Reviewer 2:

This study evaluates NOx emissions and ozone production in Texas, which covers several interesting questions, including model-satellite comparison, top-down NOx emission estimates using EMG approach, and HCHO/NO2 for identifying the ozone production regimes. Overall the results are clearly presented, and the figures are informative.

Thank you for the careful review. Please see our responses below in red. All new text added to the manuscript is described in detail in this document and has been *italicized*.

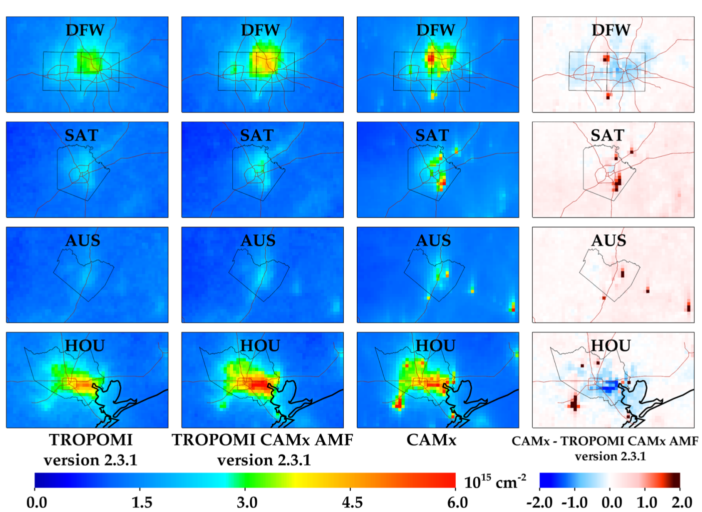
There are several interesting findings, which could well be three independent studies, but my concern is that the findings are a little disconnected, and it is not easy to tell the major take-aways from the study. I'd suggest the authors better frame the core questions, and show how each section is connected.

We have shortened the Abstract and made it clearer. We have also re-framed sections of the Conclusions to better describe the take-away points.

Below are some major comments:

1. It is not clear to me why the authors compare two different versions of TROPOMI NO2. Satellite retrieval products are updated routinely. The newest version should certainly be better, and the older version is already replaced with newer one. I don't think there is much scientific value of such comparison. Evaluation of different versions should be in the technical document of TROPOMI NO2. I'd suggest the authors stick to the newest, publicly available version.

In this revised draft, we have removed mentions of the Version 1.3 algorithm in Sections 3.4 (including Figures 6 & 7) and 4.1, as we agree that it was unnecessary to discuss. Please see new versions of Figures 6 & 7 below:



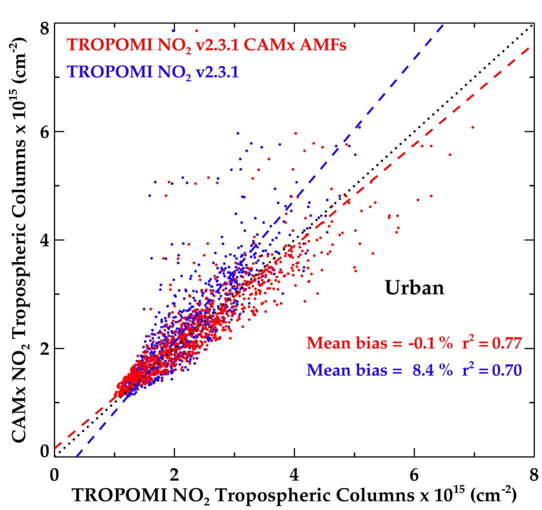


Figure 6. NO_2 tropospheric vertical column amounts averaged across April through September 2019 from TROPOMI, TROPOMI v2.3.1, TROPOMI v2.3.1 with new AMF, and CAMx for the largest four cities (Dallas, San Antonio, Austin and Houston). (Right) Scatterplot showing slope and correlation of various TROPOMI configurations and CAMx

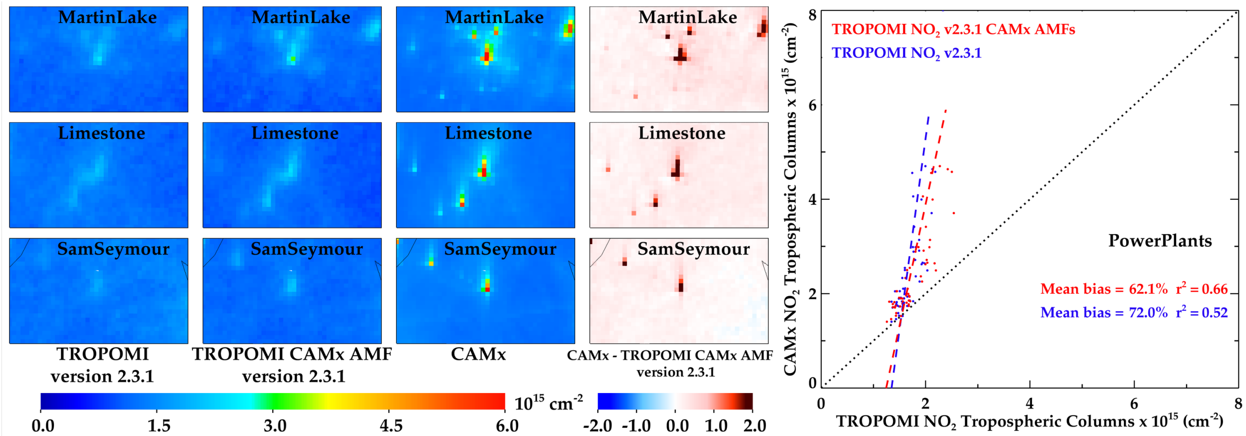


Figure 7. NO_2 tropospheric vertical column amounts averaged across April through September 2019 from TROPOMI v2.3.1, TROPOMI v2.3.1 with new AMF, and CAMx for the largest three power plants in East Texas (Martin Lake [Lat: 32.25 ° N, Lon: 94.58° W], Limestone [Lat: 31.42° N, Lon: 96.25° W], and Sam Seymour [Lat: 29.92° N, Lon: 96.75° W]). (Right) Scatterplot showing slope and correlation of various TROPOMI configurations and CAMx

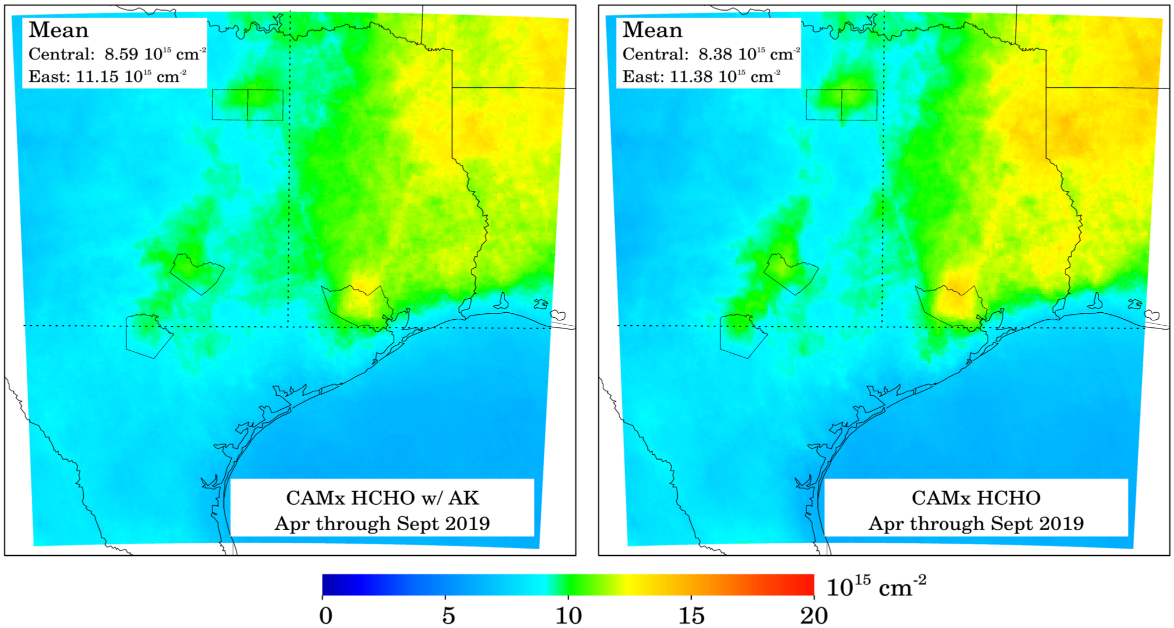
However, we decided to keep the version 1.3 algorithm versus version 2.3.1 comparison in Section 3.1 because there is value in doing a US-specific comparison between algorithms which is not covered in either technical document by developers. Inclusion of such a comparison is important because many users are most familiar with prior versions of the algorithms and have been wondering how the algorithms compare on a regional basis, rather than a global basis. We show evidence that regionally, the changes match global changes; a simple but important confirmation.

2. Related to previous question, I'd suggest the authors use newest version TROPOMI HCHO and NO2 data in Section 4.2 to evaluate ozone sensitivity. The authors have shown better performance of v2, but switch to v1 in 4.2. It is also not clear to me why the authors use an outdated version of TROPOMI HCHO (v1.1) while a newer version is already available.

A new HCHO product is not available for the timeframe in which we are investigating. The Version 2 algorithm only has data between July 13, 2020 and present, and not during 2019.

3. It's also not clear to me why satellite HCHO is less sensitive to the application of AK. While it may not help improve the spatial patterns of HCHO, application of AK may resolve the overall difference between modeled and satellite HCHO. It'd be great if authors can show a figure or two to illustrate this point.

We have now added an additional panel on to Figure 10 to show the effect of the averaging kernel. The averaging kernel affects the column HCHO amounts by ±2.5% for areawide averages. We have now added the following text: *"The difference between CAMx and CAMx with the averaging kernel applied is ±2.5% for areawide averages."*



Specific comments:

1. Figure 3 (right): What does the red line mean? Which shape profile?

The red line is the shape profile. It represents how much of the partial column is at any given altitude. The shape of the line is equivalent to the NO_2 mixing ratio show in the left panel, but normalized to a unitless quantity that integrates to unity over the depth of the troposphere. The profile on the right looks different because it is on a linear scale while the left panel is on a log-scale. We have now added the following clarifying text:

"In the right panels of Figure 3, we show *the modeled shape profiles – the NO_2 vertical distribution normalized to a unitless quantity that integrates to unity over the depth of the troposphere…*"

2. Figure 4: There is no explanation of the right figure in the caption. Why is it continuous? Are these generated for selected pixels?

Thank you for catching this oversight. We have now clarified the right hand panel with the following text:

"*The fraction of the NO_2 column attributed to different layers of the atmosphere (below 2 km, above 2 km (attributed to Other), and above 2 km attributed to lightning NO_X (LNOx)) at six locations (Gulf of Mexico, rural Central Texas, Austin, San Antonio, Dallas and Houston); the fraction attributed to lightning NO_X (LNOx) is calculated as the NO_2 addition between the two simulations without and with lightning NO_X emissions."*

3. Page 16 Lines 5 to 10: The authors show TROPOMI NO2 is significantly lower than modeled NO2, and they attribute this to issues with TROPOMI. While it may be true, but could this also be due to model unable to capture the sub-grid processes?

Thank you for the suggestion to further probe sub-grid processes, and the one most relevant here is the NOx/NO2 ratio. We have now probed the NOx/NO2 ratio using our model simulation, and included this information as Appendix C. We calculate NOx/NO2 ratios for the partial columns below 2 km and show a spatial image of the quantity at 1 PM – the TROPOMI overpass time. The NOx/NO2 ratio above 2 km is inappropriate for use in the EMG method since the column above 2 km represents "background conditions" and is subtracted out when using the EMG method. We've now added a discussion in Appendix C on this topic:

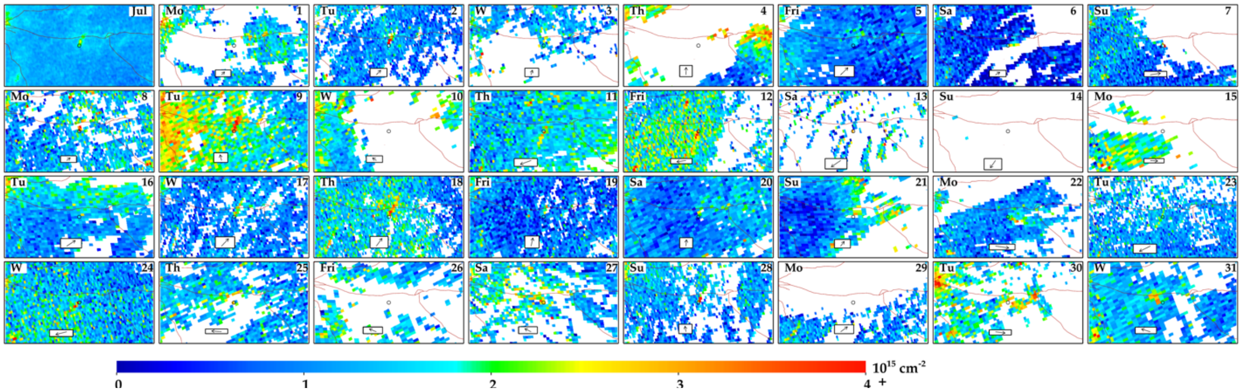
"the NO_X/NO_2 ratio can vary more substantially near large point sources. In the grid cells of the large point source itself, the NO_X/NO_2 ratio can be as large as 1.52. It is possible that the NO_x/NO_2 ratio in the model may be underestimated due to the emissions being equally spread out across the 4 km grid cell. NO_x/NO_2 ratios can be as large as 2 within 100 m downwind of major NO_X sources, especially under low ozone (< 30 ppb) conditions (Kimbrough et al., 2017). However, further downwind (>4 km) of these large point sources, the NO_X/NO_2 ratio quickly converges back to a value of ~1.31."

In addition to Appendix C, in the abstract we now state:

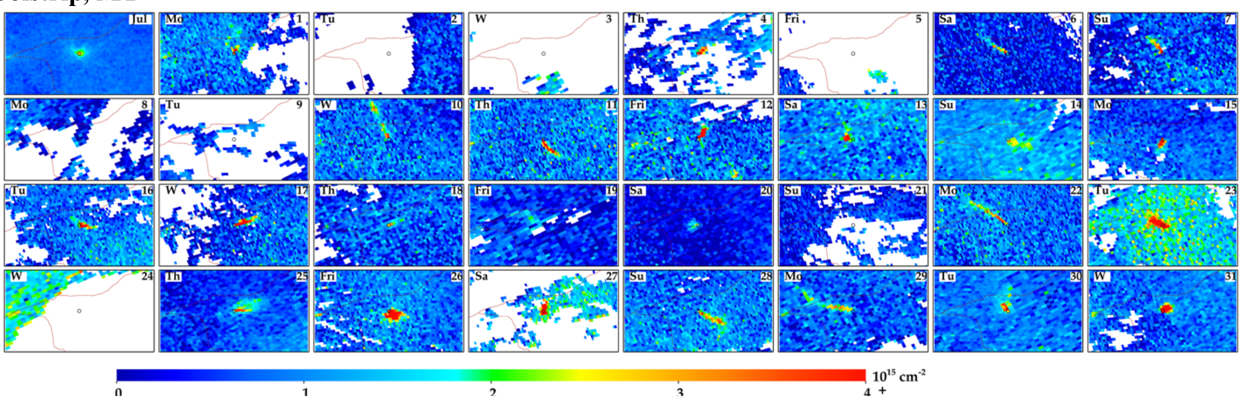
"Secondarily, the NO_X/NO_2 ratio in the model [near power plants] may be underestimated due to the 4 km grid cell size."

With that said, in Appendix D, we now show substantial evidence that the cause of the TROPOMI low bias near power plants is related to TROPOMI not capturing the NOx plumes on a day-to-day basis. TROPOMI is having difficulty distinguishing NO2 attributed to power plants from the background NO2 concentrations especially in areas with atmospheric conditions that cause short NO2 lifetimes (such as Texas). The daily images at the Martin Lake Power Plant, TX is shown below and in Appendix D. During roughly half of cloud-free days in July 2019 can a NO2 plume been easily observed from this power plant. In contrast, for the Colstrip Power Plant in Montana – a power plant with similar magnitude of NOx emissions – a plume can be easily discerned on all cloud-free days.

Martin Lake, TX



Colstrip, MT



We do also want to point out that there have been significant declines in NOx emissions from the power sector since the de Foy et al. (2015) study, which analyzed emissions in the 2005 – 2011 timeframe. NOx emissions from the power sector (https://ampd.epa.gov/ampd/) have declined by ~76% between 2005 (3.63 million tons) and 2019 (0.86 million tons). It is feasible that if TROPOMI were observing the same power plant 17 years ago, we might not have the same issue.

We have added the following text in Section 3.4:

"The reason for the substantial disagreement in Texas is still unknown, but we do not believe this repudiates the prior evaluation for urban areas. *NO_X emissions from the power sector in the U.S. have declined by 76% between 2005 (3.63 million tons) and 2019 (0.86 million tons) (https://ampd.epa.gov/ampd/). At these lower emission rates, it appears that TROPOMI is having difficulty distinguishing NO_2 attributed to power plants from the background NO_2 concentrations especially in areas, such as Texas, with atmospheric conditions that cause short NO_2 lifetimes – rapid plume dilution, high oxidation capacity due to large amounts of VOCs and water vapor, and high solar elevation angles. Secondarily, the NO_X/NO_2 ratio in the model may be underestimated due to the 4 km grid cell size (Appendix C).* The two power plants in New Mexico and Montana are located in areas with *smaller background NO_2*, lighter wind speeds, less VOCs and water vapor, and higher elevations; all of these factors cause the satellite sensor to be more sensitive to the NO_X emissions. *TROPOMI does not have* *the same difficulty over urban areas because the larger aggregated NO_X emissions are more easily distinguishable from background concentrations. Please see Appendix D for a discussion on this topic.*

4. Page 20 Line 12: This sentence is confusing.

We have modified to: "*We first compare* column HCHO comparison between CAMx and TROPOMI."

5. Table 3 shows the derived NOx emissions are sensitive to NO2 lifetime. How much confidence do you have with the NOx lifetime in EMG approach? Does it agree with the NO2 lifetime simulated from model? And how does the seasonal variation in NO2 lifetime affect the calculation of emissions?

We acknowledge the "seesaw relationship" between the NO_2 lifetime and NO_X emissions, and we make every effort to ensure that the EMG method is outputting an effective NO_2 lifetime that is in the realm of possibilities. When we visually inspect and see a NO2 plume dissipating rapidly, we expect the EMG fit to output a short effective NO_2 lifetime; in this instance a long NO_2 lifetime (~6 h) would not be appropriate. We've added the following text to Section 4.1:

"Most top-down methods fit both the effective NO_2 lifetime and NO_X emissions simultaneously and therefore have a "seesaw relationship" (Liu et al., 2022)– as lifetime increases, NO_X emissions decrease given a constant NO_2 burden. Here, we visually inspect the plume to ensure that the NO_2 effective lifetime is reasonable given the plume decay before proceeding."

When the EMG fit does not show an appropriate NO_2 lifetime, we do not report the results; this occurred for San Antonio and Houston, and is why EMG results for these cities are not included, which we noted in the text in Section 4.1: "The technique was applied to other urban areas, but those cities have large point sources at the periphery of the urban areas which adversely affected the calculation of the effective NO_2 lifetime needed to calculate the NO_X emissions."

Doing a rigorous analysis of why the NO2 effective lifetime varies by location and by platform (model vs. satellite) is beyond the scope of this project.

Through unpublished work (that is beyond the scope of this project), we have found that during winter, the NO_2 dispersion lifetime is dominant, and therefore makes this method much more difficult to apply. In this study, we only focus on the warm season months (April – Sept) when the chemical lifetime is shorter.

6. Figure 12: It's interesting to see there is almost no diurnal cycle with HCHO. I think both biogenic emissions and photolysis rate vary diurnally. It'd be interesting to illustrate why HCHO is relatively constant, and what this would mean for ozone production.

We agree that this is interesting, but an investigation of the diurnal patterns of HCHO is beyond the scope of this project, particularly because this would involve vertically-resolved HCHO measurements (which are sparse) and disentangling the complex relationship between HCHO and VOC emissions. We refer you (and the reader) to Schroeder et al. (2016) and Schwantes et al., (2022), which we now cite in this location:

"HCHO has broad peak in the afternoon, which is likely related to biogenic emissions and secondary formation. However, the HCHO diurnal cycle is flatter than we expected; this may be due to model difficulties in representing complex VOC chemistry for secondary HCHO production (Schroeder et al., 2016; Schwantes et al., 2022)."

7. Figure A1: There seems to be a large discrepancy between CAMx and observed NOx and NOy. I'd suggest the authors investigate why CAMx is biased too low, and how this would affect the interpretation of other findings.

TCEQ NO_2 ground monitors are disproportionately located near or directly adjacent to major highways; often within 1 km but sometimes within 50 m. While this is appropriate from a regulatory standpoint, it is often inappropriate to compare these monitors to a model simulation with 4 x 4 km^2 resolution. Therefore, a low bias in the model as compared to the monitors is primarily indicative of the sub-grid processes that CAMx cannot resolve. This is why a satellite comparison to the model is more appropriate than a ground monitor comparison. We now cite Souri et al. 2022 which discusses how care is needed when comparing pointwise measurements to concentrations spatially averaged over larger (>1 km) grid cells:

"As discussed in (Souri et al., 2022), care is needed when comparing pointwise measurements to concentrations spatially averaged over large (>1km) grid cells."