Reply to reviewers and the editor:

We thank the reviewers and editor for their careful reading of the manuscript and their many constructive comments. We have adopted most of the suggestions. The original comments by reviewers are in black font, and our replies are in blue.

Comments from Reviewer 1:

The manuscript titled "**Aggravated surface O_3 pollution primarily driven by meteorological variation in China during the early COVID-19 pandemic lockdown period**" assesses the individual effects of meteorology and emission reductions on O_3 changes during COVID-19 lockdown by using GEOS-Chem model. By updating the emission inventories using satellite measurements and further validation of model performance, the study provides a more solid analysis for this issue. The article is well organized. It can be accepted after considering the following suggestions.

Thanks for the positive comments. In below, we address all of your comments.

This study has validated the model performance in simulating NO_2 by comparing with observed NO_2 column. However, the VOCs simulation is bot validated. At least, the model simulated HCHO column can be validated by comparing with observed HCHO column.

We added the comparison of tropospheric HCHO vertical column density (VCD) from GEOS-Chem and TROPOMI in the Appendix D of the updated manuscript. The GEOS-Chem simulations Baseline (2019) and 2020 Adjoint agree well with TROPOMI HCHO in Feb.-Mar. of 2019 and 2020, respectively, with the correlation coefficients of 0.878 and 0.874 and the mean bias of around -24% and -27% for the two comparisons. This negative bias is primarily contributed by the low (< 0.15 DU) HCHO columns (Fig. D1 in the updated manuscript). Vigouroux et al. (2020) reported that TROPOMI tends to overestimate the HCHO by 26% for HCHO column lower than around 0.1 DU. Therefore, the negative bias of our model is in an acceptable range. Please see the Appendix D of the updated manuscript.

The study finds that the large effect of meteorological changes on O_3 changes during lockdown period. However, they only discuss the effect of temperature and solar radiation. The influence of other meteorological factors such as humidity and wind speed should be also discussed.

We added the analysis of influence of relative humidity and wind speed on the surface O_3 change as you suggested. We find that the decreased relative humidity (RH) also contributed to the surface O_3 increase in South China, but we cannot identify any significant impact of wind speed on the surface O_3 pollution. This increase of RH and the increase of temperature could be interconnected thermodynamically, and they both contribute the meteorological effect on O_3. See Section 4.2 in the updated manuscript.

In the abstract, I suggest to mention the time period that you focus on is from 2019 to 2020, rather than from before lockdown to the lockdown period in 2020.

In the updated abstract, we have emphasized that this study investigated the O_3 increase during the lockdown period in 2020 compared to the same period in 2019 instead of the period right before the lockdown. See the abstract in the updated manuscript.

The emission inventory used here is MIX 2010. Why not using MEIC inventory which have updated emissions in 2019 or 2017?

We have several practical reasons for that we used MIX 2010 instead of MEIC emission in 2019 or 2017. First, the default anthropogenic emission for East Asia used in GEOS-Chem is MIX 2010. If we want to adopt MEIC 2017 or 2019 emission in GEOS-Chem, we need to develop a new module in the source code of Harmonized Emissions Component (HEMCO) to add a new emission extension. The validation of the new emission module is also necessary. Even if we successfully implement the MEIC 2019 emission in the GEOS-Chem, we may still need to update the emission for the baseline simulation, because the bottom-up emission inventory did not consider the reduction of emission associated with the pandemic. Second, MEIC only provide emissions in mainland China, but MIX 2010 covers the whole East Asia. Finally, the MEIC team did not provide gridded emission inventory for 2019 until May 2023 (http://meicmodel.org.cn/?page_id=541&lang=en). They developed the anthropogenic emission inventory for 2019 and 2020 in China by province when we were implementing our study (B. Zheng et al., 2021), but we needed gridded emission inventory for our model simulations.

Lines 216-218: Why does the change of anthropogenic VOC emissions from 2010 to 2019 can be ignored?

In the Appendix D of the updated manuscript, we compared the tropospheric HCHO VCD from GEOS-Chem simulation Baseline (2019) and TROPOMI data in 2019 Feb.-Mar. The result indicates a decent model performance with the correlation coefficient (R), mean bias error (MBE), root-mean-square-error (RMSE), and relative mean bias (RMB) of 0.878, -0.036 DU, 0.06 DU and -24.0%, respectively. This good agreement supports our assumption of ignoring the change of VOC emissions from 2010 to 2019, since the anthropogenic VOC emission used for Baseline (2019) is equivalent to MIX 2010. On the contrary, the comparison of tropospheric NO_2 VCD from GEOS-Chem simulation with default MIX 2010 emission and 2019 GEOS-FP meteorological fields versus the TROPOMI data in 2019 Feb.-Mar. show a significant overestimation, with the R value, MBE, RMSE and RMB of 0.720, 0.164 DU, 0.24 DU and 79.2%, respectively. Consequently, we updated the anthropogenic NOx emission but ignored the change of anthropogenic VOC emissions from 2010 to 2019. See the Appendix D of the updated manuscript.

Lines 222-225: It is unclear about how you derived the VOCs emission during lockdown 2020. Do you mean that you derive VOCs emission according to the ratio of HCHO column between lockdown period and before lockdown period and the VOCs emissions in 2020 before lockdown period? If so, the VOCs emissions in 2020 before lockdown period should be introduced here. In addition, why not using the ratio of HCHO column in 2019 to that in 2020 to update the VOCs emission?

Sorry for the confusion. We used the first-order mass balance approach in which the ratio of TROPOMI HCHO in 2020 lockdown period to the same period in 2019 at 0.5° resolution is applied as the scaling factor to update the VOC emissions. Details can be found in the replies below. By "before the lockdown" we were referring to the same season in 2019, not the period in 2020 right before the lockdown. Our choice of words may cause some confusion to you. We have clarified this in Section 2.4 of the updated manuscript. We hope it is clear now.

You have mentioned that the large uncertainty in HCHO prevent accurately quantifying the emission decline. I didn't see that the scaling method that you adopted to update VOC emissions can solve this problem.

We used the first-order mass balance approach as described above. Because of the large retrieval uncertainty of HCHO due to its low optical depth, the oversampling results of HCHO for the 2-month periods are spatially noisy at the resolution of 0.01°, which is the recommended resolution for TROPOMI oversampling by Sun et al. (2018). Therefore, we spatially aggregated the ratio of TROPOMI HCHO in 2020 Feb.-Mar. to that in 2019 Feb.-Mar. to the resolution of 0.5° to reduce the spatial variation. Due to the significant biogenic source of HCHO, the scaling method using the original ratio could underestimate the decline of anthropogenic VOC emission. Therefore, we aggregated the ratio from 0.01° to 0.5° by picking the lowest 25th percentile. Certainly, our method may still have large uncertainty for the updated VOC emission, given that we did not consider other VOC species but HCHO. This is in part because HCHO is directly measured from TROPOMI. Nevertheless, our model simulations have been evaluated using satellite and ground measurements, and the magnitude of the VOC emission impact on surface O_3 is much lower than that of NOx emission and meteorology (Fig. 9 in the manuscript).

Figure 8. The observation shows O_3 concentrations in the Northern China is higher than in the southern China, while the model simulation shows an opposite spatial distribution especially for 2020. The author should clarify the reason for the inconsistency and how this affects your major conclusion about the relative importance of meteorology and emission reductions in O_3 changes.

The model simulation shows an opposite spatial distribution over North China and South China compared with ground observations, especially for 2020, which is mainly caused by the underestimation of the simulated surface O_3 over North China. One possible reason for the underestimation over North China is the underestimation of the biogenic VOC emissions as indicated by Appendix D in the updated manuscript. The underestimation of the simulated O_3

over North China will not impact our study results since this study focuses on revealing the impact of emissions and meteorology change on the surface O_3 change by each region. The bias is predominantly systematic and can be substantially cancelled when we compute the relative difference of the surface O_3. The bias of the simulated relative difference in surface O_3 over North China is acceptable as stated in the manuscript. We have updated the manuscript in Section 3.3 to clarify this part.

Figure 8. In Sichuan Basin, the relative difference from model simulation is negative while the observed relative difference is positive. The opposite trend between model simulation and observation is also displayed in Shandong province. The underlying causes should be clarified.

For Sichuan Basin, the opposite trend of surface O_3 change from the model and measurements is probably caused by the inaccurate simulation of the meteorological effects (Fig. 9) due to the complex terrain features in this region. We added this clarification in Section 3.3 of the updated manuscript. However, further investigation of the reasons for the bias of the simulated surface O_3 change in this region is out of the scope of this study since our South China domain exclude the most part of this region. For Shandong province, the deviation of the simulated trend of surface O_3 change is actually much lower than that in Sichuan Basin, and this small bias is acceptable as stated in Section 3.3. The model is always imperfect, and our main conclusions reflect overall results. We noted that localized improvement is further needed. See Section 3.3 of updated manuscript for the clarification.

Lines 497-499: The results of changes in the cloud fraction and the downward visible direct flux should be provided.

We have added the changes of the cloud fraction and the downward visible direct flux at surface in Fig. 10 of the updated manuscript as you suggested.

Lines 505-506: Why does the intercept is negative in North China?

The negative intercept indicates that the surface O_3 could decrease even if the temperature remains the same in North China. This decreasing is caused by the net effects of factors other than temperature, including chemistry, emissions, and other meteorological factors. For instance, the VOC emission decline and the increasing wind speed in North China could contribute to the negative intercept here (Fig. 9 and Fig. 11 in the updated manuscript). It is a challenge to quantify the contributions of each individual factors, because these factors are thermodynamically or dynamically related. Collectively, we refer their impact as meteorological impact and focused on the analysis of some key factors such as temperature. We added this clarification in Section 4.2 of the updated manuscript.

Figure 1. It should be specified that which is minuend and subtrahend for the calculation of relative difference.

As you suggested, we added the clarification of the minuend and subtrahend in the caption of Fig. 1 in the updated manuscript.

Comments from Reviewer 2:

This work investigated the reasons for the increase of surface ozone concentrations in perspectives of emission changes and meteorological variations. It's found that meteorological variation is the major driving force to the ozone increase in South China. The paper is informative and generally well organized. My major concern is the negative bias of TROPOMI data (~20% based on previous validations), especially over urban areas, and the effect to the results due to the bias.

Thanks for the positive comments. Our item-by-item responses can be found in below.

My detailed comments are listed below.

1. Line 126-127: not clear about what this means. Do you mean the difference between NOx, NOy, and NOz? Please specify this.

   The China National Environmental Monitoring Center measures the surface NO_2 using the chemiluminescence method, which quantifies the NO_2 concentrations by measuring the NO decomposed from NO_2, which may introduce a positive bias because NOz can also be decomposed to NO. You are right. We clarified it in Section 2.2 in the updated manuscript. We hope it is clear now.

2. Line 166-167: is the optimization done day by day, or on a monthly basis?

   The optimization was implemented by assimilating the daily TROPOMI NO_2 measurements to constrain the emission at the monthly scale. The scale factor is optimized for each month at each gridbox, as the original temporal resolution of MIX emissions is monthly. The cost function in the optimization seeks to improve the overall agreement of daily TROPOMI and GEOS-Chem NO_2 in a month. We have clarified this in Section 2.3 of the updated manuscript.

3. Line 216-217: the anthropogenic VOC emissions in China has grown significantly from 2010-2017, driven by solvent use. I don't think it can be ignored.

To further verify the assumption that anthropogenic VOC emission change can be ignored, we compared the GEOS-Chem simulation of HCHO with TROPOMI data. The results are added in Appendix D of the updated manuscript. The GEOS-Chem simulations Baseline (2019) and 2020 Adjoint, which used the default MIX 2010 and updated MIX 2020 VOC emissions respectively, agree well with TROPOMI HCHO in Feb.-Mar. of 2019 and 2020, respectively (Fig. D1 in the updated manuscript). The correlation coefficients for the two comparisons are 0.878 and 0.874, and the mean bias error are around -24% and -27%, respectively. The negative bias is primarily contributed by the low (< 0.15 DU) HCHO columns (Fig. D1 in the updated manuscript), and it is in an acceptable range considering TROPOMI tends to overestimate the HCHO by 26% for HCHO column lower than around 0.1 DU (Vigouroux et al., 2020).This good agreement supports our assumption of ignoring the VOC emission change.

Fig. D1 also shows the GEOS-Chem tends to overestimate the HCHO VCD over urban regions. It indicates default MIX 2010 may overestimate the anthropogenic emissions in 2019, at least for HCHO, although Bo Zheng et al. (2018) reported the anthropogenic emissions of non-methane volatile organic compounds (NMVOCs) increased by 11% during 2010-2017.

In addition, the uncertainty of the anthropogenic VOC emissions in this study can be smeared by the large biogenic source of VOC emissions.

See Appendix D of the updated manuscript.

4. Line 221-222: can you elaborate more on the HCHO uncertainties?

The satellite retrieval of HCHO using the differential optical absorption spectroscopy (DOAS) method has large uncertainty because of the low optical depth of HCHO. The total retrieval uncertainty is composed of the errors of slant columns and model errors (De Smedt et al., 2018). The errors of the slant columns are contributed by the characteristics of the measuring instrument and errors of the slant column fitting procedure in the DOAS method. The model errors are related to the inaccurate representation of the physical properties of the atmosphere and surface, such as the treatment of clouds and aerosols and the surface albedo (De Smedt et al., 2018). Vigouroux et al. (2020) reported that TROPOMI HCHO tends to be overestimated by 26% for HCHO column lower than 0.093 DU (2.5×10^{15} molecules cm^{-2}) and underestimated by 31% for HCHO column higher than 0.297 DU (8.0×10^{15} molecules cm^{-2}). We added this quantification of the TROPOMI HCHO uncertainty in the Section 2.4 of the updated manuscript.

5. Figure 2. For (c), it seems that the VOC emission changes are mainly from biogenic or open biomass burning for Southeast Asia, instead of anthropogenic sources.

TROPOMI HCHO data cannot distinguish the anthropogenic emissions from biogenic and biomass burning sources for the Indochinese Peninsula in Southeast Asia because of the dense vegetation in this region. However, this study investigated the O_3 pollution in China, the Southeast Asia with the dense vegetation is out of our study domain. The impact of VOC emission bias in Southeast Asia on the surface O_3 pollution in China is negligible considering the lifetime of biogenic VOC is generally short. For example, the lifetime of the most

prevalent biogenic VOC species, isoprene, is only 1-2 hours (Atkinson, 2000). For the populated urban regions in China, where the surface O_3 pollution exerts more significant health impacts, the anthropogenic source dominates the VOC emissions (Williams & Koppmann, 2007). We made this clarification in Section 3.1 of the updated manuscript.

6. Line 294: previous validations have found that TROPOMI NO_2 product has negative bias of ~20%, especially over urban areas (Verhoelst et al., 2021, Judd et al., 2020, Li et al., 2021). How will this bias affect the validation and your conclusions?

Verhoelst et al. (2021) reported that TROPOMI tropospheric NO_2 column displays a negative bias of -23% to -37% in clean to slightly polluted areas but reaching up to -51% over highly polluted regions compared to the ground-based measurements, but they did not consider the averaging kernel in the comparison. Judd et al. (2020) reported that the negative bias of TROPOMI NO_2 was reduced from 19%-33% to 7%-19% by recalculating the tropospheric air mass factor (AMF) using the averaging kernel to replace the priori NO_2 profiles. Li et al. (2021) also found the TROPOMI NO2 displays a negative bias up to -20% over polluted regions even with the AMF correction applied. In our study, we applied the AMF correction using the TROPOMI averaging kernel for both data assimilation and model evaluations. We expect a negative bias up to -20% of TROPOMI NO_2 even for urban regions. This low bias may cause an underestimation of NOx emission in our study.

To evaluate the influence of the underestimation of the NOx emissions on our main conclusions, we conducted sensitivity simulations with the anthropogenic NOx emissions increased by 20%. The validation of simulated NO_2 column against TROPOMI NO_2 would not be significantly affected since we are using TROPOMI NO_2 for both data assimilation and model evaluation, and the bias in TROPOMI data will be cancelled out. The impacts on the O_3 simulation would be a major concern. Fig. R1 below displays the comparison of MDA8 surface O_3 from the simulation 2020 Adjoint (Table 1 in the Manuscript) and the one with NOx emission increased by 20%. The relative difference is within 6% in North China and within 2% in South China. Overall, this difference is lower than the inter-annual variation of surface O_3 between 2019 and 2020 by one order of magnitude (Fig. 8 in the manuscript).

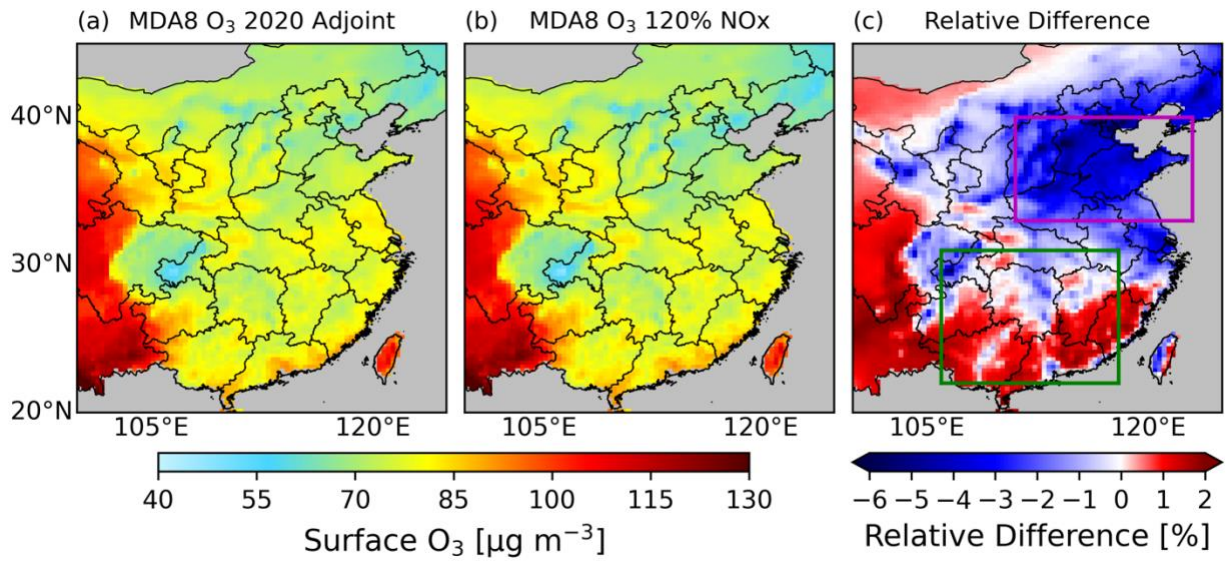

Figure R1. (a) GEOS-Chem mean MDA8 O_3 at 9 m above the surface under standard temperature and pressure (273.15 K, 101.325 kPa) from 2020 Adjoint simulation (Table 1 in the manuscript); (b) Same as (a) but with anthropogenic NOx emission increased by 20%; (c) Relative difference between (b) and (a), which is calculated as dividing the difference of (b) minus (a) by (a).

We further investigated how this difference affect the validation of O_3 simulation against ground measurements. Fig. R2 shows the scatter plot for comparing the surface MDA8 O_3 over South China from ground measurements versus GEOS-Chem simulations 2020 Adjoint and the one with NOx emission increased by 20%. Over South China, the model performance of surface O_3 simulation after increasing NOx emission by 20% is similar to that of default simulation, with mean bias error (MBE) decreased by 40.2% (0.553 μg m^{-3}) but root mean square error (RMSE) increased by 5.51% (0.484 μg m^{-3}). Over North China, the underestimation of the surface O_3 becomes more pronounced as NOx emission increased by 20% (Fig. R3). The negative mean bias increased from -12.620 μg m^{-3} to -15.624 μg m^{-3} by 23.8%, but the slope of the regression line increased from 0.336 to 0.372. Overall, the impact of increasing NOx emission by 20% on the surface O_3 simulation is not significant.

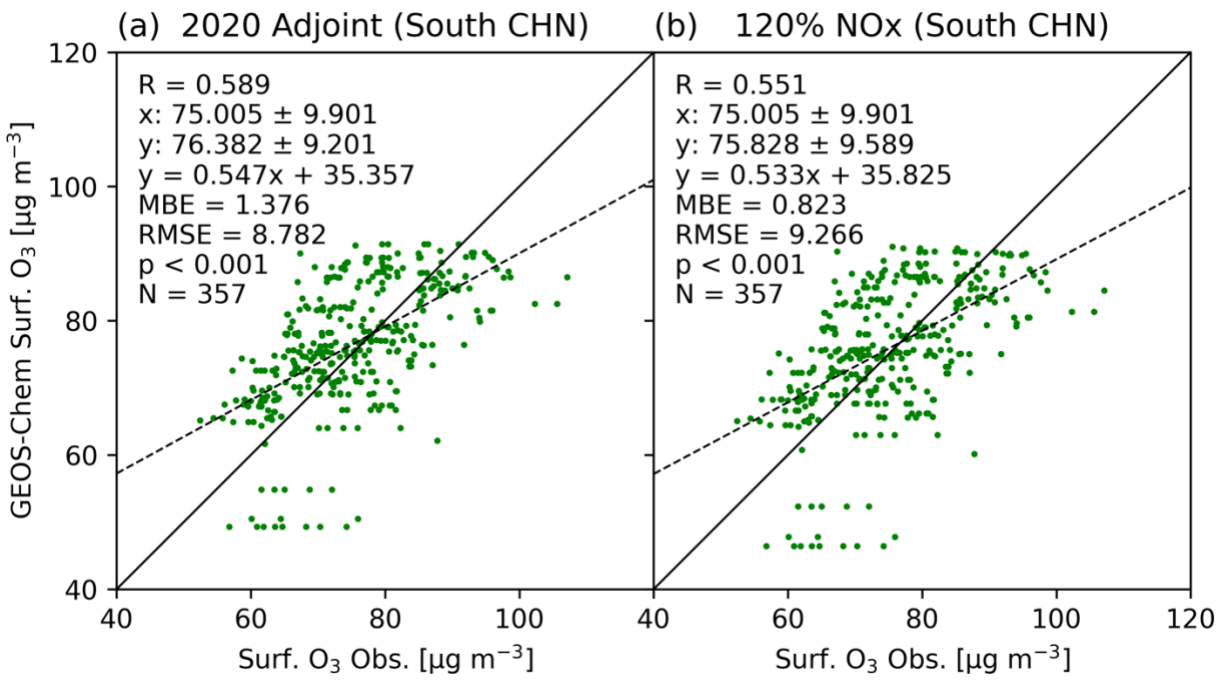

Figure R2. Scatter plot for comparing the surface MDA8 O_3 from GEOS-Chem simulations and ground measurements over South China (green box in Fig. R1). (a) GEOS-Chem simulation 2020 Adjoint versus ground measurements in 2020 Feb.-Mar. (b) The simulation same as 2020 Adjoint but with anthropogenic NOx emission increased by 20% versus ground measurements in 2020 Feb.-Mar.

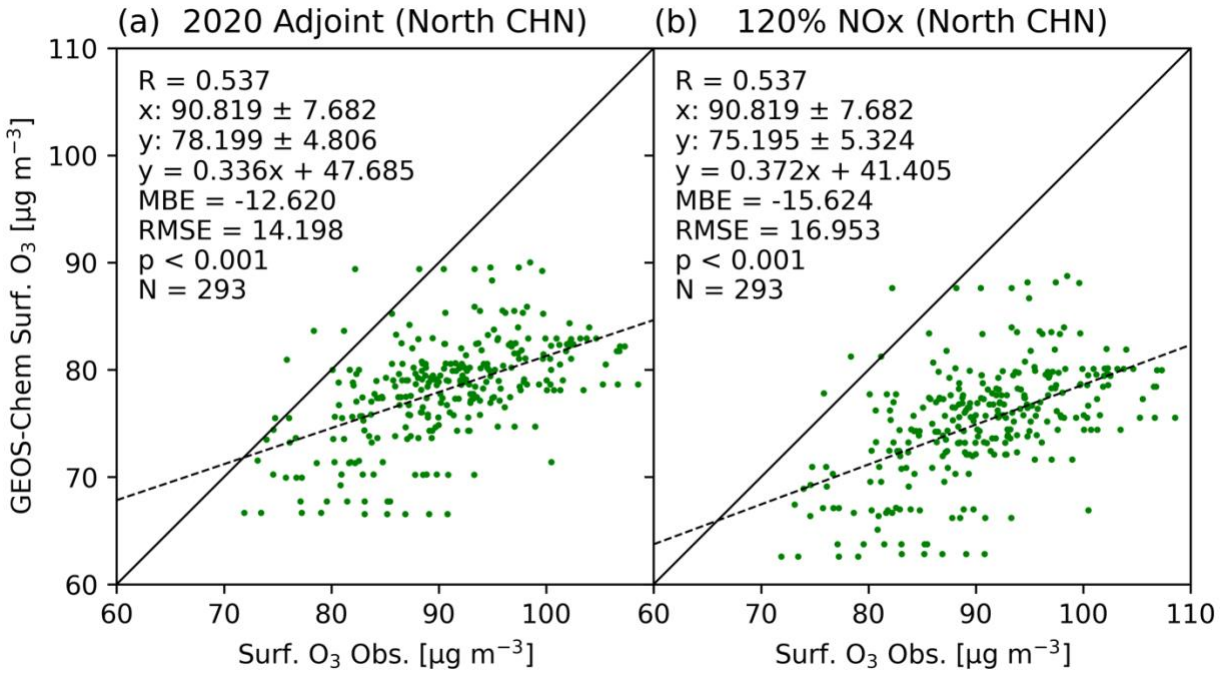

Figure R3. Same as Fig. R2 but for North China.

For the conclusion on the impacts of emission change and meteorological variation on the surface O_3 pollution, which is calculated from the relative difference of various model sensitivity simulations, the influence of underestimation of NOx emission is not significant. The reason is that we updated the NOx emission using TROPOMI NO_2 data for each simulation and the impacts of underestimation of NOx emission would be partly cancelled out. We also conducted the sensitivity simulations by increasing the anthropogenic NOx emission by 20% for both 2019 Baseline and 2020 Adjoint and computed their relative difference. The variation of the relative difference is within 4% compared to that of default simulations. The main conclusion that inter-annual meteorological variations have a larger impact than emission reductions on the aggravated surface O_3 pollution in China during the lockdown period of COVID-19 pandemic still holds considering the low bias in the TROPOMI NO_2 data. We noted this part in Section 4.1 in the updated manuscript.

References for reviewer comments:

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. Atmospheric Measurement Techniques, 14(1), 481–510. https://doi.org/10.5194/amt-14-481-2021

Judd, L. M., Al-Saadi, J. A., Szykman, J. J., Valin, L. C., Janz, S. J., Kowalewski, M. G., et al. (2020). Evaluating Sentinel-5P TROPOMI tropospheric NO2 column densities with airborne and Pandora spectrometers near New York City and Long Island Sound. Atmospheric Measurement Techniques, 13(11), 6113–6140. https://doi.org/10.5194/amt-13-6113-2020

Li, M., McDonald, B. C., McKeen, S. A., Eskes, H., Levelt, P., Francoeur, C., Harkins, C., He, J., Barth, M., Henze, D. K., Bela, M. M., Trainer, M., de Gouw, J. A., and Frost, G. J.: Assessment of Updated Fuel-Based Emissions Inventories Over the Contiguous United States Using TROPOMI NO2 Retrievals, Journal of Geophysical Research: Atmospheres, 126, e2021JD035484, https://doi.org/10.1029/2021JD035484, 2021.

References for our replies:

Atkinson, R. (2000). Atmospheric chemistry of VOCs and NOx. *Atmospheric Environment, 34*(12-14), 2063-2101. 10.1016/s1352-2310(99)00460-4

De Smedt, I., Theys, N., Yu, H., Danckaert, T., Lerot, C., Compernolle, S., et al. (2018). Algorithm theoretical baseline for formaldehyde retrievals from S5P TROPOMI and from the QA4ECV project. *Atmospheric Measurement Techniques, 11*(4), 2395-2426. 10.5194/amt-11-2395-2018

Judd, L. M., Al-Saadi, J. A., Szykman, J. J., Valin, L. C., Janz, S. J., Kowalewski, M. G., et al. (2020). Evaluating Sentinel-5P TROPOMI tropospheric NO2 column densities with

airborne and Pandora spectrometers near New York City and Long Island Sound. *Atmospheric Measurement Techniques, 13*(11), 6113-6140. 10.5194/amt-13-6113-2020

Li, M., McDonald, B. C., McKeen, S. A., Eskes, H., Levelt, P., Francoeur, C., et al. (2021). Assessment of Updated Fuel-Based Emissions Inventories Over the Contiguous United States Using TROPOMI NO2 Retrievals. *Journal of Geophysical Research-Atmospheres, 126*(24). 10.1029/2021jd035484

Sun, K., Zhu, L., Cady-Pereira, K., Chan Miller, C., Chance, K., Clarisse, L., et al. (2018). A physics-based approach to oversample multi-satellite, multispecies observations to a common grid. *Atmospheric Measurement Techniques, 11*(12), 6679-6701. 10.5194/amt-11-6679-2018

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. *Atmospheric Measurement Techniques, 14*(1), 481-510. 10.5194/amt-14-481-2021

Vigouroux, C., Langerock, B., Aquino, C. A. B., Blumenstock, T., Cheng, Z. B., De Mazière, M., et al. (2020). TROPOMI-Sentinel-5 Precursor formaldehyde validation using an extensive network of ground-based Fourier-transform infrared stations. *Atmospheric Measurement Techniques, 13*(7), 3751-3767. 10.5194/amt-13-3751-2020

Williams, J., & Koppmann, R. (2007). Volatile Organic Compounds in the Atmosphere: An Overview. In *Volatile Organic Compounds in the Atmosphere* (pp. 1-32).

Zheng, B., Tong, D., Li, M., Liu, F., Hong, C., Geng, G., et al. (2018). Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions. *Atmospheric Chemistry and Physics, 18*(19), 14095-14111. 10.5194/acp-18-14095-2018

Zheng, B., Zhang, Q., Geng, G. N., Chen, C. H., Shi, Q. R., Cui, M. S., et al. (2021). Changes in China's anthropogenic emissions and air quality during the COVID-19 pandemic in 2020. *Earth System Science Data, 13*(6), 2895-2907. 10.5194/essd-13-2895-2021