Responses to the comments of Reviewer #2:

Thank you for your comments and suggestions. Based on your opinion, we have revised our manuscript comprehensively and carefully. The items lined out in your report are responded as follows:

Specific Comments:

1. Comparison of the one-step method with the two-step method.

Actually, the claim of the reviewer was not to support the one step method over the two-step method. There is no claim on my part that the one step method, as practiced in the authors' study, is to be preferred. The authors seem to take THEIR 1-step method as a benchmark, aimed to be outperformed by their preferred 2-step method. Rather, my claim is, that any coupled method, done properly should be superior to any method split in separate steps. This is simply a reference to numerical principles and results which give error estimates of operator split approaches. There is no evidence for the reader to understand why the two-step method should be better than a joint synthesized method. If the authors were claiming that numerical efficiency would be higher with the split method, then this would be an argument worth discussing. However, for a paper in a journal like GMD, I think there should at least be some explanation for the purported analytical superiority of the two-step method. So, if the reference is constructed with respect to the one-step method, could it be that the study implementation is still suboptimal and not using the full potential of a fully coupled system? So in summary, the authors have presented a comparison between the one and two-step methods, which is an internal comparison of their own implementations. In terms of the journal's objectives, this is not an indication of advances in model design for data assimilation in the sense of (GMD objectives citation):

• development and technical papers, describing developments such as new parameterizations or technical aspects of running models such as the reproducibility of results;

- new methods for assessment of models, including work on developing new metrics for assessing model performance and novel ways of comparing model results with observational data;

- papers describing new standard experiments for assessing model performance or novel ways of comparing model results with observational data

Therefore, what is presented is in fact only episodic evidence. If the proposed two-step procedure is superior to a properly unified coupled approach, the authors should be able to prove this on the more mathematically rigorous basis. This request of my previous review has not been responded in any resilient way.

Response: Thank you for this comment. In the previous round of revisions, we have modified the system to optimize both the initial conditions and emissions simultaneously within the unified framework of EnKF (fully coupled, one-step method), instead of separately optimizing the initial conditions and emissions in separate steps. In our two-step approach, it is not a separation of the aforementioned one-step method, but rather the first step focuses on optimizing emissions only, and in the second step, we introduce emission error propagation. We have also analyzed the superiority of the two-step approach in optimizing emissions from 1) the evaluation of the inversion results, 2) the OSSE experiment, and 3) the convergence after inversion with different priors.

We strongly agree with your comments, and we also recognize that the “one-step” and “two-step” methods actually have their own advantages. Theoretically, the “one-step” method can reduce the influence of the initial field error on the inversion results and ensure the accuracy of the inversion at each step; the “two-step” method makes the whole inversion process mass-conserving, the changes in concentration are all caused by the emission changes, and the inversions between adjacent windows can compensate each other. Therefore, it is not appropriate for us to try to show that the “two-step” method is better than the “one-step” method through sensitivity experiments, and it is difficult for us to give a theoretical (mathematical) proof. Therefore, in the revised

manuscript, we have removed the contents about the comparisons between the "one-step" and "two-step" approaches.

See lines 53, page 52; lines 162-165, page 6; lines 204-208, page 8; lines 556-561, page 23; lines 951-1015, pages 49-51; lines 1148-1151, pages 55-56.

2. Notational confusion between concentration and emission:

The authors' response cannot be correct because the following definition of X remains ambiguous: We find in the revised manuscript:

Line 417:" $\boldsymbol{\delta X_i^b}$ is the randomly perturbed samples that are added to the prior emissions $\boldsymbol{X_0^b}$ is …"

However, on the other hand,

Line 440: $\mathbf{y} - \boldsymbol{H}\overline{\boldsymbol{X^b}}$ reflects the differences between the simulated and observed concentrations.

In all related formulae in the paper, the notational use of $\boldsymbol{HX}$ follows the well-known standard, except for the extension of $\boldsymbol{X}$ and delta $\boldsymbol{X}$ to include emissions. Thus, contrary to the authors' response, it appears that vector $\boldsymbol{X}$ does also include concentrations. Please provide a clear notational definition of concentrations and emissions, and on what the observation operator H acts. Further, make clear which parameters are to be optimized in the definition, for discriminability.

Response: Thank you for this comment. In this study, treating emissions as state variables, so $\boldsymbol{X}$ solely consists of emissions. The observation operator $\boldsymbol{H}$ performs the mapping from model space to observation space. For optimizing concentrations, $\boldsymbol{H}$ can be directly applied to the simulated concentration field. However, when it comes to optimizing emissions, $\boldsymbol{H}$ also consists of the model integration process (Houtekamer and Zhang, 2016), which converts emissions $\boldsymbol{X}$ into concentrations. Similar formulations for emission inversion can also be found in other studies (Kong et al., 2021; Miyazaki et al., 2012; Tang et al., 2016).

The prior state $\overline{\boldsymbol{X^b}}$ is optimized, and the analyzed state $\overline{\boldsymbol{X^a}}$ is considered the best estimate of the emissions.

We have rephrased as follows, and see lines 369-382, pages 16-17 in the revised manuscript.

"Combined with the observational vector $\boldsymbol{y}$, the state vector $\overline{\boldsymbol{X^b}}$ was updated by minimizing the analysis variance… … where $\overline{\boldsymbol{X^b}}$ is the mean of the ensemble samples $\boldsymbol{X_i^b}$; $\boldsymbol{H}$ is the observation operator that maps the model space to the observation space, consisting of the model integration process converting emissions into concentrations and spatial interpolation matching the model concentration to the locations of the observations; $\boldsymbol{y} - \boldsymbol{H}\overline{\boldsymbol{X^b}}$ reflects the differences … …"

Houtekamer, P. L. and Zhang, F.: Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation, Monthly Weather Review, 144, 4489-4532, 2016.

Kong, L., Tang, X., Zhu, J., Wang, Z., Li, J., Wu, H., Wu, Q., Chen, H., Zhu, L., Wang, W., Liu, B., Wang, Q., Chen, D., Pan, Y., Song, T., Li, F., Zheng, H., Jia, G., Lu, M., Wu, L., and Carmichael, G. R.: A 6-year-long (2013-2018) high-resolution air quality reanalysis dataset in China based on the assimilation of surface observations from CNEMC, Earth System Science Data, 13, 529-570, 2021.

Miyazaki, K., Eskes, H. J., and Sudo, K.: Global NOx emission estimates derived from an assimilation of OMI tropospheric NO2 columns, Atmospheric Chemistry And Physics, 12, 2263-2288, 2012.

Tang, X., Zhu, J., Wang, Z., Gbaguidi, A., Lin, C., Xin, J., Song, T., and Hu, B.: Limitations of ozone data assimilation with adjustment of NOx emissions: mixed effects on NO2 forecasts over Beijing and surrounding areas, Atmospheric Chemistry and Physics, 16, 6395-6405, 2016.

3. Omission of the O_3 observation and the NOx-VOC-O_3 balance

The authors still refrain from assimilating ozone in order not to degrade the performance of their model. On the other hand, they claim that their method improves emission inventories, especially for nitrogen oxides. Again, the relationship between volatile organic compounds, nitrogen oxides and ozone is critical for photochemistry.

In particular, ozone is driven by emissions and should therefore be an important source of information on its emitted precursors. By neglecting ozone observations, the authors tacitly admit that the inversion system set-up is not capable of inferring emissions.

As they acknowledged, this may also be the result of a model resolution that is too coarse to account for the fine-scale point and line sources that make up the bulk of nitrogen oxide emissions. At the very least, this problem could be demonstrated to be manageable with a sufficiently high resolution simulation. Unfortunately, the authors have not presented a model run that takes this suggestion into account.

Response: Thank you for this comment. To demonstrate the limitations of assimilating O_3 data with adjustments of NO_x emissions, even at high resolutions of 3 km, we conducted an OSSE experiment. Considering the challenges of operationalizing at such high resolutions on a national scale, we focused research area over Hangzhou, a big city in East China. In the OSSE, a local high-resolution emission inventory was considered as a "true" emission, and the assimilated O_3 observations were simulated using this emission inventory. We then reduced the "true" emission by 40% to represent the a priori emissions. Cross-variable adjustment of the NO_x emissions was conducted through assimilating the artificial O_3 observations. After assimilation, the deviation in NO_x emissions was reduced in most regions (Figure R1c VS. Figure R1b). However, there is also a further decrease in NO_x emissions in many regions (Figure 1d), indicating an increase in emission deviations. The above results suggest that the adjustment directions of NO_x emissions at different places by the O_3 data assimilation are significantly different, because the response of O_3 on NO_x emissions changes is quite complex, even at high resolutions. To investigate the cause of observed negative effects, we conducted another OSSE experiment in a 'box' model to perform inversion on a representative grid with a negative adjustment. Figure R2 shows the changes in NO_x emissions before and after assimilating O_3. The "true" condition is VOC-limited, and the simulation with prior emissions overestimates O_3 significantly. The ensemble simulated concentrations with the corresponding ensemble emissions during DA show positive correlation (blue dots). This implies that, to match the observed O_3

concentrations (red dot), the system will adjust NO_x emissions in the direction indicated by this positive correlation. Therefore, despite the ensemble mean O_3 concentration after DA being closer to the observations, assimilating O_3 results in even lower NO_x emissions and larger errors. As explained in the previous revisions, improving the resolution can indeed provide better prior information on O_3-NO_x, but it is still challenging to determine whether the condition is NO_x-limited (positive correlation) or VOC-limited (negative correlation) in the real atmosphere. This limitations of assimilating O_3 to adjust NO_x have been studied in detail in Tang et al. (2016, ACP).

We have added related discussions in the revised paper (Lines 1084-1085, page 53). The changes are also listed as follows:

"Elbern et al. (2007) emphasized that assimilating O_3 to correct NO_x or VOC emissions must follow the EKMA framework derived based on observations, otherwise, even if the resolution is improved to sufficiently solve point and line sources, precursor emissions may be still adjusted in an opposite direction. This can be demonstrated in our OSSE experiment at high resolution of 3 km (Figure S11). … …"

Tang, X., Zhu, J., Wang, Z., Gbaguidi, A., Lin, C., Xin, J., Song, T., and Hu, B.: **Limitations of ozone data assimilation with adjustment of NOx emissions: mixed effects on NO2 forecasts over Beijing and surrounding areas**, Atmospheric Chemistry and Physics, 16, 6395-6405, 2016.

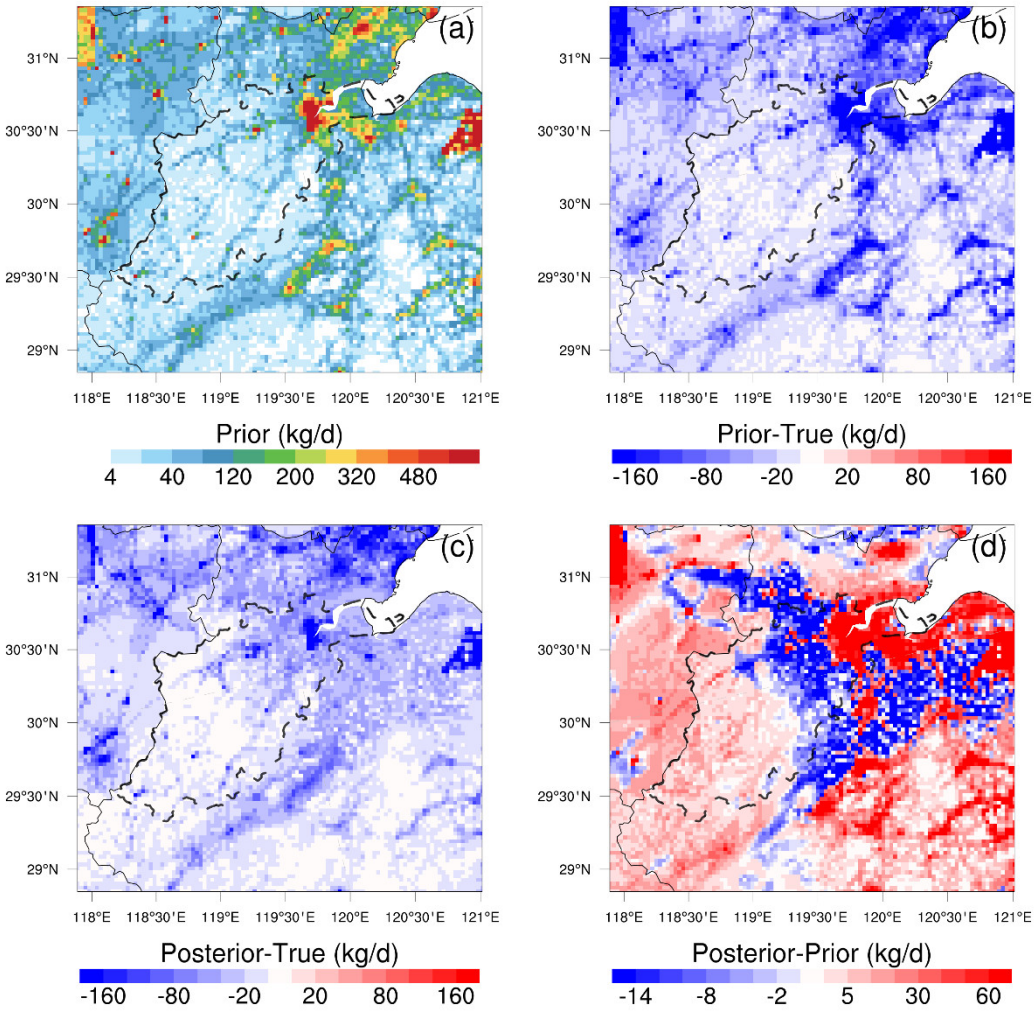

Figure R1 Spatial distribution of the prior, posterior, 'true' emissions and their differences (kg/d).

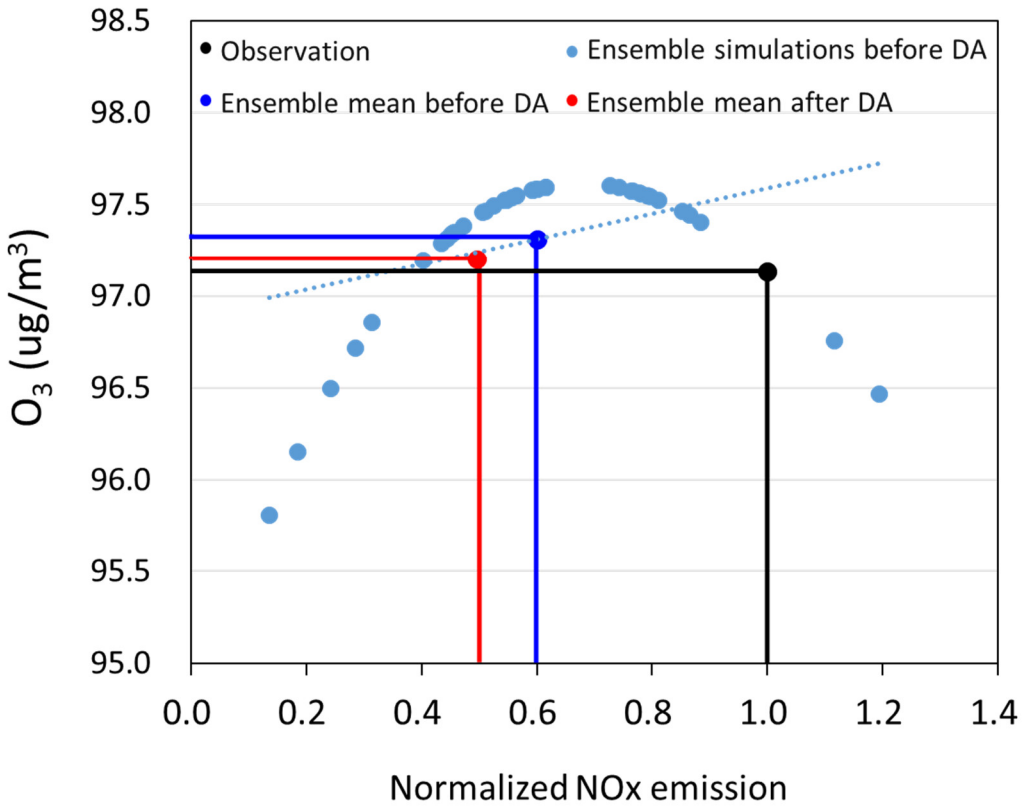

Figure R2 O_3 concentrations (ug/m^3) and normalized NO_x emissions before and after data assimilation.