Response to review 2

We have responded to each comment below. Our replies are in blue, and the revised manuscript text is written in bold.

This manuscript has presented top-down estimates of global NO_x emissions using two OMI satellite NO_2 products over 2005-2016 and using the GEOS-Chem adjoint inversion method. Considerable differences are found between the two top-down emission estimates. Implementing the top-down NO_x emissions to the GEOS-Chem atmospheric chemistry model shows some improvements on the model simulation of tropospheric ozone. The study also points out that model improvements largely depend on the top-down emissions, the ozone metrics used, and model versions. The manuscript is in general well organized and meets the scope of ACP. One main concern is that the manuscript has been presented as a model evaluation paper that comparing several model simulations with different NO_x emissions with surface and sonde ozone measurements. It lacks some analyses in depth to understand the driving factors of the differences. The key findings of this study are also not clear. Do we have a better understanding of the NO_x emission trends as constrained by the satellite measurements, or how NO_x emission changes affect tropospheric ozone? I think the concern and the following specific comments should be addressed before considering publish.

We appreciate the comments from the reviewer. We added the following sentences in the abstract to address the concern on the NO_x emission trends:

"Posterior NO_x emissions show consistent trend over China, US, India, and Mexico constrained by the two retrievals. Emission trends are less robust over South America, Australia, Western Europe and Africa, where the two retrievals show less consistency."

Limited by the availability of surface measurements, we cannot claim that NO_x emission trends are improved everywhere. However, we demonstrate in this study that there are several regions where top-down NO_x emission trends are consistent across different retrievals and we are more confident about these.

The impact of NO_x emission on ozone simulations have spatial heterogeneity due to the nonlinear response of ozone to NO_x and our different understanding of local sources, physics, and chemistry. So, there is no generalized conclusion at global scale. We added the following sentences to the abstract to summarize our findings from this work:

"The performance of posterior ozone simulations is spatially heterogeneous from region to region. On a global scale, ozone simulations using NASA-based emissions remove the double peak in the prior simulation of global ozone. The higher abundances of NO_2 from the DOMINO posterior increase the global background ozone concentrations and therefore reduce the negative biases more than the NASA posterior in the GEOS-Chem v12 simulations at remote sites. Compared to surface ozone measurements, posterior simulations have more consistent magnitude and interannual variations than the prior, but the performance from the NASA-based and DOMINO-based emissions varies across ozone metrics. The current hard-constraints on NO_x diurnal variations and limited availability of

remote sensing data hinder improvement of ozone diurnal variations from the assimilation, and therefore have mixed performance on improving different ozone metrics. Additional improvements in posterior NO_2 and ozone simulations require more precise and consistent NO_2 retrieval products, more accurate diurnal variations of NO_x and VOC emissions, and improved simulations of ozone chemistry and depositions."

Specific comments:

1) Page 1, Line 24-25 in the Abstract: The statement "using NO_x emission datasets that have the best performance . . ." is not clear. As ozone simulation is affected by many other factors, the NO_x emissions that have the best performance on ozone simulation may not be the correct one. Some results in this study also showed that satellite constrained NO_x emissions did not necessarily improve ozone simulation (e.g., China daytime surface ozone in Figure 5)

We have revised the abstract, see response above.

2) Page 3, Section 2.1: What was the spin-up time for the model simulations? Were you using the same initial conditions? Please clarify.

The initial conditions are different for each NO_x emission datasets. We added the following sentence to the last paragraph of Section 2.1:

"For each NO_x emission dataset, the model is spun-up for 6 months, starting from July 2005. Therefore, we derive NO_x emissions from 2005, but only evaluate simulations with measurements from 2006."

3) Page 6, Line 179: Should here "the average of GEOS-Chem simulated NO_2 column density" be OMI observed NO_2 column density over 2x2.5 grid cell? Here you are generating pseudo measurements in the statement. The ratio should be calculated by OMI observations to avoid the OMI vs. model biases.

Thanks for pointing this out. We are calculating in the way the reviewer suggested, but did not describe it correctly. We changed the sentence to:

"…by the ratio of OMI NO_2 column density gridded at 0.1° × 0.1° to the **OMI NO_2 column density gridded at** 2° × 2.5° grid cell"

4) Page 8, Line 240-245: The large differences in seasonal variations of DOMINO and NASA posterior NO_x emissions seem interesting. Here you explained that the DOMINO posterior may better constrain soil emissions. Do you have any evidence or support for that?

We changed the cited sentence to:

"The peak of the DOMINO posterior NO_x emissions in the United States and Mexico shifted earlier in the year to June and July compared to the prior and NASA posterior emissions, similar to the results from Miyazaki et al. [2017]. **The peak in DOMINO posterior emissions corresponds to the time of high soil NO_x emissions, which are reported to be underestimated in high-temperature agricultural systems in the bottom-up inventory [Oikawa et al., 2015; Miyazaki et al., 2017].**

5) Page 8, Line 250-256: Here you showed that prior simulated surface ozone concentrations had double maxima in April and August, and the posterior results partly corrected the biases. What cause the double maxima in the prior simulation? And how NO_x emission changes correct the August maximum? Please clarify.

We added the following sentences to the cited paragraph:

"The August ozone peak in the prior simulation is mainly due to the high ozone concentrations in North China, Southwest China, and North India. The NASA and DOMINO posterior simulations have both reduced surface ozone concentrations in North China Plain and Northeast China in August due to the larger posterior NO_x emissions than the prior in these high-NO_x regions. Both posterior ozone simulations are also smaller than the prior in Tibet and North India due to the reductions of posterior NO_x emissions in low-NO_x region. The August ozone peak in the DOMINO posterior comes from the higher ozone concentrations in Angola and Democratic Republic of the Congo compared to the NASA posterior and prior simulations in the same month and DOMINO posterior simulations in the previous months. This can be explained by the larger upward adjustment of DOMINO posterior NO_x emissions in South Africa in August. These results show the large spatial heterogeneities on the responses of ozone seasonality to the changes in NO_2 abundances on a global scale."

6) Page 9, Line 269-271: As indicated in Figure 6, interannual changes in the two posterior NO_x emissions in Australia over 2005-2016 are not that consistent. The DOMINO results show large reduction over 2006-2010 and then increase afterwards. Do you have any explanation why the two satellite products show different interannual variation and trends over some regions?

The different trends in posterior NO_x emissions are propagated from the different trend in NO_2 column densities retrieved from these two products, as shown in Figure R4. This could possibly be caused by the differences in scattering weight / averaging kernel, but it is hard for us to pinpoint what is the exact cause. We made the following modification to the cited sentence:

"In **Mexico**, the two posterior NO_x emissions consistently increased by 6% (NASA) and 13% (DOMINO) from 2005 to 2016. **The DOMINO posterior shows more obvious increase** in Mexico from 2010 to 2016. ... **In Australia, the NASA posterior increases by 10% from 2005 to 2016. In comparison, the DOMINO posterior decreases from 2005 to 2010 and increases afterwards, consistent with the posterior trend from Miyazaki et al. [2017]. The different trends in posterior NO_x emissions are propagated from the trends in the two OMI NO_2**

retrieval products. The discrepancies are likely due to the different surface albedo and cloud products used in the two retrievals, which affect averaging kernel sensitivities."

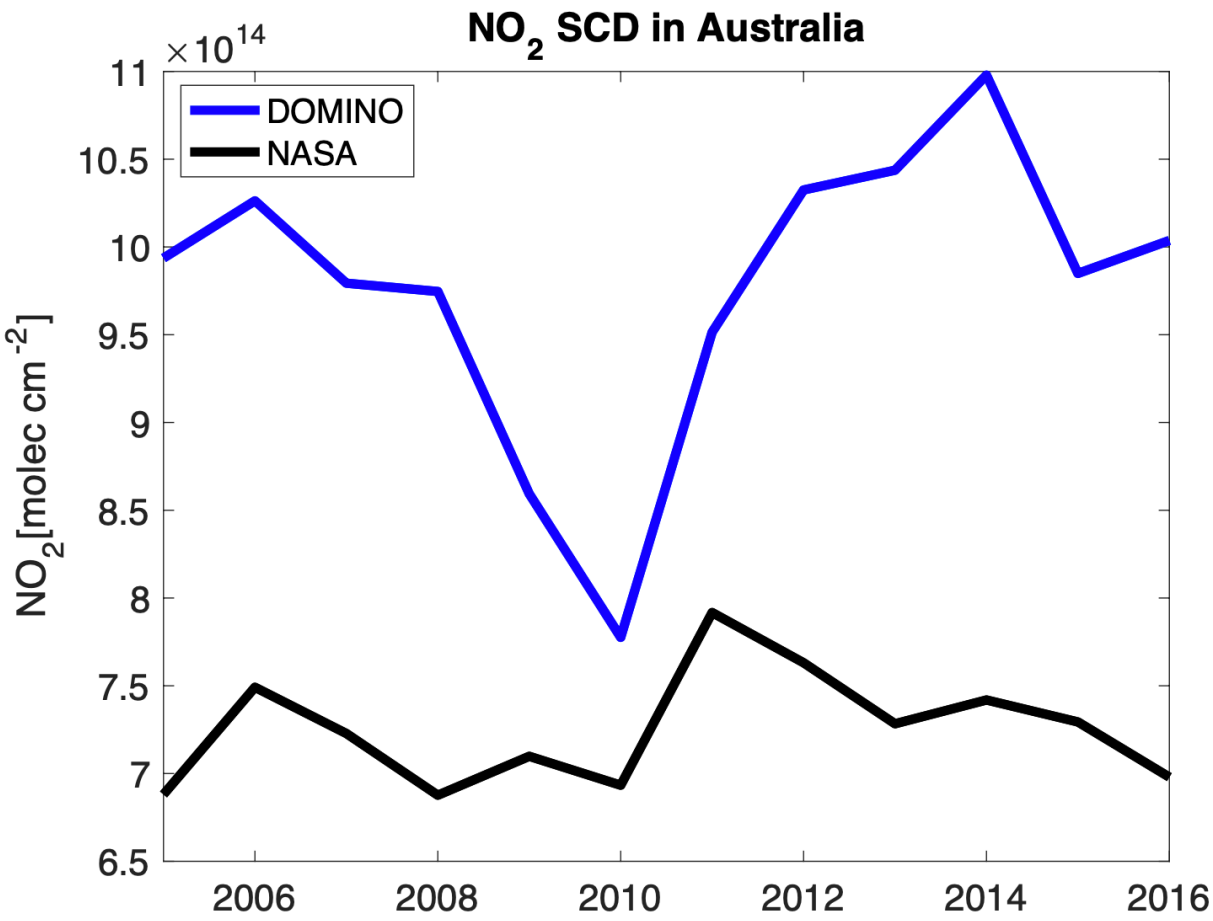

Figure R4. NO_2 column densities in Australia from OMI.

7) Page 10, Line 319: "Ozone measurements in 2014 decreased compared to the 2006 level in China, the US, South America and Mexico". I do not see from Figure 9 that in China ozone concentration in 2014 was lower than 2006.

Thanks for pointing this out. That statement comes from an earlier analysis that used all available TOAR sites at each year, not just sites that have continuous measurements throughout the years. We removed China and South America from that sentence.

8) Page 10, Line 314-316: How did you separate the ozone trends caused by NO_x emissions vs. meteorology? A description in the main text is needed. Also, you may calculate the meteorology (non-NO_x) effects using either GC-adj or GCv12 results? Which one did you use in Figure 9, and how they differed?

We added the following sentences to this paragraph:

"The second trend is calculated through simulations that use constant NO_x emissions throughout the studied years. It has a similar trend from GCv12 and GC-adj as shown in the green lines in Fig. 9. The trend caused by NO_x emissions is obtained by subtracting the second trend from the ozone trend simulated using NO_x emissions at each corresponding year."

We also added dotted green lines in Fig. 9 to separately show simulated trend from non-NO_x sources from GCv12 and GCadj.

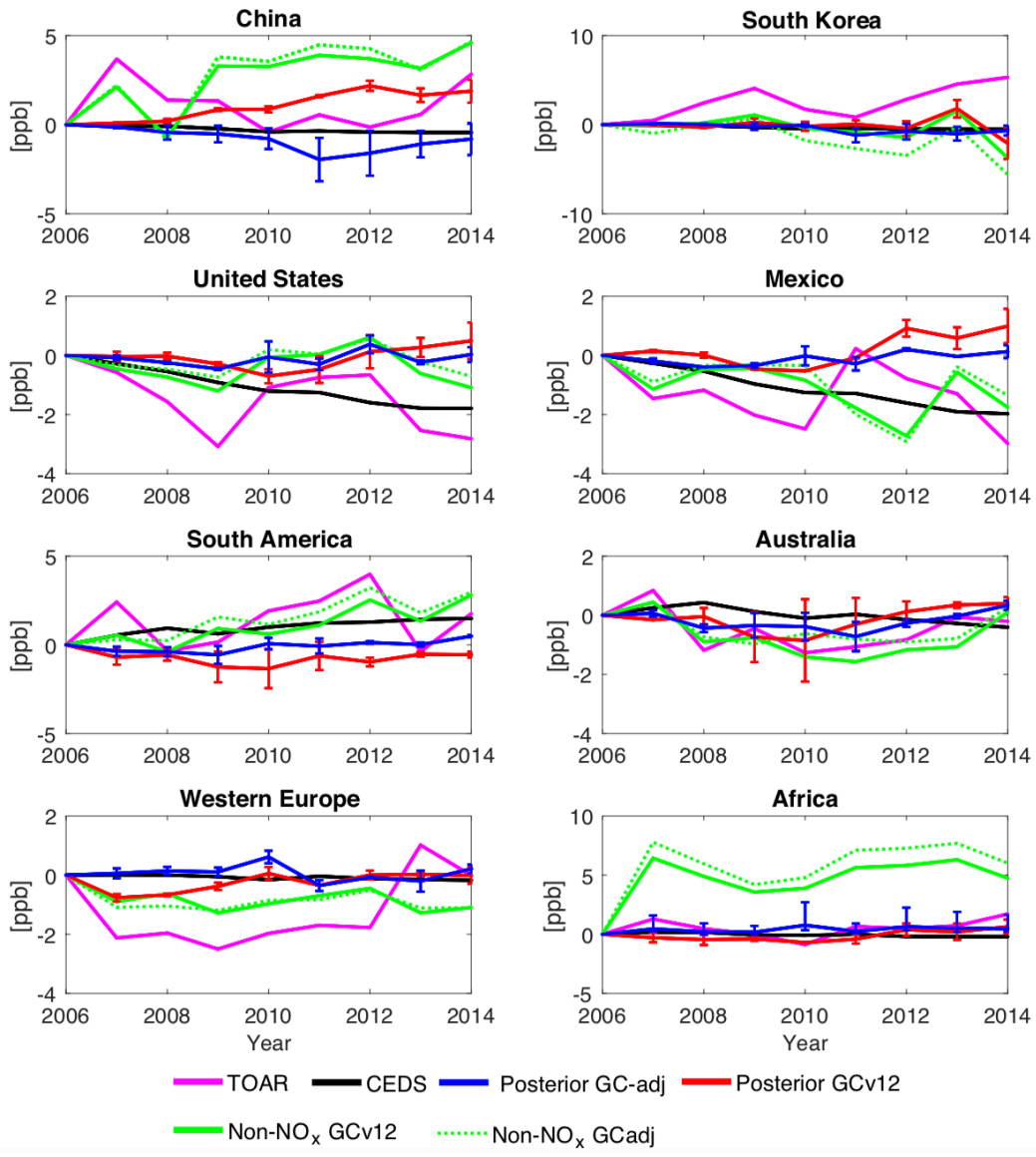

Figure 9. Changes of regional mean annual MDA8 ozone concentrations compared to 2006 from TOAR measurements (magenta line), due to changes in bottom-up NO_x emissions (black), due to changes in top-down NO_x emissions (blue lines for simulations from GC-adj and red lines for simulations from GCv12), and due to changes in meteorology and non-

NO_x emissions (green lines). Only sites that have continuous measurements throughout the 9 years are included. The vertical bars represent the spread of changes from simulations using the NASA and the DOMINO posterior NO_x emissions. The impact of meteorology and natural sources are removed from black, blue and red lines by subtracting simulations using 2010 bottom-up anthropogenic emissions for all years from simulations that use bottom-up NO_x emissions corresponding to each year.

9) Page 11, Line 338: It is surprising that the model versions (GCadj and GCv12) simulate very different ozone vertical profiles. GCv12, which is a more updated version, has much large biases in the upper troposphere, in particular with the updated NO_x emissions. Can you explain why in GCv12 changes in surface NO_x emissions would lead to large ozone changes in the upper troposphere?

GCv12 includes halogen chemistry, which is not included in GCadj. This chemistry depletes ozone. Its impact is especially larger at locations away from NO_x sources, e.g., upper troposphere, leading to much lower ozone concentrations in the GCv12 simulations.

We modified the following sentences in the cited paragraph:

"The different biases in ozone simulations **close to surface** can be explained by the usage of different emission inventories (e.g., different biogenic emissions) **and** different boundary layer mixing scheme (non-local mixing [Lin and McElroy, 2010] in GCv12 and full mixing in GCadj). **The different** chemical **mechanisms in the two model versions affect the different model biases especially in the upper troposphere**. For instance, …"

We also added the following sentences to the cited paragraph:

"The prior simulations in GCv12 applies NO_x emissions at different altitude, whereas the posterior GCv12 and all GC-adj simulations apply all NO_x emissions to the surface. This leads to different transport and formation of ozone at different model layers and therefore causes larger differences in ozone simulations in the upper troposphere."