Responses to the comments of Reviewer #1:

The authors have addressed my comments to my satisfaction. I cannot see any remaining substantial issues beyond the requirement for a thorough check of the English. There are many minor grammatical errors that do not affect the overall readability, but will need to be addressed. I would be happy to proofread the entire article again after the authors have addressed the remaining comments from reviewer 2.

Response: Thank you very much for your valuable comments and suggestions. We have checked through our paper and improved our English writing comprehensively and carefully in the revised manuscript.

Responses to the comments of Reviewer #2:

Thank you very much for your very valuable 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:

General comments:

A Regional multi-Air Pollutant Assimilation System (RAPAS v1.0) for emission estimates: system development and application by Shuzhuang Feng et al. has been re-evaluated. While a part of recommended modifications have been provided by the authors, there are still substantial questions, which either are not directly addressed (probably misunderstood) or insufficiently dispelled.

In summary (for more detail see items below) the presentation quality of the proposed two-step emission inversion method does not demonstrate sufficient mathematical rigour or convincing evidence of a favourable performance, which could be taken as a substantial contribution to scientific progress within the scope of GMD. Therefore the scientific approach and applied methods cannot be confirmed as valid.

The paper cannot be recommended to be published in its present form. As for the application results resting on case study simulations, a submission to a less algorithm or model design oriented journal as GMD may be considered.

Response: Thank you for this comment. In the revised manuscript, we have adopted the same approach of EnKF to compare the performances of the "one-step" and "two-step" schemes, and shown comprehensive evidences that the "two-step" scheme has better performance on the estimates of anthropogenic emissions. Details are given in the response of the specific comment 2.

The “one-step” scheme has been widely used in previous studies. Our finding shows that this scheme may cause considerable overestimates for emissions, especially for the species with longer atmospheric lifetime during the pollution period, which are mainly related to the avoidless bias in the initial field (the optimization of initial field using either 3DVar or EnKF cannot be perfect) and the lacks of compensation or trade-off mechanism. Therefore, we believe that it is a substantial contribution to the scientific progress.

Specific Comments:

1. Point 1: Literature survey is still substantially incomplete. The authors might wish to consider the following selection, which mostly share key objectives of their study.

Bocquet, M.: Parameter-field estimation for atmospheric dispersion: application to the Chernobyl accident using 4D-Var, QUARTERLY JOURNAL OF THE ROYAL METEOROLOGICAL SOCIETY, 138, 664–681, doi:10.1002/qj.961, 2012.

Bocquet, M. and Sakov, P.: Joint state and parameter estimation with an iterative ensemble Kalman smoother, NONLINEAR PROCESSES IN GEOPHYSICS, 20, 803–818, doi:10.5194/npg-20-803-2013, 2013.

Elbern, H., A. Strunk, H. Schmidt, and O. Talagrand, Emission rate and chemical state estimation by 4-dimensional variational inversion, ACP, 3749-3769, 2007.

Meirink, J. F., Bergamaschi, P., and Krol,M. C.: Four-dimensional variational data assimilation for inverse modelling of atmospheric methane emissions: method and comparison with synthesis inversion, ATMOSPHERIC CHEMISTRY AND PHYSICS, 8, 6341–6353, 2008.

Muller, J. F. and Stavrakou, T.: Inversion 10 of CO and NOx emissions using the adjoint of the IMAGES model, Atmos. Chem. Phys., 5, 1157–1186, 2005.

Navon, I. M.: Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography, Dyn. Atmos. Oceans, 27, 55–79, 1997.

Yumimoto, K., Uno, I., Sugimoto, N., Shimizu, A., Liu, Z., and Winker, D. M.: Adjoint inversion modeling of Asian dust emission using lidar observations, ATMOSPHERIC CHEMISTRY AND PHYSICS, 8, 2869–2884, 2008.

Response: Thanks for this suggestion. We have added related literatures in the revised manuscripts.

See lines 123-184, pages 5-7.

"… …Additionally, Jiang et al. (2017) and Stavrakou et al. (2008) also used the 4DVAR algorithm to estimate global CO and NO_x emission trends using MOPITT and GOME/SCIAMACHY retrievals, respectively. Using NIES LiDAR observations, Yumimoto et al. (2008) applied the 4DVAR DA to infer dust emissions over eastern Asia and the results agreed well with various satellite data and surface observations. Based on surface observations, Meirink et al. (2008) developed a 4DVAR system to optimize monthly methane emissions, which showed a high degree of consistency in posterior emissions and uncertainties when compared with an analogous inversion based on the traditional synthesis approach.

Although considerable progress has been made to reduce large uncertainties in emission inventories, the drawback of the 4DVAR method is the additional development of adjoint models, which are technically difficult and cumbersome for complex chemical transport models (Bocquet and Sakov, 2013). Instead, EnKF uses flow-dependent background error covariance… … Tang et al. (2011) adjusted the emissions of NO_x and VOCs through assimilating surface O_3 observations and achieved an better performance in O_3 forecasts. However, such a revision may encounter the problem of model error compensation rather than a retrieval of physically meaningful quantities, which should be avoided from overfitting for emission inversion purposes (Bocquet, 2012; Navon, 1998; Tang et al., 2011). The EnKF has also been widely applied to optimize emissions of carbon dioxide (Jiang et al., 2021; Liu et al., 2019), carbon monoxide (Feng et al., 2020a; Mizzi et al., 2018), sulfur dioxide (Chen et al., 2019), ammonia (Kong et al., 2019), etc.

Multi-species data assimilation can efficiently reduce the uncertainty in emission inventories... … Muller and Stavrakou (2005) also found that the simultaneous optimization of the sources of CO and NO_x led to better agreement between simulations and observations compared to the case where only CO observations are used.

The deviation in the chemical initial condition (IC) is an important source of error that affects the accuracy of emission inversion … …For example, Elbern et al. (2007) adjusted O_3 ICs, NO_x ICs and emissions, VOCs ICs and emissions jointly through assimilating surface O_3 and NO_x observations. Although the forecast skills of O_3 were improved, due to the coarse model resolution and the strong nonlinear relationship between O_3 and NO_x, the assimilation of O_3 observation worsened emission inversion and forecast of NO_x. Peng et al. (2018) assimilated near-surface observations ... …”

2. Point 3 a.: The authors were encouraged to provide mathematical facts on their two-step mixed method approach, or resort to a consistent method with verified BLUE property (Best Linear Unbiased Estimate). Yet, the two-step method, despite its single steps is not shown to have this property. There is no further evidence provided than ad hoc description based on a single case study, without sound comparison on the same weather case with other methods, not to mention any basic explanation. Rather, the authors offer no further supporting matter than multiple "we believe phrases"

"Although the biases in the high levels were not evaluated, with only ground observations, we believe that the performance of the EnKF method in the high levels is similar. .... Therefore, we believe that in this comparison, a combinatorial assimilation approach used in the "one-step" scheme is an acceptable approach ..."

"... we believe that the “two-step” scheme has better performance than the “one-step” scheme in emission inversions"

The latter claim of superiority is therefore not at all justified to be adopted as a step forward in aerosol inversion modelling.

Response: Thank you for this suggestion. We agree that the “two-step” procedure combining 3DVAR and EnSRF algorithm cannot ensure convergence toward the same result as the one-step procedure does. Therefore, we reconstructed our system based on EnSRF with simultaneous adjustment of initial conditions and emissions. We then compared the performances in emission inversion using the "two-step" and "one-step" schemes. In the “one-step” scheme, except for updating the initial field of each window based on EnSRF, other settings were the same as those in the “two-step” scheme. The spatial localization radius for updating initial conditions was set to 90 km in horizontal and lowest 5 layers in vertical. The selection of horizontal and vertical scales was similar to Kong et al. (2021) and Tang et al. (2016). Figure R1 showed the performance of optimization in initial field using the EnSRF method. The assimilation experiment (red dots) showed much better agreement with observations than the control experiment. Figure R2 also showed the comparisons of EnSRF (red) and 3DVAR analysis (blue) fields. Overall, the EnSRF and 3DVAR have similar performance in the optimization of initial fields. The difference between the two may be caused by the different background error construction methods in 3DVAR and EnKF.

We compared the performances of the "one-step" and "two-step" inversions from three aspects: 1) the evaluation of the inversion results, 2) the OSSE experiment, and 3) the convergence after inversion with different priors. Details are shown as follows:

1) Compared with “two-step” method (EMDA), the posterior emissions obtained from

the “one-step” method (EMS1) were increased by 7.9%, 9.6%, 2.7%, 27.1%, and 22.8% for CO, SO_2, NO_x, $PPM_{2.5}$ and PMC, respectively. The large emission increase was mainly distributed in the northern China (Figure R3). Overall, compared to the VEP, the RMSEs of CO, SO_2, NO_2, $PM_{2.5}$ and PMC in VEP1 increased from 0.56 mg m^{-3} and 17.7, 12.3, 29.6, and 24.6 μg m^{-3} to 0.58 mg m^{-3} and 18.3, 12.9, 34.9, and 25.9 μg m^{-3}, respectively. From the perspective of spatial distribution, the evaluation results become worse in areas where emissions increase (Figure R3). Additionally, it can be seen from the Figure R4 that the results of the VEP and VEP1 were relatively close at the beginning. However, in the heavy pollution (16–21 December) and later period, the VEP1 had higher concentrations than the observations and larger RMSE than VEP. The results verified against the independent sites showed a similar situation (Figure S8). The possible reasons of the worse performance are that: on the one hand, the inversion between each window in the "one-step method" is independent, and there is no compensation mechanism between windows (Figure R5); on the other hand, the assimilation for initial fields cannot be perfect (Figure R1). As shown in Figure R5, during the heavy pollution episode, there were negative biases in the optimized ICs every day, which lead to a larger positive and a smaller negative emission increment at a certain extent, and result in a larger emission in the end.

2) To remove the effect of this imperfect initial field, we conducted another OSSE experiment (OSSE_TRUEIC) using “one-step” scheme, in which the IC of each window was directly taken from the "true" simulation. We compared the emission error reductions between the OSSE experiment (Section 3) and the OSSE_TRUEIC experiment. The results show that during the last ten days, the error reductions of OSSE_TRUEIC were 70.7%, 78.6%, 73.3%, 72.4%, and 63.6% for CO, SO_2, NO_x, $PPM_{2.5}$, and PMC, respectively, which were smaller than those in the OSSE experiment (Figure R6), indicating that even with a perfect IC at each window, the inversion performance of “one-step” scheme is still not as good as that of the “two-step” method.

3) We also tested the convergence of the posterior emissions in the "one-step" inversion. Except for $PPM_{2.5}$, the relative differences of other species in posterior emissions were slightly larger than those inverted with the “two-step” scheme (Figure R7), which further underscores the advantages of the “two-step” scheme in emission inversion.

We have revised the contents related to the comparison between “one-step” scheme and “two-step” scheme. See Sec. 4.3.5.

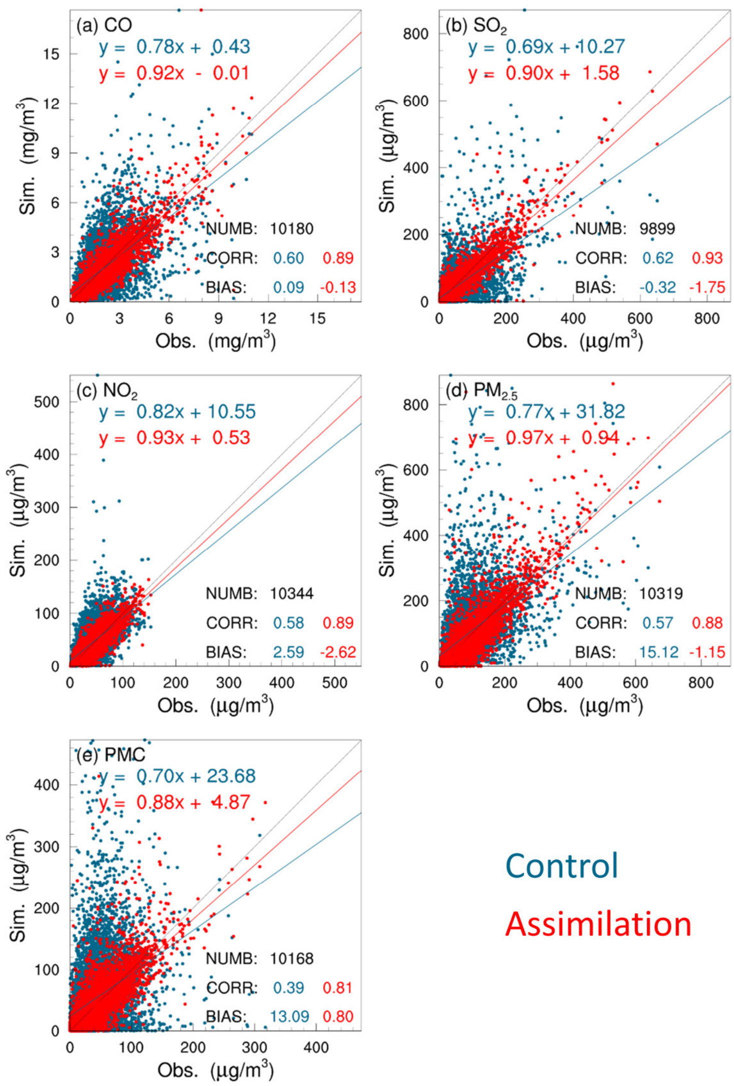

Figure R1. Scatter plots of simulated versus observed (a) CO, (b) SO_2, (C) NO_2, (d) $PM_{2.5}$, and (e) PMC mass concentrations aggregated over all 0000 UTC initializations in December from the background (blue) and EnSRF analysis (red) fields. (Figure S11 in the revised manuscript)

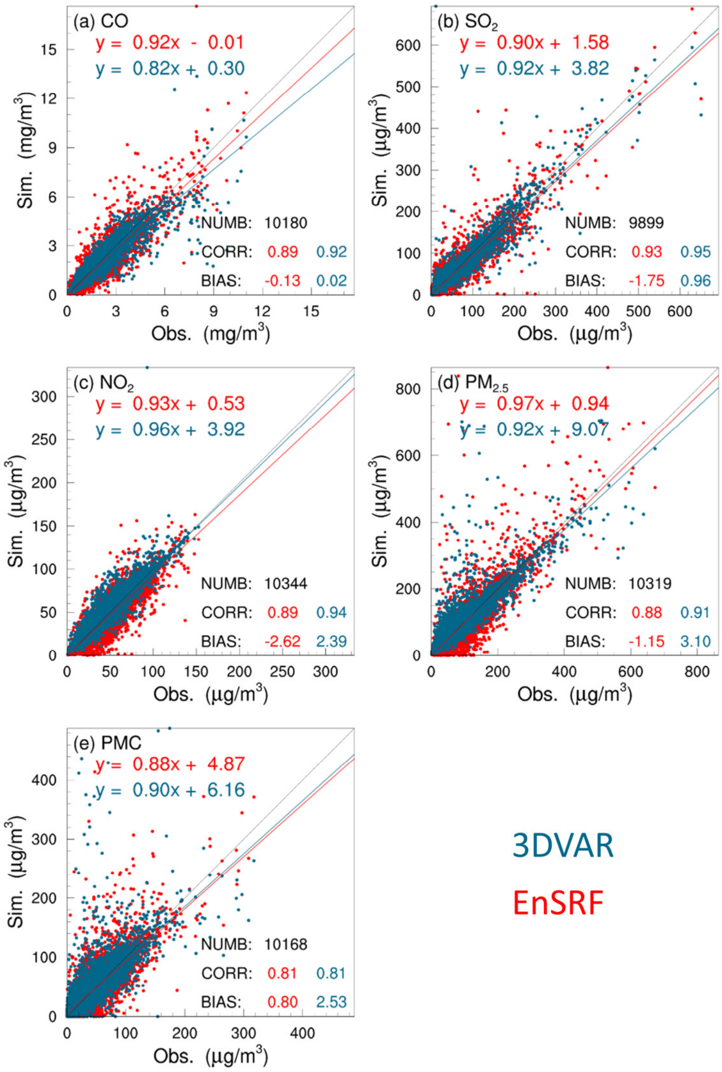

Figure R2. Scatter plots of simulated versus observed (a) CO, (b) SO_2, (C) NO_2, (d) $PM_{2.5}$, and (e) PMC mass concentrations aggregated over all 0000 UTC initializations in December from the EnSRF (red) and 3DVAR analysis (blue) fields.

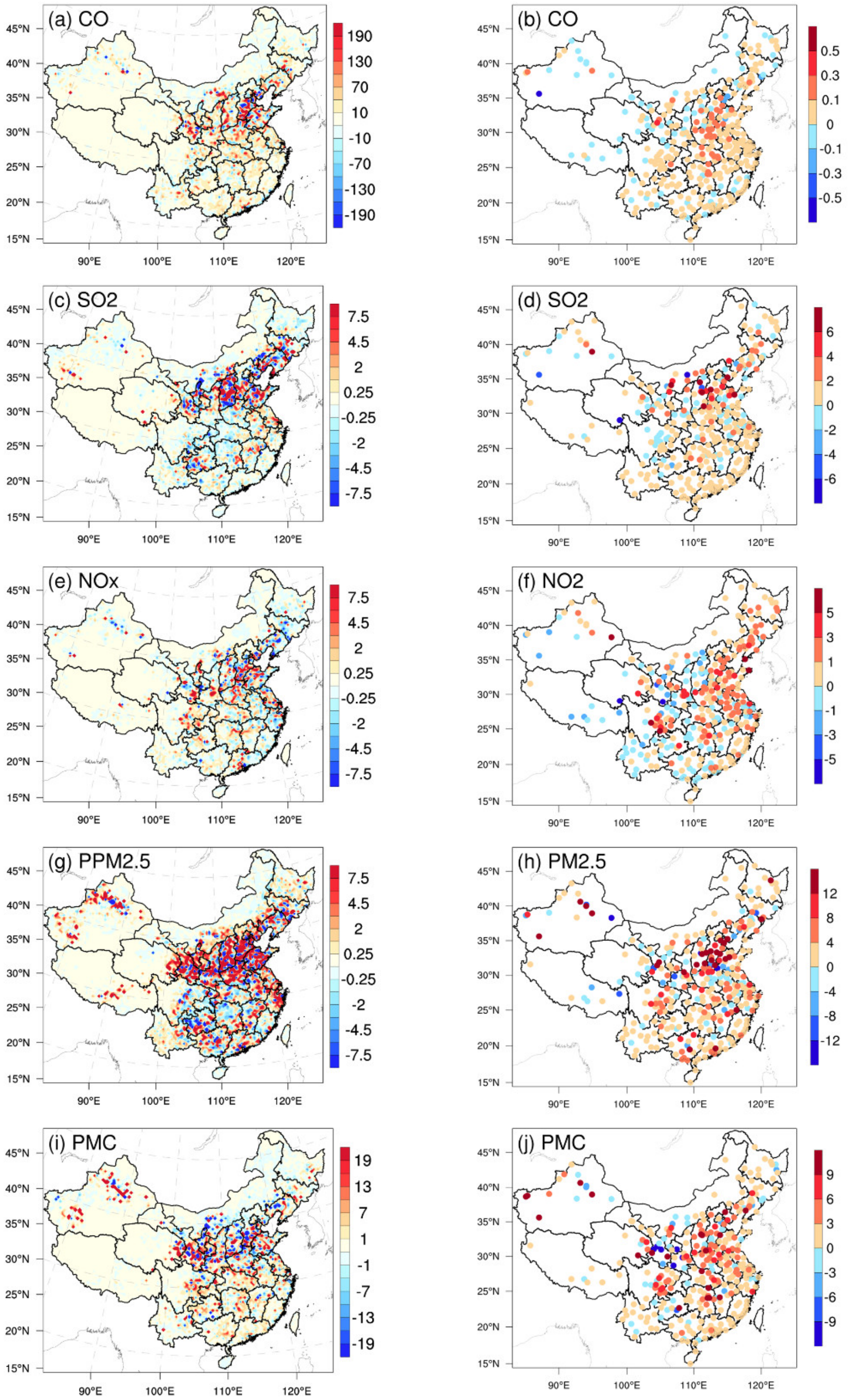

Figure R3. The differences of the posterior emissions between EMDA and EMS1 (left panel, EMS1-EMDA, %) and the differences of RMSE between VEP and VEP1 experiments (right panel, VEP1-VEP, CO: mg m^{-3}; others: μg m^{-3}). (Figure S13 in the revised manuscript)

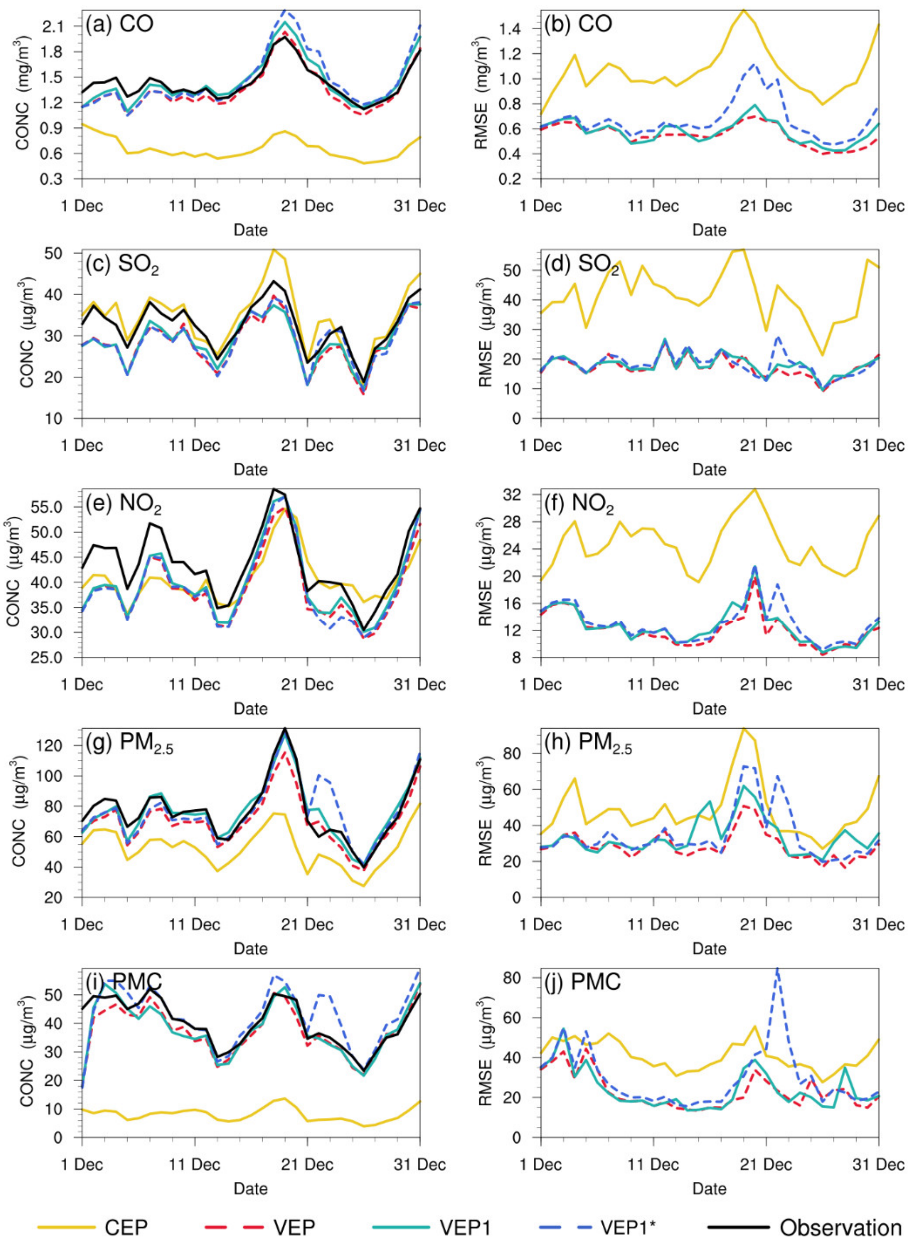

Figure R4. Time series of the daily concentrations (CONC, left) and root mean square error (RMSE, right) obtained from CEP, VEP (two-step), VEP1 (one-step based on EnSRF), and VEP* (one-step based on EnSRF and 3DVAR).

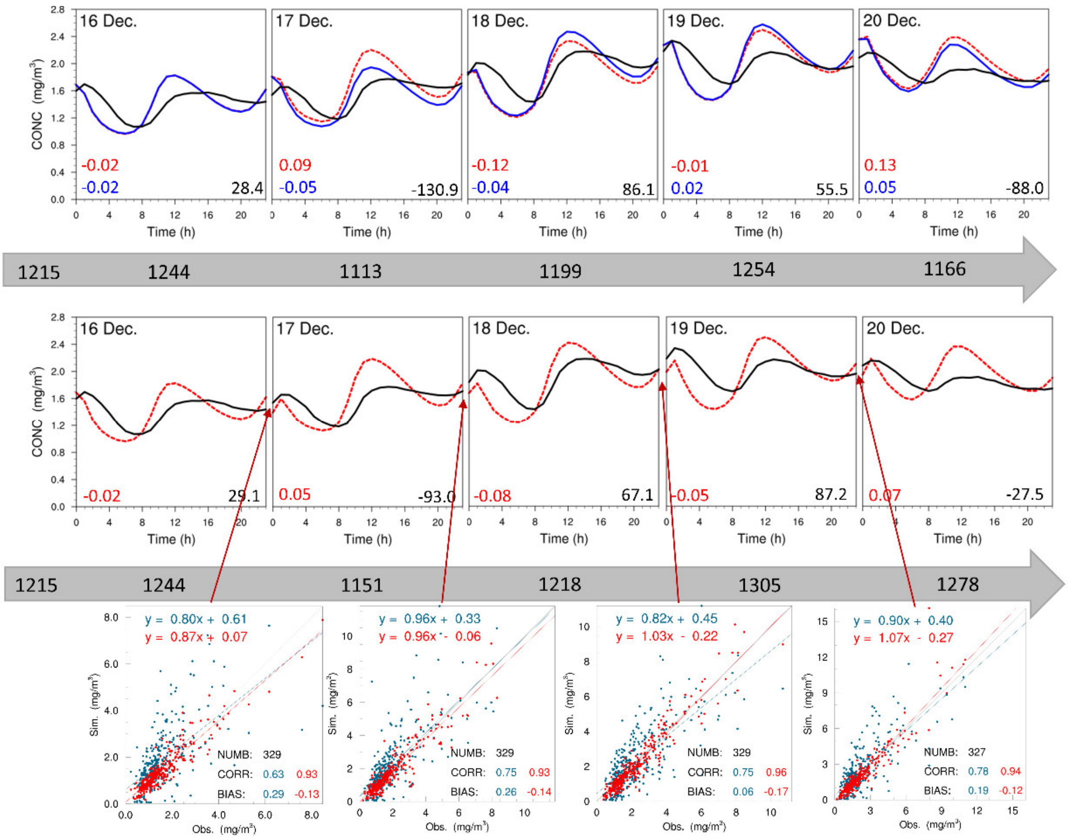

Figure R5. The difference in the daily simulation and inversion of CO between the "one-step" (bottom) and "two-step" (top) experiments. The red line, blue line and black line represent prior simulation, posterior simulation (forward simulation in the second step) and observation, respectively. Red and blue numbers represent biases of prior and posterior simulations, respectively. Black numbers represent emissions and changes. The original emissions (1215 kton/day) and ICs of the two experiments are derived from the inversion and forward simulation for December 15 in the EMDA experiment. The bottom row shows the simulated versus observed CO concentrations aggregated over 0000 UTC initializations from the background (blue) and analysis (red) fields. (Figure S14 in the revised manuscript)

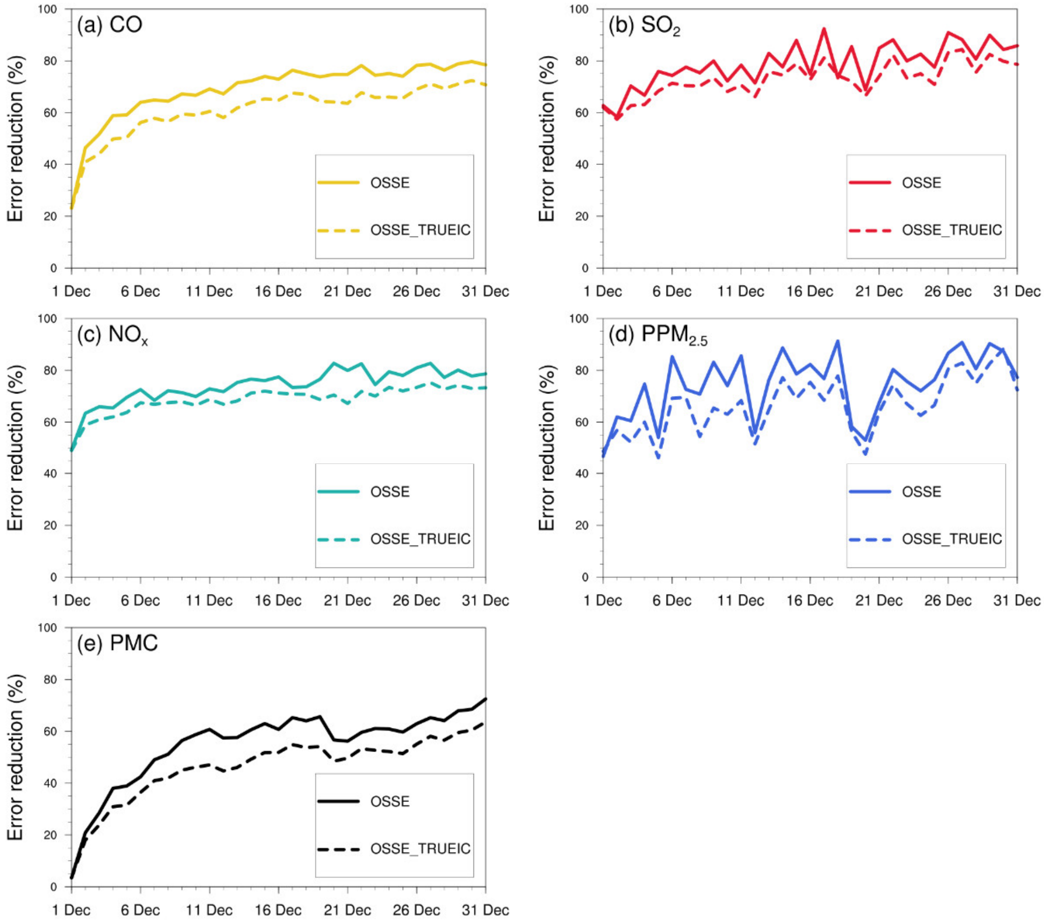

Figure R6. Comparison of emission error reduction between OSSE experiment and OSSE_TRUEIC experiment. (Figure S15 in the revised manuscript)

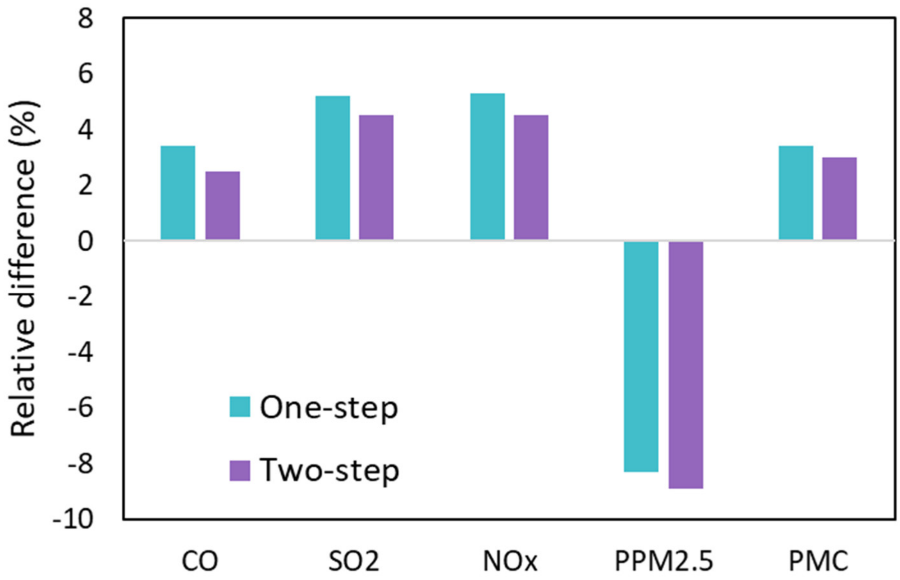

Figure R7. The relative differences (%) of posterior emissions with MEIC2016 and MEIC2012 as prior emissions. (Figure S16 in the revised manuscript)

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.

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. Point 4. The authors' response remains unclear. They have added " H reflects the combined information of emissions, the physics and chemistry processes in simulations and the transformation of different species from model space to observation space..." in phase space denoted X for the vector of concentrations. So H acts on X. The problem is not in understanding H. The confusion is whether it only applies to concentrations X, or also on emissions, where it remains from the first version: "delta X_b represents the randomly perturbed samples that are added to the prior emissions" So here it is emissions, not concentrations.

The authors should clarify the scope of X (and delta X), whether it comprises concentrations only (for 3D-var this is obvious), or emissions only, or both concentrations and emissions (feasible in EnKF). The explanation of H is clearly not required.

Response: Thank you very much for your suggestion. We have rephrased the sentence as follows (See lines 414-417, page 18):

"… inversion accuracy (Figure S1). In contrast to the estimation of parameters based on the augmentation of the conventional state vector (e.g. concentrations) with the parameter variables, $\boldsymbol{X}$ only comprises emissions in this study (similarly hereafter). $\delta\boldsymbol{X}_i^b$ is…"

4. Point 6: Authors' approach: "... O3 observations are not assimilated to improve NOx and VOC emissions using cross species information due to the strong nonlinear effects within the O3-NOx-VOC relationship,..."

Degradation of performance upon assimilation of additional data (here O3) is a typical indication of system inconsistency, which in this case is presumably the model resolution. NOx emissions are often point sources (industrial, domenstic) or line sources (traffic). In any case, sources are small compared to the model resolution and the assimilation degrades the simulation. The authors make reference to other studies. Yet in assimilation and ensuing forecasts, the proof of concept can be demonstrated, but is not done. Therefore the emission inversion cannot be improved! Therefore I uphold my recommendation to increase the resolution.

Response: Thanks for this comment. Yes, we strongly agree that the NO_x emissions are often point or line sources, which are all small compared to the model resolution. With a coarse spatial resolution, the model cannot accurately simulate the relationships between O_3 and its precursors, namely NO_x-limited, VOC-limited, or mixed-limited. When assimilating O_3 observations to infer NO_x or VOC emissions, the inaccurate relationships simulated by model would worsen the inversion of NOx emissions. In the case of high resolution emission inventory, improving the resolution can indeed improve the detailed simulation and provide better prior information on O_3-NO_x-VOC, but it is still difficult to determine whether the condition is NO_x-limited or VOC-limited in the real atmosphere using prior emissions (Liu and Shi, 2021). Elbern et al. (2007) also emphasized that assimilating O_3 to correct NO_x or VOC emissions must follow the EKMA framework (their Sec. 5.5), otherwise, even if the resolution is improved to sufficiently solve point and line sources, precursor emissions may be adjusted in the opposite direction.

In this study, the spatial resolutions of the prior emission inventory (MEIC) is 0.25° × 0.25°, which is appropriate for modeling at regional scales (Zheng et al., 2017). With this emission inventory, it is unable to accurately simulate the O_3-NO_x-VOC relationships. Therefore, to avoid the impact of inaccurate O_3-NO_x relationship on emission inversion, in our system, we did not assimilate O_3, but directly assimilate NO_2 to optimize the NO_x emissions. Although we do not assimilate O_3 observation, model resolution still has some influence on inversion results. In our previous study (Feng et al., 2022), we have inferred the NO_x emissions over Yangzte River Delta (YRD) in China using the NO_2 observations, which has a spatial resolution of 12 km. The study period, assimilated observations, and inversion settings are the same as this study. We compared the posterior emissions of YRD between this study and Feng et al. (2022). The results showed that the spatial distribution of the posterior emissions inferred using these two resolutions (36 km vs. 12 km) were similar (Figure R8), but the total NO_x emission in YRD inferred using 36 km resolution was about 8.8% higher than that inferred using 12 km resolution. The differences are mainly caused by meteorological differences at different resolutions. With this spatial resolution (36 km), we also have successfully inferred the NO_x emission changes in China during the COVID-19 Epidemic, and the spatiotemporal variation of inverted emissions were consistent with the lockdown policies conducted over China (Feng et al., 2020).

We have rephrased that paragraph as follows, and see lines 1204-1252, pages 59-60 in the revised manuscript.

"… …In addition, O_3 observations were not assimilated to improve NO_x and VOC emissions using cross-species information. O_3 concentration and NO_x (VOC) emissions were positively correlated in the NO_x (VOC)-limited region and negatively correlated in the VOC (NO_x)-limited region (Tang et al., 2011; Wang et al., 2019b). Hamer et al. (2015) successfully used O_3 observations to estimate NO_x and VOC emissions within the 4DVAR framework within an ideal model. However, the NO_x emissions are often point or line sources, which are all small compared to the model resolution. With a coarse spatial resolution, the model cannot accurately simulate the relationships between O_3 and its precursors. When assimilating O_3 observations to infer NO_x or VOC emissions, the inaccurate relationships simulated by model would worsen the inversion of NO_x emissions (Inness et al., 2015). In general, improving the model resolution can improve the detailed simulation and provide better prior information on O_3-NO_x-VOC, but it is still difficult to determine whether the condition is NO_x-limited or VOC-limited in the real atmosphere using prior emissions. 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. In this study, the spatial resolutions of the prior emission inventory (i.e., MEIC) is $0.25° \times 0.25°$, which is appropriate for modeling at regional scales (Zheng et al., 2017). With this emission inventory, it is unable to accurately simulate the O_3-NO_x-VOC relationships. Therefore, to avoid the impact of inaccurate O_3-NO_x relationship on emission inversion, in our system, we did not assimilate O_3, but directly assimilate NO_2 to optimize the NO_x emissions. This work will be followed by an ongoing study using the available VOC observations.

Although we do not assimilate O_3 observation, model resolution still has some influence on inversion results. In our previous study (Feng et al., 2022), we have inferred the NO_x emissions over YRD in China using NO_2 observations, which has a spatial resolution of 12 km. The study period, assimilated observations, and inversion settings are the same as this study. We compared the posterior emissions of YRD between this study and Feng et al. (2022). The results showed that there was similar spatial distribution of posterior emissions inferred using the two resolutions (36 km vs 12 km) (Figure R17), but the total NO_x emission in YRD inferred using 36 km resolution was about 8.8% higher than that inferred using 12 km resolution. The differences are mainly caused by meteorological differences at different resolutions. This indicates that coarse model resolution may lead to some overestimation of the inverted emissions. In addition, as shown previously, the concentrations after DA … …"

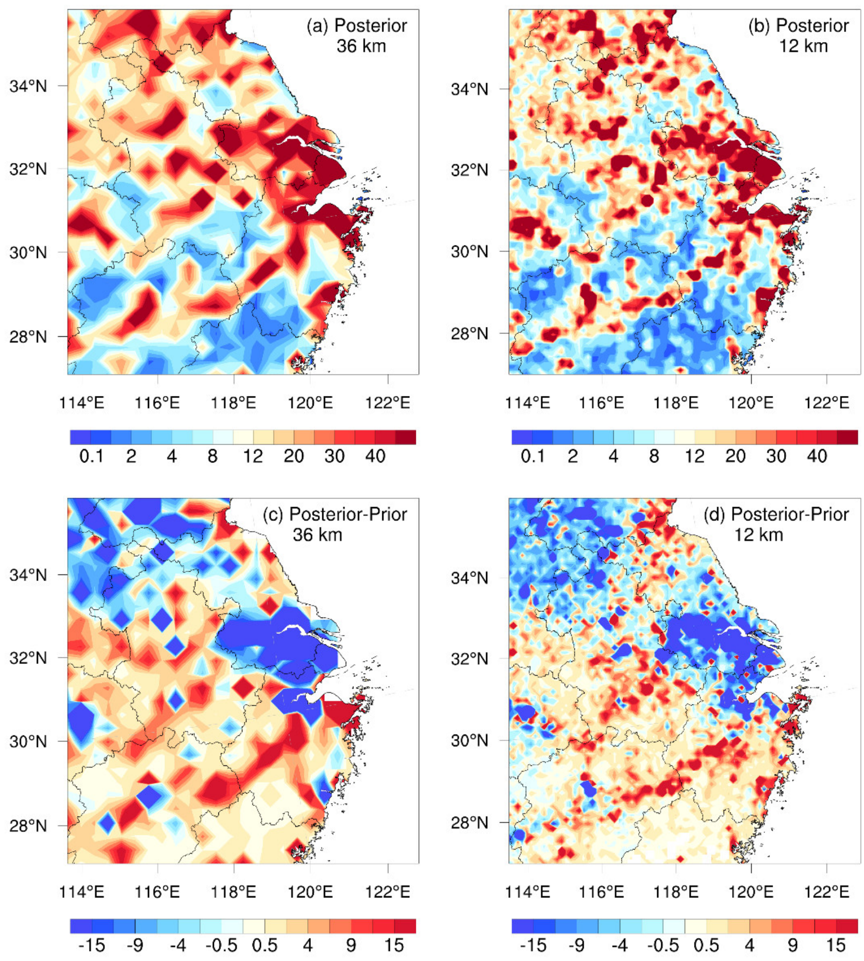

Figure R8. Spatial distribution of the time-averaged posterior emissions (kg $day^{-1}km^{-2}$) and differences between posterior and prior emissions (posterior minus prior) at 36 and 12 km resolutions. (Figure S17 in the revised manuscript)

Elbern, H., Strunk, A., Schmidt, H., and Talagrand, O.: Emission rate and chemical state estimation by 4-dimensional variational inversion, Atmospheric Chemistry and Physics, 7, 3749-3769, 2007.

Feng, S., Jiang, F., Wang, H., Shen, Y., Zheng, Y., Zhang, L., Lou, C., and Ju, W.: Anthropogenic emissions estimated using surface observations and their impacts on PM2.5 source apportionment over the Yangtze River Delta, China, Science of The Total Environment, 828, 154522, 2022

Feng, S., Jiang, F., Wang, H., Wang, H., Ju, W., Shen, Y., Zheng, Y., Wu, Z., and Ding, A.: NOx Emission Changes Over China During the COVID-19 Epidemic Inferred From Surface NO2 Observations, Geophysical Research Letters, 47, 2020.

Liu, C. and Shi, K.: A review on methodology in O3-NOx-VOC sensitivity study, Environmental Pollution, 291, 118249, 2021.

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, Atmospheric Chemistry and Physics, 17, 921-933, 2017.