Response to Reviewer #2's comments on He et al. 2019 Atmospheric Physics and Chemistry manuscript

We thank the anonymous reviewer for thoroughly reading our manuscript and providing helpful comments and suggestions, which lead a significant improvement of the manuscript. The detailed responses to comments are listed below (text in italic is the reviewer's comments, and the blue text highlighted is our response):
**All line numbers are from the clean version of revised manuscript.*
***Please see the revised manuscript in the attachment.*

This manuscript focuses on the comparison of ground/airborne measurements and satellite retrievals of a variety of air quality relevant trace gases with CMAQ model products. The authors attempt to adjust emissions in CMAQ based on column model/satellite ratios, thus improving the modelled ozone agreement with ground/airborne measurements. The dataset described in the manuscript is certainly novel, but there are enough omissions in the analysis that would prohibit its publication in its current state. I recommend the manuscript to be rejected, while encouraging the authors to resubmit after the address of the below issues.

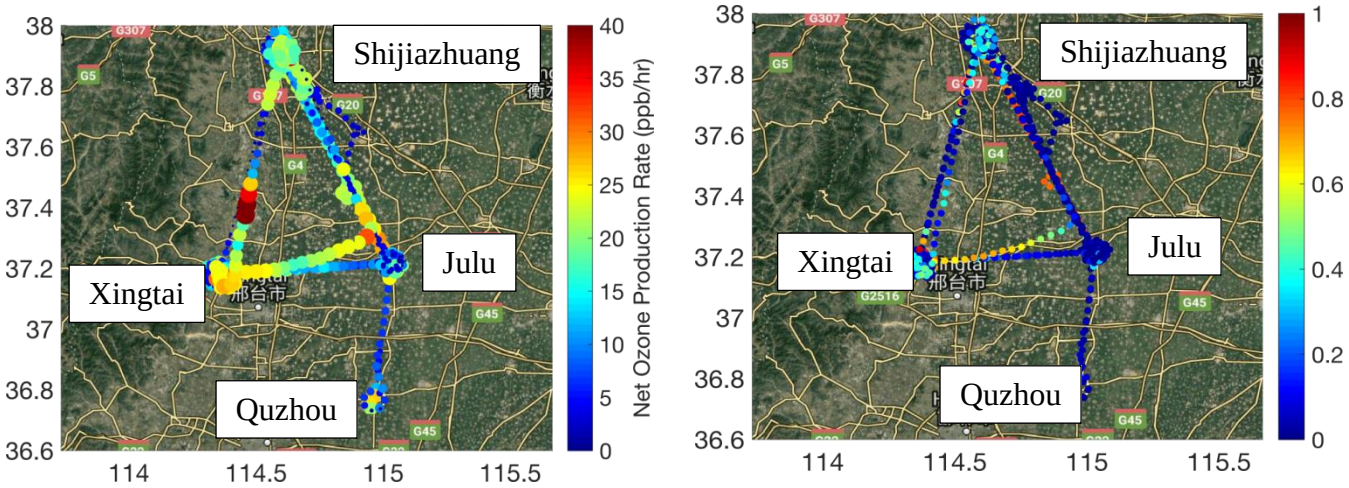
Response: We appreciate the comments from the anonymous reviewer. We have significantly revised the manuscript to improve the data analysis and the discussion according to these comments as discussed below.

Major comments:
-The improvement in ozone agreement is quite modest, to the point I am interested as to how significant it is. How does this compare with the various uncertainties in measurements and the EI? What ratios would be required to force greater agreement? It seems there would be many more sensitivity studies needed to try to evaluate this. As is, this suggests to me that the ozone discrepancy cannot be explained by a simpler large scale emissions adjustment and resolving it would require a more intensive study of the VOC chemistry.

Response: We agree that moderate agreement was achieved even after adjusting the emissions based on satellite products. There are some major reasons discussed in the original manuscript. For instance, we could only access to surface observations at one supersite which was significantly influenced by the nearby sources and local topography (the site was located on a small hill near mountains and downwind of Xingtai with ~1 million population and heavy industries; we added Figure S4 in the supplementary material to show the location). These fine characteristics cannot be resolved by our 12 km CMAQ simulations. Also, the latest available anthropogenic emission inventory was developed for 2010, while the intensity and distribution of emissions in China had changed drastically between 2010 and 2016 with migration of industrial sectors to less developed regions and strict regulations on air quality. Our comparison of NO_2 columns from CMAQ and OMI supported our hypothesis that the NO_2 column over highly developed regions such as Beijing, Yangtze River Delta, and Pearl River Delta decreased, while NO_2 column over less developed regions increased. For VOC emissions, the EDGAR inventory only included the total emissions instead of detailed emission rates for each VOC species or group, so our approach using column HCHO as a proxy to scale VOCs emissions could

introduce some uncertainties. The VOC chemistry of ozone photochemical production over the NCP has been thoroughly studied in a companion paper (Benish et al., JGR-Atmosphere, under review) using airborne VOC measurements and box modeling, which has confirmed a VOC-sensitive environment over the NCP (See Figure 9 adapted from the manuscript). The full manuscript is available upon request.

Figure 9. Left: Net production rates of O_3 calculated using box model results along the Y-12 flight track during ARIAs. The size of the dots is proportional to the production rate of O_3. Right: O_3 production sensitivity indicator, L_N/Q, along the Y-12 flight track during ARIAs. Ozone production is VOC-sensitive when $L_N/Q>0.5$ and NO_x-sensitive when $L_N/Q<0.5$.



In summary, we agree that one set of sensitivity experiments may not be sufficient to correct the discrepancy of emissions, but our study shows the correct direction to improve CMAQ performance qualitatively. To better explain our goal, we significantly revised the manuscript especially the abstract, introduction, and conclusion sections to emphasize the goal of our study. Please see details in the track-and-change version.

-A prevalent assertion of the authors is that discrepancies between model and observations are driven by reduced emissions, but the authors do not address any validation of the 2010 emissions. It is difficult to claim difference is due to emissions reduction without evidence that CMAQ and observations, airborne or satellite, agreed in the first place. Consistent, yes, but not necessarily suggestive or conclusive. The authors do not cite other studies to support this. Language should be adjusted throughout the manuscript to reflect that the adjustments needed to improve the model agreement are inherent, a component of which is consistent with emissions reduction.
Response: We appreciate the reviewer pointed out this problem. As discussed above, the 2010 EGDAR emission inventory is the latest dataset available to our group, even though China's emissions changed significantly between 2010 and 2016. Since we did not have observations and modeling results for 2010, we cannot validate the EDGAR emission inventory. Li et al. (2017a) discussed anthropogenic emissions in East China for the 2010 EGDAR inventory and reported the uncertainty of NO_x as ~20%. Li et al. (2017a) also stated "*Emission estimates from bottom-up inventories are uncertain due to lack of complete knowledge of human activities and emission from different sources. Uncertainty ranges of an emission inventory could be estimated using*

propagation of error or Monte Carlo approaches (e.g., Streets et al., 2003; Zhao et al., 2011). However, in a mosaic emission inventory like MIX, a normalized quantitative assessment of uncertainty ranges is difficult because detailed information for emission inventory development is not collected". So we assumed the EDGAR inventory as the true emissions for China in 2010, and treated the differences between modeling results and satellite products as the change of NO_x, VOCs, and CO between 2010 and 2016. We followed Canty et al. (2015) approach which investigated the NO_x emissions in the eastern United States using the 2011 NASA DISCOVER-AQ campaign data, CMAQ simulations, and OMI NO_2 products. Details about this methodology can be found in this paper.

Canty, T. P., Hembeck, L., Vinciguerra, T. P., Anderson, D. C., Goldberg, D. L., Carpenter, S. F., Allen, D. J., Loughner, C. P., Salawitch, R. J., and Dickerson, R. R.: Ozone and NOx chemistry in the eastern US: evaluation of CMAQ/CB05 with satellite (OMI) data, Atmospheric Chemistry and Physics, 15, 10965-10982, 10.5194/acp-15-10965-2015, 2015.

During the KORUS-AQ campaign, independent studies using satellite products revealed that NO_x and CO emissions were substantially underestimated in the Korean emission inventories, which are also based on MIX and EGDAR inventory in East Asia.

Miyazaki, K., Sekiya, T., Fu, D., Bowman, K. W., Kulawik, S. S., Sudo, K., Walker, T., Kanaya, Y., Takigawa, M., Ogochi, K., Eskes, H., Boersma, K. F., Thompson, A. M., Gaubert, B., Barre, J., and Emmons, L. K.: Balance of Emission and Dynamical Controls on Ozone During the Korea-United States Air Quality Campaign From Multiconstituent Satellite Data Assimilation, Journal of Geophysical Research: Atmospheres, 124, 387-413, 10.1029/2018jd028912, 2019.

Goldberg, D. L., Saide, P. E., Lamsal, L. N., de Foy, B., Lu, Z., Woo, J. H., Kim, Y., Kim, J., Gao, M., Carmichael, G., and Streets, D. G.: 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, 1801-1818, 10.5194/acp-19-1801-2019, 2019.

Therefore, even evaluating the emission inventory in China is beyond the scope of this study, we believe there could be significant discrepancy in the current inventory which could impact the CMAQ performance. To better explain it, we added the following sentence "*Independent studies using KORUS-AQ observations and satellite products suggested that major ozone precursor emissions such as CO and NO_x could have large discrepancies as compared with emission inventory in East Asia (Goldberg et al., 2019; Miyazaki et al., 2019)*" in Line 445 of the revised manuscript.

-Line 296-306: the EF technique is not very well described here. How were the cutoffs that the authors state (R^2, altitude, etc) justified? A fuller description of the technique would help significantly.

Response: The EF technique was developed by our colleagues at NASA Langley research center, and the paper was recently published in JGR-Atmosphere. We added this paper in our bibliography and the reference to this technique is:

Halliday, H. S., DiGangi, J. P., Choi, Y., Diskin, G. S., Pusede, S. E., Rana, M., Nowak, J. B., Knote, C., Ren, X., He, H., Dickerson, R. R., and Li, Z.: Using Short-Term CO/CO_2 Ratios to Assess Air Mass Differences over the Korean Peninsula during KORUS-AQ, Journal of Geophysical Research: Atmospheres, 0, 10.1029/2018jd029697, 2019.

Also we added the following sentences "*To evaluate the emission inventory data in the NCP, we used a plume recognition method to calculate the emission enhancements (EEs) from Y12*

observations. We first selected 60 1-s aircraft measurements with a 60-s moving window. Then we conducted linear regression of observed air pollutant (CO, NO_x, etc.) concentrations vs. CO_2 concentrations in each 60-s window and calculated the slope (i.e. $\Delta CO/\Delta CO_2$ and $\Delta NO_x/\Delta CO_2$) and correlation (R). The slope is defined as EEs in each window, standing for a 'plume' tested in the 60-s window. Lastly, we only selected EEs that are within the PBL (below 1.5 km AGL in this study) and statistically significant ($R^2 > 0.6$), so the values of these selected EEs can act as a proxy of EFs in the air mass observed" in Line 326 of the revised manuscript.

-Line 296-306: Since these come a variety of emission sources, a simple mean seems simplistic to describe the population. In looking at Fig 6, there seem to be multiple overlapping populations contributing to the mean. A broader discussion of these populations seems warranted.

Response: Thanks for raising this concern, which was also pointed out by another reviewer. The EDGAR emissions treated emissions such as NO_x and VOCs in four sectors: Energy, Industry, Transportation, and Residential. It is difficult for us to determine which sector contributed most to the uncertainty and discrepancy. Therefore, we used the mean of ratios like $\Delta CO/\Delta CO_2$ from EDGAR and observations to evaluate the possible cause of emission discrepancy. We added the following sentences "*The 2010 EDGAR inventory has emissions for 4 sectors: Energy, Industry, Transportation, and Residential. We calculated the CO/CO_2, NO_x/CO_2, and NO_x/CO ratios through averaging the EFs from these 4 sectors (Fig. 6).*" in Line 323 to better illustrate it.

-Line 277-293: Where do the Y12 NOy measurements come from? There is no description in the instrument section. Without the details of the measurement, this entire discussion of reactive nitrogen is not possible to evaluate.

Response: We are sorry for the missing information. We did have NO and NO_y measurements onboard of the Y-12 aircraft. NO was measured using a commercially available Thermo Environmental NO/NO_y analyzer, and we used a hot molybdenum converter (at 375 ^oC) next to the gas inlet to convert all reactive nitrogen to NO_y. However, due to high electricity demand from the converter and limited power supply on the airplane, we only turned on this instrument for few research flights such as 06/11/2016. Due to limited observations, we didn't compare NO_y measurements vs. CMAQ simulations. We added the description of NO_y measurements into the revise manuscript as "*Nitrogen oxide (NO) and reactive nitrogen compounds (NO_y) concentrations were analyzed using a commercial available NO analyzer (Model 42C, Thermo Environmental Instruments) with a hot molybdenum convertor working at 375 ^oC (Luke et al., 1992; Stehr et al., 2000). Ambient gas input was switched with and without the convertor frequently to measure NO and NO_y simultaneous. However, due to high power demand of the instrument and convertor, NO and NO_y were only measured during some research flights*" in Line 154 of the revised manuscript.

-Sect 3.3: It seems there is should be a before/after version of this analysis: one with the baseline CMAQ model (which is to my understanding what is used for this analysis) and with at least the CMAQ_all case. That seemed to be where the analysis was heading, and without it, the manuscript feels incomplete.

Response: This section included the evaluation of column NO_2, HCHO, and CO between satellite observations and CMAQ simulations, and we found discrepancies which could be caused by the emission change between 2010 and 2016. These results inspired us to develop the sensitivity experiments with adjusted emissions based on satellite products, but we did not

anticipate the adjustments would fully correct the emissions. We added a similar evaluation of CMAQ and satellite products in the supplementary material. In Line 459 of the revised manuscript, we added the following sentences "*We conducted a similar analysis of Figure 4 to investigate the discrepancy between satellite products and CMAQ simulations over the eastern China (Figure S7 in the supplementary material). CMAQ_all run successfully reproduced the column NO_2 and near surface CO in eastern China and column HCHO in the NCP. In southern China where biogenic VOCs dominate, adjusting anthropogenic VOCs emissions showed limited improvements in column HCHO simulations*".

Minor comments:
-Fig 1a probably unnecessary, label fig 2 more clearly
Response: We followed the suggestion and moved Figure 1a to the supplementary material and added the flight area in Figure 1 (previous Figure 2).

-Fig 3: change zeros to O
Response: We have corrected them in the figure.

-Fig S1 & S3: these have very similar captions, though the difference is clearer from the manuscript. I would edit the captions to highlight the differences
Response: We followed the suggestion and modified the caption of Figure S9 (previously S3 in the original manuscript) to highlight the difference.

-Line 251, "We observed isolated plumes: : :": Where are these plumes with respect to the surface layer structure? What about the secondary plumes at 800-1200 m? The authors need to add context on the atmospheric structure to describe the transport impact of these plumes. Perhaps a summary plot of potential temperature as well?
Response: To better explain it, we added one figure (Figure S2) in the supplementary material as an example of elevated plumes over Xingtai.

-Line 259, "In summary: : :": This sentence was somewhat confusing, is unclear whether authors mean exclusively east-west gradients observed or both north-south and east-west.
Response: Thanks for raising this question. We generally observed both east-west and north-south gradients in our campaign region. We revised the sentence as "*In summary, we found both south-north gradient and east-west gradient of air pollution in the campaign region, i.e., higher concentrations of air pollutants in the west XT-LC corridor near the mountain as compared with east side of JL and QZ, and higher concentrations in the north LC as compared with the south XT*" in Line 275 to better explain it.

-Fig 7: highlight here the aircraft campaign area, is difficult to eyeball where it is using fig 1b
Response: We followed the suggestion and add a red star showing the aircraft campaign area in Figure 7a.

-Fig 7 & 8: having both the differences and ratios seems redundant. As ratios are used in the later analysis, the authors should keep those figures and get rid of the difference figures, with some minor rewording in section 3.2.

Response: We followed the instruction and moved Figure 8 to the supplementary material as Figure S6.

-Fig. 9: it is not clear to me from the discussion the need for both panels A & B
Response: Figure 9a shows the $HCHO/NO_2$ ratio calculated from OMI observations, and Figure 9b has the ratios based on CMAQ simulations. They look similar, but demonstrate the ozone production sensitivity from two independent datasets.

-Fig 10: these are difficult to evaluate with the high variability. Perhaps a difference plot between observed and modeled, or ratios of the same, in order to clarify the improvement
Response: We appreciate the suggestion. However, even the CMAQ_all case still underestimated the surface concentrations of air pollutants. We added Figure S8 in the supplementary materials showing the differences between observed and modeled results. Slightly improvements were found, suggesting the Xingtai supersite is substantially influenced by the nearby emission sources which cannot be resolved by 12-km CMAQ.