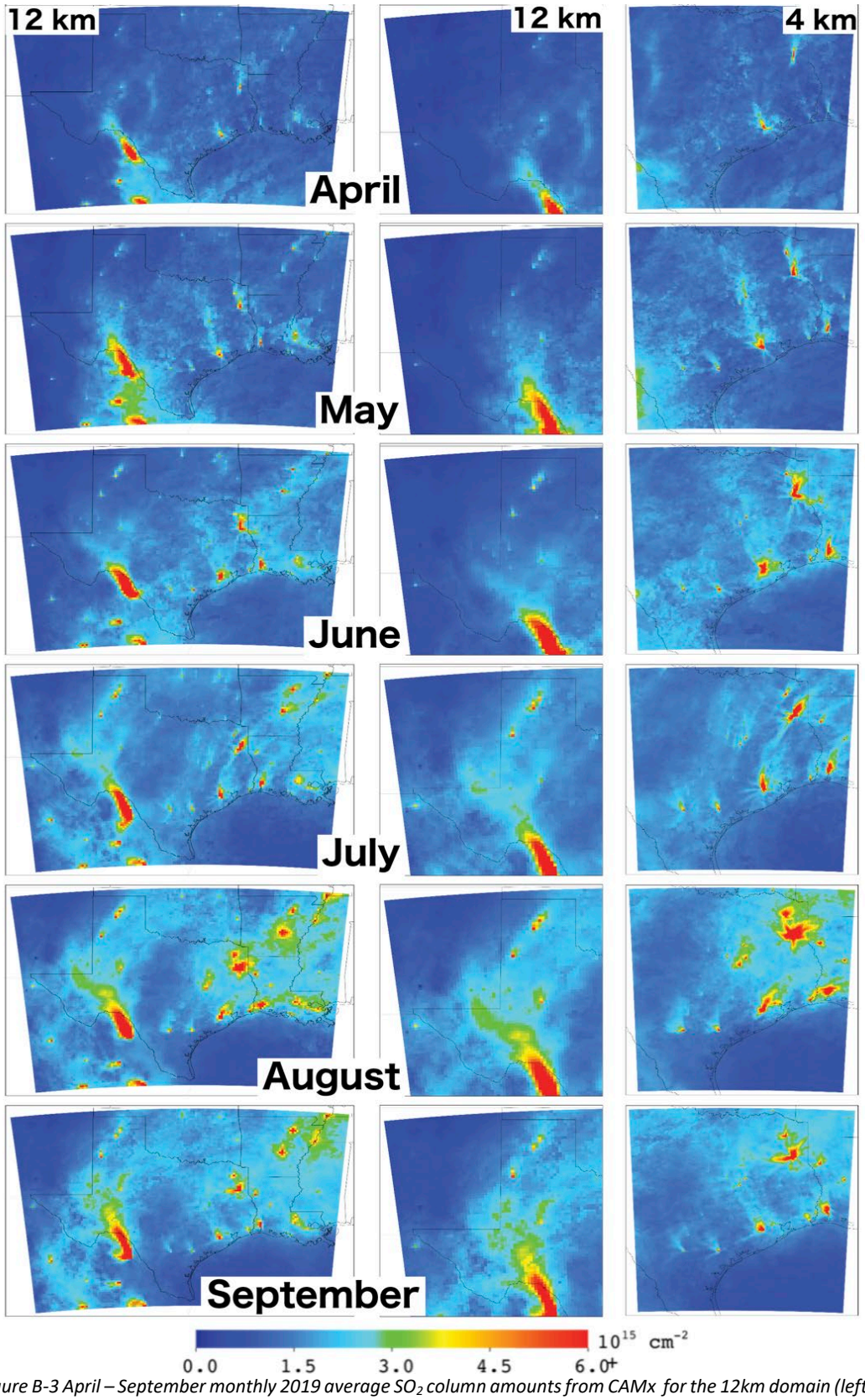



*Figure B-3 April – September monthly 2019 average $SO_2$ column amounts from CAMx for the 12km domain (left), the 12km domain zoomed in to western Texas (middle), and for the 4 km domain (right).*