# **Reply to reviewers and editors:**

We thank all of the reviewers for their careful reading of the manuscript, and for their many constructive feedbacks. The original comments by reviewers are in black font, our replies are in blue.

# Reviewer #2

This manuscript presents down-scaling results of SO2 and NOx emissions and concentrations based on the coarse-resolution joint emission inversion results from Part 1. The downscaling approaches used information from TROPOMI NO2 observations, MIX inventories, and VIIRS nighttime light observations. The downscaling results were compared against surface in-situ observations. The impact on regional air quality forecasting is also addressed. The prosed approaches are unique and could contribute to improving regional air quality modeling. I would, however, advise the authors to revise the manuscript. These revisions should be made before the manuscript can be considered for publication.

Thanks for the positive comments and constructive reviews. We've done our best to address the comments in the revision.

## [Major comments]

As I suggested for Part 1, all the results need to be revised using higher resolution (at  $0.5 \circ x 0.667 \circ$  degree resolution) joint inversion results for this type of regional study. The  $2 \circ x 2.5 \circ$  resolution inversion could lead to large systematic biases in both local and regional emissions and concentrations in the downscaling analysis, associated with the non-linear chemistry. Ideally, inversion calculations should be done at  $0.25 \circ$  or  $0.05 \circ$  degree resolution to provide reference information for the downscaling results, but this could be difficult.

It is more straightforward to conduct high-resolution inversions using regional models. There are already several high resolution regional inversion frameworks, for instance, using WRF-Chem. The benefit of using the current coarse resolution global joint inversion framework (rather than regional high resolution inversion systems) to improve regional NOx and SO2 emissions and air quality forecast for China needs to be discussed.

Thanks for the comments. Please see our replies to your comments on part I. The main reason is that the coarser-resolution OMPS data cannot resolve the emission sources smaller than its resolution on a monthly basis. In addition, there are practical issues such as computational cost and the availability of data at fine resolution – the earlier the emissions can be updated by satellite, the better the outcomes of these satellite observations for air quality forecast. The goal of this study is to develop methods potentially for improve forecasting in real time rather than to reconstruct the best historical analysis of high-resolution emissions (that was the goal of (Qu et al., 2019)). Thus, the quality of the forecast is the ultimate test of this study.

While regional model is best suited for air quality forecast, to our knowledge, when this manuscript was prepared, GEOS-chem adjoint model still remains the only CTM that has complete thermodynamic description of the secondary inorganic sulfate-nitrate-ammonium aerosol system and has no need to deal with the issues of chemical boundaries for the regional model. Our understanding is that the full-chemistry 4D-Var is not yet possible in WRF-Chem as its adjoint is made only for GOCART scheme at this point, while CMAQ-adjoint (Zhao et al., 2019) model were in open review after this paper was submitted. Again, treating the chemical boundary conditions in the regional adjoint model need to be further studies. All of this are the reasons that ended up to use GEOS-Chem. We plan to use CMAQ-adjoint in near future.

Zhao, S., Russell, M. G., Hakami, A., Capps, S. L., Turner, M. D., Henze, D. K., Percell, P. B.,
Resler, J., Shen, H., Russell, A. G., Nenes, A., Pappin, A. J., Napelenok, S. L., Bash, J. O.,
Fahey, K. M., Carmichael, G. R., Stanier, C. O., and Chai, T.: A Multiphase CMAQ Version 5.0
Adjoint, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-287, in review, 2019.

The current manuscript is technical and does not seem provide sufficient scientific implications for ACP (not for GMD). It is required to provide scientific implications based on the proposed approaches. For instance, more detailed information on differences in the spatial patterns between VIIRS nighttime lights and MIX inventories and possible biases in the MIX emission inventories for each emission category would be interesting. Such information will be essential to determine the best downscaling approach for right reason.

Thanks for the good suggestion. Please see our reply to your comments for the first part regarding the scientific merit of the paper here. ACP is very broad and methods to improve emission estimates and air quality forecasting are within its scope. Furthermore, we demonstrate the potential application of our method (monthly update of emission at coarse resolution and downscaling it) for regional air quality forecast (for the next month). Finally, following your good suggestion, we have added more discussion to Sect. 4.4. New materials about the relationship among Volatile Organic Compound (VOC), NO<sub>x</sub> emission, and O<sub>3</sub> air pollutions in Sect 4.6 is also added to make the manuscript have more scientific merits. Below is what we have add to Sect. 4.4.

MIX-DE-POS has improved values of R and NCRMSE than NL-DE-POS; here we discuss the possible reasons and propose future works to improve NL-DE. MIX is a mosaic bottom-up emission inventory, and it is actually the MEIC emission inventory for NO<sub>x</sub> emissions over China (Li et al., 2017). The MIX (or MEIC) NO<sub>x</sub> emission inventory over China consists of emissions from four sectors including coal-fired power plant, industrial, transport, and residential sectors. Coal-fired power plant emissions in MEIC are derived through extensively using detailed information (including locations of individual units) of 7657 generation units in China (Liu et al., 2015); coal-fired power plant emissions can be accurately placed to grids according to source location information (Li et al., 2017). Thus, if we can allocate posterior total anthropogenic NO<sub>x</sub> emissions into the four sectors, we expected that it is better to use the MIX coal-fired power plant NO<sub>x</sub> emission inventory rather than nighttime lights to downscale the posterior coal-fired power plant  $NO_x$  emissions. For the other sectors in MIX (or MEIC) over China, population density is used to allocate industrial and residential emissions to grids (Li et al., 2017), and transport emissions are distributed according to road networks (Li et al., 2017). Using population density to downscale industrial and residential NO<sub>x</sub> could underestimate emissions over urban region, compared with the approach of using nighttime light which could better represent economic development levels (Geng et al., 2017). Whether it is better to use road networks or nighttime lights to downscale NO<sub>x</sub> emissions from the transport sector requires future investigations. In this study, the posterior NO<sub>x</sub> emission inventory to be downscaled is

total anthropogenic NO<sub>x</sub> emissions, which is not allocated into different source sectors. Thus, if we assume that the ratios of every sectoral emissions to total anthropogenic emissions do not change between prior and posterior emission inventories, MIX-DE has an advantage for coalfired power sector, while NL-DE could benefit the downscaling for the industrial and residential sectors. In future work, we could optimize sectoral emissions rather than total anthropogenic emissions, and subsequently downscale posterior coal-fired power emissions through prior MIX coal-fired power emissions, and ultimately use VIIRS night time light data to downscale posterior industrial and residential emissions.

The evaluations of forecasts in Section 4.6 are not informative in the current form. Because the purpose of this study is to improve regional air quality forecasts, evaluations of simulated ozone (one of the most important air pollutants) using in-situ observations would add important information.

Both SO<sub>2</sub> and NO<sub>2</sub> are criteria pollutants in the atmosphere defined by US EPA and China, thus we think it is appropriate to evaluate the forecasts of two trace gases and state the improvements in regional air quality forecasts. We also agree that evaluations of simulated ozone are very important, and this part is now added into the revision. As shown below, using downscaled posterior emission inventory helps to improve spatial distribution in terms of NCRMSE, and the improvement of NMB depends on region. We have added the text and Fig. 14 below in Sect. 4.6 and Fig. S5 in the supplement.

In addition to the improvement of SO<sub>2</sub> and NO<sub>2</sub>, AQF-MIX-DE-POS enhances on AQF-PRI in the forecast of surface O<sub>3</sub> concentrations (Fig. 14). If all O<sub>3</sub> in situ observations in the research domain are used for evaluation, a spatial distribution improvement is shown with NCRMSE decreasing from 1.08 for AQF-PRI to 1.05 for AQF-MIX-DE-POS, but NMB changes from -3.1% to 5.0% (Fig. 14c). Indeed, whether bias becomes smaller or larger depends on region. In the North China Plain and Eastern China where NO<sub>x</sub> emissions (or NO<sub>2</sub> surface concentrations) are large (the black box in Fig. 14a), forecasts of surface O<sub>3</sub> concentration are much lower than other regions; and the NMB is -16.7% for AQF-PRI and -6.3% for AQF-MIX-DE-POS with NCRMSE decreasing from 1.20 to 1.16 (Fig. 14c). In this relatively NO<sub>x</sub>-rich region, the increase of O<sub>3</sub> concentration in AQF-MIX-DE-POS is caused by the decrease of NO<sub>2</sub> concentrations; the change of  $SO_2$  concentrations has negligible impacts on  $O_3$  concentrations (Fig. S5). This implies that if Volatile Organic Compound (VOC) concentrations remain constant, emission control of  $NO_x$  emissions will exacerbate  $O_3$  pollutions. For the region that is out of the black box, although NCRMSE decreases from 0.82 for AQF-PRI to 0.80 for AQF-MIX-DE-POS, NMB increases from 19.0% to 23.3% (Fig. 14c).



Figure 14. Evaluation of GEOS-Chem surface  $O_3$  forecasts with in situ observations for November 2013. (a) is AQF-PRI  $O_3$  forecasts with in situ observations overlapped. (b) is the difference between and AQF-MIX-DE-POS and AQF-PRI  $O_3$  forecasts (c) is the Taylor diagram of evaluations of surface  $O_3$  forecasts in (a) and (b) with in situ observations. Circles and squares represent the AQF-PRI and AQF-MIX-DE-POS forecasts, respectively. Labels 1, 2, and 3 represent that all sites, only sites that are within the black box in (a), and only sites that are out of the black box in (a) are used for evaluations.



Figure S5. (a) is similar to Fig. 14c, but in the posterior forecasts, the prior MIX  $NO_x$  emission inventory and the posterior MIX-DE  $SO_2$  emission inventory is used. (b) is similar to Fig. 14c, but in the posterior forecasts, the prior MIX  $SO_2$  emission inventory and the posterior MIX-DE  $NO_x$  emission inventory is used.

The use of GCv12.0.0 model instead of GC adjoint v35m could provide some insights into the model dependent posterior emission inventory. Nevertheless, the usefulness of the proposed downscaling approach should first be evaluated in a consistent framework (GC v35m) to avoid too much complications. Otherwise, it is required to demonstrate the model performance difference in detail.

Well, the robustness of the emission inventory should be independent of the CTMs. This is the original motive for us to use a different version of GC to assess the value of the optimized emission. Following your suggestion, we also conducted some evaluations in a consistent framework. Fig. S2 and Fig. S3 are similar to Fig. 4 and Fig. 5, respectively, but using the GC adjoint v35m rather than GCv12.0.0. Apparently, both MIX-DDC and NL-DC works when GC adjoint v35m is used for coarse resolution simulation. All conclusions about downscaling through MIX-DDC and NL-DC from analyzing GCv12.0.0 results can also be drawn from GC adjoint v35m results. It is not surprising that when a consistent framework (GC adj v35m) is used for coarse resolution simulation, all downscaled results show better spatial pattern (smaller NCRMSE) than using GCv12.0.0 for coarse resolution simulation. Considering the manuscript

has shown MIX-DE results are similar to MIX-DDC results, we can expect that MIX-DE should also work in a consistent framework (GC adj v35m). We have added the two figures to the supplement and corresponding text to Sect. 4.1 and Sect. 4.2.







Figure S3. It is similar to Fig. 5, but GC adjoint v35m rather than GCv12.0.0 is used.

#### More specific comments:

3.2.2 I'm wondering if this approach can be applied to SO2. If not, please explain the reason. We did not apply this approach to SO2.

We use VIIRS nighttime to downscale NO<sub>2</sub> concentrations as there is strong linear correlation between NO<sub>2</sub> VCD and nighttime light as shown in Fig 2c. The strong linear correlation is caused by two reasons: (1) nighttime lights are good spatial proxy for allocating NO<sub>x</sub> emissions (Geng et al., 2017); and (2) NO<sub>2</sub> lifetime is short (several hours), which means the distribution of NO<sub>2</sub> concentration hot spots are highly affected by source locations. We do not expect this approach can be used to downscale SO<sub>2</sub> concentrations for the two reasons: (1) nighttime lights are not very good spatial proxy for allocating SO<sub>2</sub> emissions as SO<sub>2</sub> emissions from traffic sector are very small while nighttime lights are strong over rush traffic road; (2) SO<sub>2</sub> lifetime is 1-2 days, which is much longer than NO<sub>2</sub> lifetime. We have added the explanation to the section.

L350 "Thus, for SO2. . ." This suggests that the overall spatial pattern was degraded, while capturing hot spots. What emission sources were actually degraded? This would provide important implications into the emission inventories.

Yes, compared with coarse-resolution simulations, the overall spatial patterns of fine-resolution simulations are degraded, although this conclusion is based on the ground-based observation data that are also in coarse resolution as a whole for describing the spatial pattern. The spatial pattern degradation implies that current chemistry transport simulations of surface SO<sub>2</sub> concentrations can capture regional spatial pattern (coarse-resolution) well, but it is difficult to simulate local spatial pattern (fine-resolution); the weakness for describing the local spatial pattern simulation suggests the uncertainties of either bottom-up SO<sub>2</sub> emission estimates at fine resolution or locally-resolved meteorological fields (Ge et al., 2017), or both. This uncertainty in bottom-up emission inventories can further stem from distributing SO<sub>2</sub> emissions at provincial level to fine-resolution grid. We have added the discussion to Sect. 4.1.

Xing et al. (2015) also showed the difficulty of simulating local  $SO_2$  pollution. In Xing et al. (2015)'s research, in situ  $SO_2$  observations from US-CASTNET and US-AQS were used for evaluation. US-CASTNET sites are mainly located in rural areas to represent regional air pollution, while US-AQS sites are mainly close to urban areas to represent much smaller area

(local air pollution) (Xing et al., 2015). The linear correction coefficients between WRF-CMAQ simulations of surface  $SO_2$  concentrations (108 km x 108 km resolution) over the US and in-situ observations were 0.67 and 0.2 when observations from US-CASTNET and US-AQS were used for evaluation, respectively (Xing et al., 2015).

In this study, observational sites are mainly over urban area, and linear correlation coefficients between GEOS-Chem fine-resolution simulations and observations are in the range of from 0.26 to 0.36 (Fig. 4g and h), which is comparable to the value of 0.2 in Xing et al. (2015)'s research. For coarse resolution simulations, the same sites are used for evaluation, but linear correlation coefficients are in the range from 0.62 to 0.66. In the process of evaluating coarse resolution simulations, there are usually several observational sites in a coarse grid box, and observations from these sites are averaged to compare with the simulation of the coarse grid box. The better spatial pattern at coarse resolution also means it is much easier for GEOS-Chem simulations to capture regional spatial pattern of surface  $SO_2$  concentrations than local spatial pattern.

L360 "The MIX-DDC-POS...". It is not clear to me that the POS is better and the CGS effect still exists (how did you know?).

Thanks for pointing out this. We would like to express that the MIX-DDC-POS simulation is better than the MIX-DDC-PRI simulation **in terms of spatial pattern (NCRMSE)**, although the MIX-DDC-POS simulation has larger negative bias than the MIX-DDC-PRI simulation. We partly ascribe the negative bias to the CGS effect.

We acknowledge that simulation bias is at least affected by emission bias and the CGS effect Thus negative NMB may be CGS effect as well as emission bias. Zheng et al. (2017) showed that surface  $SO_2$  (NO<sub>2</sub>) concentration simulations from WRF-CMAQ, when evaluating with in situ observations, have a NMB of -23% (%0), 7% (32%), and 41% (45%) at the resolutions of 36 km (~0.36°), 12 km (~0.12°), and 4 km (~0.04°), respectively, which suggests that (1) the CGS effect and other non-linear resolution-dependent processes can affect the results and (2) these problems are alleviated at the resolution of 0.25°x0.3125°, but are not completely avoided. We have added that CGS effect is only reduced in part, and other factors needs to be investigated (section 4.1 and section 4.2). L365 "Thus MIX-DDC-POS".. Why did the MIX-DDC approach show good spatial pattern for NO2 and not for SO2? The MIX SO2 and NO2 spatial pattern should look similar.

The performance of the MIX-DDC approach is largely affected by fine-resolution simulations of surface species concentration spatial pattern using prior MIX emission inventory. Spatial patterns of SO<sub>2</sub> and NO<sub>2</sub> are comparable at coarse resolution. When come to fine-resolution simulations, SO<sub>2</sub> spatial pattern degrade much stronger than NO<sub>2</sub> spatial pattern. Thus, MIX-DDC approach show good spatial pattern for NO<sub>2</sub> but not for SO<sub>2</sub>. The reason why the simulation of SO<sub>2</sub> spatial patter at fine resolution has been discussed for the question "L350 "Thus, for SO2. . ." This suggests that the overall spatial pattern was degraded, while capturing hot spots. What emission sources were actually degraded? This would provide important implications into the emission inventories. ", as shown above.

## L378 Why are there large positive biases?

The bias of surface NO<sub>2</sub> concentrations are 45.3% and 25.5% for NL-DC-PRI and NL-DC-POS, respectively, which could come from total emission bias as well as the downscaling process through the NL-DC approach. The sites used for validation are mainly over urban region, and we lack sites that are located over rural region to evaluate if positive or negative bias persists over rural region. Thus, we are not able to determine how much of positive bias in NL-DC-PRI and NL-DC-POS is caused by the NL-DC approach. We have added the discussion in Sect. 4.2.

#### L400 The correlation is very low. Please discuss it.

As we answer the question above for "L350", we have shown that the weakness for describing the local spatial pattern simulation suggests the uncertainties of either bottom-up SO<sub>2</sub> emission estimates at fine resolution or locally-resolved meteorological fields (Ge et al., 2017), or both. This uncertainty in bottom-up emission inventories can further stem from distributing SO<sub>2</sub> emissions at provincial level to fine-resolution grid. The low correlation here also implies this problem. We have added the following discussion in Sect. 4.3

We also noticed that R is 0.14 in MIX-DE-POS, which is even smaller than 0.23 in F-PRI. Thus, in the simulations, using prior emissions inventories shows better linear correlation than using posterior emissions inventories. Conversely. in the forecast, using posterior emission inventories (AQF posterior, R=0.11, Fig. 12b) has better linear correlation than use prior emission inventories (AQF prior, R=0.06, Fig. 12a). The contrast may be caused by the fact that linear correlation coefficient is not a robust metric and should be used together with other metrics to evaluate the model.

L416 I'm not sure if this is really caused by the CGS effect only. For instance, what happens when posterior emissions are biased?

Thanks for pointing out this. We acknowledge that simulation bias is at least affected by emission bias and the CGS effect. Thus, negative NMB may be CGS effect as well as emission bias. Zheng et al. (2017) showed that surface SO<sub>2</sub> (NO<sub>2</sub>) concentration simulations from WRF-CMAQ, when evaluating with in situ observations, have a NMB of -23% (%0), 7% (32%), and 41% (45%) at the resolutions of 36 km (~0.36°), 12 km (~0.12°), and 4 km (~0.04°), respectively, which suggests that (1) CGS effect and other non-linear resolution-dependent processes can affect SO<sub>2</sub> simulation results and (2) these problems are alleviated at the resolution of 0.25°x0.3125°, but are not completely avoided. We have added the discussion above to Sections 4.1 and 4.2. We also have let this sentence replaced by "which should be partly caused by the CGS effect, although emission bias and other non-linear resolution-dependent processes could play a role."

L417 "which may be attributed. . ." I don't understand the sentence. We have replaced the sentence by the text below.

MIX-DE-POS has improved values of R and NCRMSE than NL-DE-POS; here we discuss the possible reasons and propose future works to improve NL-DE. MIX is a mosaic bottom-up emission inventory, and it is actually the MEIC emission inventory for NO<sub>x</sub> emissions over China (Li et al., 2017). The MIX (or MEIC) NO<sub>x</sub> emission inventory over China consists of emissions from four sectors including coal-fired power plant, industrial, transport, and residential sectors. Coal-fired power plant emissions in MEIC are derived through extensively using detailed information (including locations of individual units) of 7657 generation units in China

(Liu et al., 2015); coal-fired power plant emissions can be accurately placed to grids according to source location information (Li et al., 2017). Thus, if we can allocate posterior total anthropogenic NO<sub>x</sub> emissions into the four sectors, we expected that it is better to use the MIX coal-fired power plant NO<sub>x</sub> emission inventory rather than nighttime lights to downscale the posterior coal-fired power plant NO<sub>x</sub> emissions. For the other sectors in MIX (or MEIC) over China, population density is used to allocate industrial and residential emissions to grids (Li et al., 2017), and transport emissions are distributed according to road networks (Li et al., 2017). Using population density to downscale industrial and residential  $NO_x$  could underestimate emissions over urban region, compared with the approach of using nighttime light which could better represent economic development levels (Geng et al., 2017). Whether it is better to use road networks or nighttime lights to downscale NO<sub>x</sub> emissions from the transport sector requires future investigations. In this study, the posterior NO<sub>x</sub> emission inventory to be downscaled is total anthropogenic NO<sub>x</sub> emissions, which is not allocated into different source sectors. Thus, if we assume that the ratios of every sectoral emissions to total anthropogenic emissions do not change between prior and posterior emission inventories, MIX-DE has an advantage for coalfired power sector, while NL-DE could benefit the downscaling for the industrial and residential sectors. In future work, we could optimize sectoral emissions rather than total anthropogenic emissions, and subsequently downscale posterior coal-fired power emissions through prior MIX coal-fired power emissions, and ultimately use VIIRS night time light data to downscale posterior industrial and residential emissions.

#### Reference

- Geng, G., Zhang, Q., Martin, R. V., Lin, J., Huo, H., Zheng, B., Wang, S., and He, K.: Impact of spatial proxies on the representation of bottom-up emission inventories: A satellite-based analysis, Atmos. Chem. Phys., 17, 4131-4145, 10.5194/acp-17-4131-2017, 2017.
- Ge, C., Wang, J., Reid, J. S., Posselt, D. J., Xian, P., and Hyer, E.: Mesoscale modeling of smoke transport from equatorial Southeast Asian Maritime Continent to the Philippines: First comparison of ensemble analysis with in situ observations, J. Geophys. Res., 122, 5380-5398, 10.1002/2016JD026241, 2017.
- Li, M., Liu, H., Geng, G., Hong, C., Liu, F., Song, Y., Tong, D., Zheng, B., Cui, H., Man, H., Zhang, Q., and He, K.: Anthropogenic emission inventories in China: a review, National Science Review, 4, 834-866, 10.1093/nsr/nwx150, 2017.

- Qu, Z., Henze, D. K., Theys, N., Wang, J., and Wang, W.: Hybrid Mass Balance/4D-Var Joint Inversion of NOx and SO2 Emissions in East Asia, J. Geophys. Res., 124, 8203-8224, 10.1029/2018JD030240, 2019.
- Xing, J., Mathur, R., Pleim, J., Hogrefe, C., Gan, C. M., Wong, D. C., Wei, C., Gilliam, R., and Pouliot, G.: Observations and modeling of air quality trends over 1990–2010 across the Northern Hemisphere: China, the United States and Europe, Atmos. Chem. Phys., 15, 2723-2747, 10.5194/acp-15-2723-2015, 2015.
- Zheng, B., Zhang, Q., Tong, D., Chen, C., Hong, C., Li, M., Geng, G., Lei, Y., Huo, H., and He, K.: Resolution dependence of uncertainties in gridded emission inventories: a case study in Hebei, China, Atmos. Chem. Phys., 17, 921-933, 10.5194/acp-17-921-2017, 2017.