Responses to the comments of Reviewer #2:

We are truly grateful to the positive comments and thoughtful suggestions. Those comments are all valuable and very helpful for revising and improving our paper, as well as the important guiding significance to our research. We have studied the comments/suggestions carefully and have made corresponding corrections. All changes made to the text are marked in red color. The point-by-point responses to the reviewers' comments are listed as follows:

General comments:

1. The study "A Regional multi-Air Pollutant Assimilation System (RAPAS v1.0) for emission estimates: system development and application by Shuzhuang Feng and colleagues describes a proposed method to estimate chemical and particulate matter emissions from observations by inversion. Emission rates of trace gases both of biogenic sources and anthropogenic chemical pollutants cannot be measured directly (apart from very special cases). Yet by evident reasons a precise knowledge is utterly important. While it is a long-discussed topic in atmospheric chemistry data assimilation and inversion, a solution of which is hampered by other factors, as initial and boundary data, deposition rates and many other parameters exert significant influence on the simulation result of chemistry transport models (CTM), which are the link between observations and emission estimates.

As their central objective the authors claim this study offer "a useful tool for accurately quantifying multi-species anthropogenic emissions at large scales and near-real time, which will serve better for monitoring emission changes and designing future emissions regulations and pollution control."

While the aspirations of the authors are remarkable, I could not find sound justifications for both a validated methodological approach and the practical proof of concept by the presented case study. I cannot recommend publication prior to a presentation of supporting evidence of the central claim.

Response: Thank you very much for this comment. We strongly agree that the conclusion that the system can reveal emissions changes well and provide effective assistance for pollution control can only be concluded after sufficient validation of the spatial and temporal changes in emissions inferred from the system. In this study, several methods were adopted to evaluate our system and the inversion results.

First, the uncertainty reduction rate (UR) is an important quantity to evaluate the performance of an assimilation system and the effectiveness of assimilating observations (Chevallier et al., 2007; Takagi et al., 2011). In this system, the overall UR of CO, SO2, NOx, PPM2.5 and PMC emissions were 44.4%, 45.0%, 34.3%, 51.8% and 56.1%, respectively, indicating that our system has good and comparable performance in optimizing the emissions of the 5 species.

Second, an Observing Systems Simulation Experiment (OSEE) was conducted to evaluate the performance of the RAPAS system. We used the MEIC 2016 inventory as a "true" emission, and reduced the "true" emission by 30% over the mainland of China as a prior emission. Overall, the errors of posterior emissions of CO, SO2, NOx, PPM2.5 and PMC over the mainland of China can be reduced by 78.4%, 86.1%, 78.8%, 77.6%, and 72.0%, respectively. Results of the OSSE show that the RAPAS can significantly reduce errors in prior emissions, and the method adopted in the RAPAS is reasonable and feasible.

Third, to diagnose the performance of the EnKF analysis, the chi-squared ($\chi 2$) statistics was also calculated. Since we imposed a same uncertainty of prior emission at each DA window to partially compensate for the influence of model errors, $\chi 2$ statistics showed small fluctuations, indicating that the system updates emissions consistently and stably.

Fourth, as only organized emissions from stacks have direct online monitoring in China, direct validation of the optimized emissions is impossible, and instead, we indirectly evaluate the posterior emissions by comparing the forward simulated atmospheric mixing ratios against measurement. The results showed that the RMSEs of the simulated concentrations with posterior emissions decreased by 40.1-56.3%, and the

CORRs increased from 0.26-0.66 to 0.69-0.87 for different species.

Fifth, according to the suggestion of another reviewer, the weekend effect of emissions was examined in the revised manuscript, which is a good way to test the reasonability of the day-to-day variation of the inverted emissions. But unfortunately, the weekend effect of China's emissions is not significant (Wang et al., 2014; Wang et al., 2015). We examined the day-to-day variations of the posterior NOx emissions, which is mainly emitted by transportation (Li et al., 2017), and found that there is no significant weekend effect, but the daily variation of the posterior emissions is in good agreement with the observations.

Finally, the emission changes of NOx during the 2020 COVID-19 Epidemic have been widely studied with satellite observations (Bauwens et al., 2020; Liu et al., 2020), and were relatively well known because of the lockdown policy. In our previous study (Feng et al., 2020), with the RAPAS system, we also estimated the spatial and temporal changes of the NOx emissions, and found that the inverted daily variations of NO2 emissions in different areas of China were consistent with their lockdown policies.

In this study, we further extended our system to synchronously assimilate the conventional ground observations of SO2, CO, NO2, PM2.5 and PM10 and infer the SO2, CO, NOx, primary PM2.5 and coarse PM10 emissions. Since the observations are released by the National Environmental Monitoring Station of China in real time, with this system, near-real time emissions could be estimated. Therefore, we believe that this system could serve as a useful tool for accurately quantifying the changes of anthropogenic emissions at near-real time, which will be helpful for the air pollution control in China, and the other regions around the world with ground observation networks.

Bauwens, M., Compernolle, S., Stavrakou, T., Müller, J.-F., van Gent, J., Eskes, H., Levelt, P. F., van der A, R., Veefkind, J. P., Vlietinck, J., Yu, H., and Zehner, C.: Impact of Coronavirus Outbreak on NO2 Pollution Assessed Using TROPOMI and OMI Observations, 47, e2020GL087978, 10.1029/2020gl087978, 2020.

Chevallier, F., Bréon, F.-M., and Rayner, P. J.: Contribution of the Orbiting Carbon Observatory to the estimation of CO2 sources and sinks: Theoretical study in a variational data assimilation framework, 112, 10.1029/2006JD007375, 2007.

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, 10.1029/2020gl090080, 2020.

Li, M., Zhang, Q., Kurokawa, J.-i., Woo, J.-H., He, K., Lu, Z., Ohara, T., Song, Y., Streets, D. G., Carmichael, G. R., Cheng, Y., Hong, C., Huo, H., Jiang, X., Kang, S., Liu, F., Su, H., and Zheng, B.: MIX: a mosaic Asian anthropogenic emission inventory under the international collaboration framework of the MICS-Asia and HTAP, Atmospheric Chemistry and Physics, 17, 935-963, 10.5194/acp-17-935-2017, 2017.

Liu, F., Page, A., Strode, S. A., Yoshida, Y., Choi, S., Zheng, B., Lamsal, L. N., Li, C., Krotkov, N. A., Eskes, H., van der A, R., Veefkind, P., Levelt, P. F., Hauser, O. P., and Joiner, J.: Abrupt decline in tropospheric nitrogen dioxide over China after the outbreak of COVID-19, Science Advances, 6, 2020.

Takagi, H., Saeki, T., Oda, T., Saito, M., Valsala, V., Belikov, D., Saito, R., Yoshida, Y., Morino, I., Uchino, O., Andres, R. J., Yokota, T., and Maksyutov, S.: On the Benefit of GOSAT Observations to the Estimation of Regional CO₂ Fluxes, SOLA, 7, 161-164, 10.2151/sola.2011-041, 2011.

Wang, Y. H., Hu, B., Ji, D. S., Liu, Z. R., Tang, G. Q., Xin, J. Y., Zhang, H. X., Song, T., Wang, L. L., Gao, W. K., Wang, X. K., and Wang, Y. S.: Ozone weekend effects in the Beijing-Tianjin-Hebei metropolitan area, China, Atmospheric Chemistry and Physics, 14, 2419-2429, 2014.

Wang, Z., Li, Y., Dong, X., Sun, R., Sun, N., and Pan, L.: Analysis on weekend effect of air pollutants in urban atmosphere of Beijing, Journal of University of Chinese Academy of Sciences, 32, 843-850, 2015.

Specific comments:

1. The introduction comprises related literature on emission inversion, however with a clear bias toward Kalman filtering. The variational approach, which plays a key role in radiatively active trace gases for greenhouse effects, is also used in reactive chemistry inversion. Yet this methodological branch is disposed of by remarking that this is “technically difficult and cumbersome for complex chemical transport models”,

without addressing its features. Literature survey may be complemented to a appropriate level here.

Response: Thank you for this suggestion. We supplement some literature reviews on the characterization of the 4D-var method and related inversion studies in the revised manuscript. See lines 108-119, pages 4-5.

"4DVAR provides a global optimal analysis through minimizing a cost function. It shows implicit flow-dependent background error covariance and can reflect complex nonlinear constraint relationship (Lorenc, 2003). Additionally, the model error can be partly accounted for with a weak constraint 4DVAR method through the definition of a systematic error term in a cost function (Derber, 1989). For example, GEOS-Chem and TM5 4DVAR frameworks have been used to estimate CH_4 (Alexe et al., 2015; Schneising et al., 2009; Stanevich et al., 2021; Wecht et al., 2014) and CO_2 fluxes (Basu et al., 2013; Nassar et al., 2011; Wang et al., 2019) from different satellite retrieval products. Monteil et al. (2013) showed that the global patterns of CH_4 emissions derived from SCIAMACHY (with bias correction) and GOSAT retrievals are in remarkable agreement based on 15 months observations. Additionally, Jiang et al. (2017) used 4DVAR algorithm to estimate global CO emission trends from 2000–2015 using MOPITT retrievals. Kurokawa et al. (2009) and Stavrakou et al. (2008) …"

2. A central position next to emission optimization is initial value optimization as a prerequisite for unbiased emission rate estimates. A discussion on sensitivities of CTMs to additional parameter controls (see above), which demonstrate the authors' awareness of competing impacts on model simulations is lacking however. As it is presented, it is tacitly assumed that these sensitivities are minor. In this context, the authors' approach may well be valid. But it is a matter of scientific scrutiny to meticulously expose the underlying assumptions, which would not affect the value of the study.

Response: Thank you very much for this comment. We fully agree that the model-data mismatch error comes not only from the emissions, but also from the inherent model

errors arising from model structure, discretization and parameterizations, such as boundary condition error, deposition rates, lack of chemical reactions and processes of gas-particle transformation, etc. At present, there is still a lack of reasonable and effective algorithms to solve the model error in atmospheric data assimilation (Houtekamer and Zhang, 2016). Usually, it was assumed that the atmospheric chemical transport model is perfect. Neglecting model errors in the EnKF would attribute all uncertainties to emissions. We have added the following discussions about model error in the revised manuscript, which could also be found in lines 1096-1129, pages 58-59.

"The model-data mismatch error not only comes from the emissions, but also from the inherent model errors arising from model structure, discretization, parameterizations and the biases in the simulated meteorological fields. Neglecting model errors would attribute all uncertainties to emissions, and lead to considerable biases in the estimated emissions. In the version of CMAQ model used in this study, there is no heterogeneous reactions (Quan et al., 2015; Wang et al., 2017), the parameterization scheme for the formation of secondary organic aerosol (SOA) is imperfect (Carlton et al., 2008; Jiang et al., 2012; Yang et al., 2019), no feedback between chemistry and meteorology is considered, and we used an idea profile for chemical lateral boundary conditions. All of the above problems can lead to underestimated concentrations of pollutants, which in turn require more emissions to compensate, leading to overestimations in emissions. In addition, previous studies have shown that the emission of ammonia in the MEIC inventory was underestimated (Kong et al., 2019b; Paulot et al., 2014; Zhang et al., 2018). Due to lack of ammonia observations, our system does not include emission estimates of ammonia, which means that the concentration of ammonium aerosol was underestimated in this system, also resulting in an overestimation in the $PPM_{2.5}$ emission. Wind-blown dust was also not simulated here, thus the PMC emission inverted in this system do not only come from anthropogenic activities, but also from natural sources. Although some of these shortcomings could be solved in the future by updating the CTM model, there will still be errors in each parameterization and each process. Generally, parameter estimation method was used to reduce the model errors,

in which, some uncertain parameters were included in the augmented state vector and were optimized synchronously based on the available observations (Brandhorst et al., 2017; Evensen, 2009). However, it is still quite difficult to identify the key uncertain parameters of different species in different models, which generally comes not only from the complex atmospheric chemical model, but also from hundreds of model inputs (Tang et al., 2013). Another method is bias correction, which treats the model error as a bias term, and includes it in the augmented state vector (Brandhorst et al., 2017; De Lannoy et al., 2007; Keppenne et al., 2005). In addition, the weak-constraint 4D-Var method can also be used to reduce the model errors, which adds a correction term in the model integration to account for the different sources of model error (Sasaki, 1970). Although reliable diagnosis of model error is still a challenge at present (Laloyaux et al., 2020), it should be considered in an assimilation system. We will consider model errors in our system in the future to obtain better emission estimates."

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

Methodology comments:

3.a. Inversion and data assimilation methods perceived in the referenced literature address Best Linear Unbiased Estimators (BLUE). This property should also be granted by the two-step procedure combining three-dimensional variational (3DVAR) algorithm and the ensemble square root filter (EnSRF) algorithm, which is claimed to show "that the "two-step" scheme clearly outperformed the simultaneous assimilation of ICs and emissions ("one-step" scheme), …" (line 72). While splitting tasks into sequential steps is a well established strategy in numerics to reduce complexity and increase efficiency (e.g. ADI methods), essential care must be taken to ensure the convergence of both split (two-step) and combined (one-step) algorithms toward the

same result. The claimed robustness of the method is demonstrated here not in a sense of same result. Rather different background emissions are probed for convergence (e.g. line 153), which does not experimentally validate the two-step method. Moreover, it is a problem, that the two-step method is established by two different methods. Unless there is a sophisticated multivariate (that is multi-species crosscorrelating) background error covariance matrix (which is evidently not the case in this study) initial value estimation by 3DVAR cannot account for some realistic chemical consistency. Even monovariate cross-correlations are essential to be defined, both horizontally and vertically, to optimize unobserved adjacent areas/height levels. Any estimation errors made in step one are compensated by step two provided biased emission optimization, which is, by EnKF, also performed in time. (Say, too high/low estimated vertical concentrations aloft and mixed down to the observation site are adjusted by too low/high emissions). So the system is prone to propagate errors made in a single step, the methodologically weakest of which limits the overall success. So better analyse why (see L 71-72) it can be possible that "the "two-step" scheme clearly outperformed the simultaneous assimilation of ICs and emissions ("one-step" scheme)".

The authors should either proof the BLUE property of their two-step method with a unique solution by mathematical rigour, or adopt a one-step procedure based on their EnKF approach, or even better, an ensemble Kalman Smoother.

Response: Thank you for this comment. We want to prove the superiority of the system in the inversion strategy ("two-step") by comparing the performances of the "two-step" and "one-step" schemes. The latter has been widely used in previous studies (Miyazaki et al., 2017; Peng et al., 2018). However, in this study, for the "one-step" scheme, we use a combination assimilation method, namely 3DVAR for the optimizations of initial fields and EnKF for emission inversions in each DA window, which is similar as Jiang et al., (2017), but different from most previous studies. Because most previous "one-step" assimilation studies used only one method (i.e., EnKF). We agree that this combination method may cause the comparison less than perfect. However, It should be noted that, even using the same method (such as EnKF) to optimize the emission of

the current window and the initial field of the next window simultaneously (Peng et al., 2018), the initial field estimation errors will still be mixed in the simulated concentration field, resulting in unreasonable emission compensation in the next window. In “one-step” scheme, the essence is to build a good initial field through data assimilation for each DA window. The 3DVAR method used in this study has good performance in optimizing the initial field, which has been shown in section 4.1.2. 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. Schwartz et al. (2014) compared the performances of EnKF and 3DVAR in optimizing initial fields, and found that 3DVAR method can obtain a better initial field than EnKF method. Therefore, we believe that in this comparison, a combinatorial assimilation approach used in the "one-step" scheme is an acceptable approach.

During the comparison, three results were analyzed, including the evaluation of posterior emissions against observations, the daily variations of the posterior emissions, and the convergence performances with different prior emissions. The evaluation against observations shows that the RMSEs of the simulated concentrations with posterior emissions decreased by 40.1-56.3%, and the CORRs increased from 0.26-0.66 to 0.69-0.87 for different species; The daily variations show that the posterior emissions is in good agreement with the observations; and the convergence performances with different prior emissions show that the differences between the posterior emissions gradually decrease over time in the "two-step" scheme, but it is not found in the "one-step" scheme. Therefore, we believe that the “two-step” scheme has better performance than the “one-step” scheme in emission inversions.

Following descriptions are added in the revised manuscripts.

Lines 1046-1048, page 56.

“Overall, there is no significant difference between the two methods for NOx and SO2, but for CO, it can be clearly seen that the difference increases with the inversion (Figure S11).”

Lines 1072-1079, page 57.

“On the contrary, this overestimation will be corrected quickly in the subsequent inversion using the “two-step” inversion scheme in this study (Figure S11), so as to ensure the stability of the system. Additionally, the other “one-step” experiment, taking MEIC 2012 as prior emissions, was conducted. However, the relative differences (Figure S12) in posterior emissions between this experiment and the EMS1 did not converge like that between EMDA and EMS2 with “two-step” scheme (Figure 13), which further demonstrates the advantages of the “two-step” scheme.”

Lines 1163-1179, pages 60-61.

“When comparing the performances of the “two-step” and “one-step” schemes, for the “one-step” scheme, we use a combination assimilation method, namely 3DVAR for the optimizations of initial fields and EnKF for emission inversions in each DA window, which is similar as Jiang et al., (2017), but different from most previous studies (Miyazaki et al., 2017; Tang et al., 2013). Because most previous “one-step” assimilation studies used only one method (i.e., EnKF). This combination method may cause the comparison less than perfect. However, it should be noted that, even using the same method (such as EnKF) to optimize the emission of the current window and the initial field of the next window simultaneously (Peng et al., 2018), the initial field estimation errors will still be mixed in the simulated concentration field, resulting in unreasonable emission compensation in the next window. In “one-step” scheme, the essence is to build a good initial field in the high levels. Schwartz et al. (2014) compared the performances of EnKF and 3DVAR in optimizing initial fields, and found that 3DVAR method can obtain a better initial field than EnKF method. Therefore, we believe that in this comparison, a combinatorial assimilation approach used in the "one-step" scheme is an acceptable approach, and the conclusion is credible, that the “two-step” scheme has better performances than the “one-step” scheme in emission estimates.

”

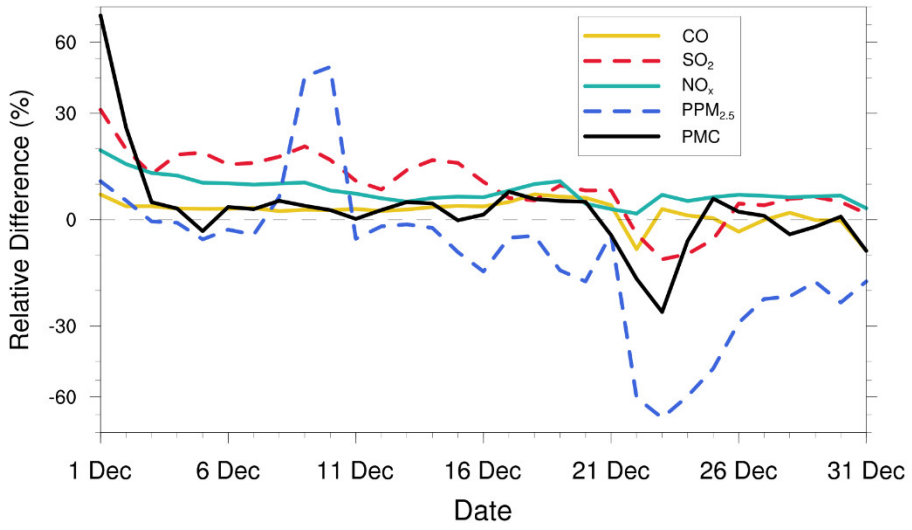

Figure R1. Relative differences in CO, SO2, NOx, PPM2.5 and PMC emissions (%) between the EMS1 and the other "one-step" experiment taking MEIC 2012 as prior emissions. Note that the X-axis scale is different from Figure 13. (Figure S12 in the revised manuscript)

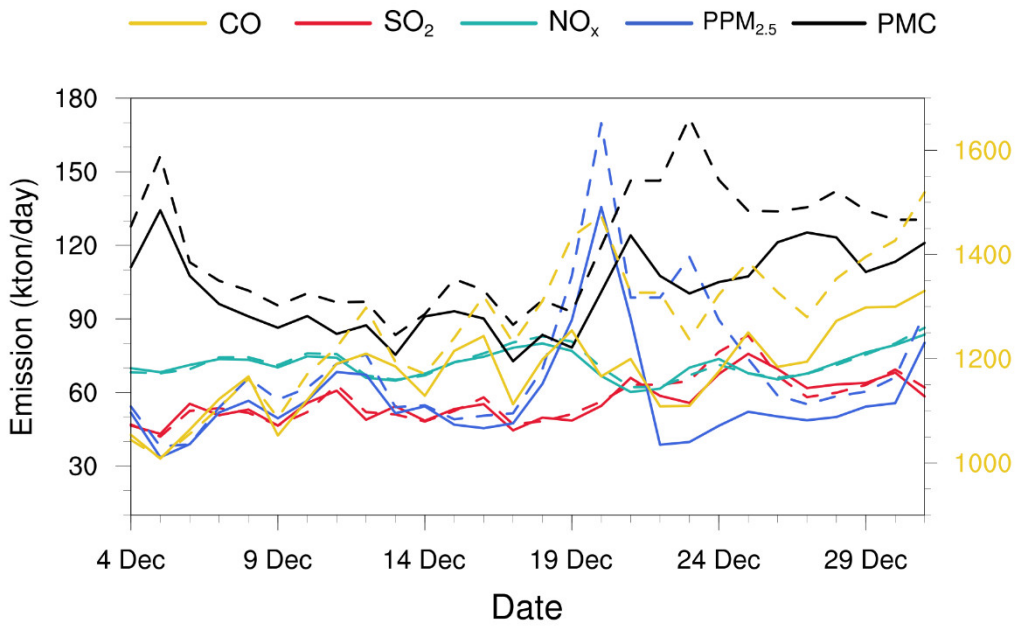

Figure R2 The temporal variability of the posterior emissions (kton/day) in EMSD (solid line) and EMS1 (dotted line) experiments over the whole mainland China. Labels on the right Y-axis refer to CO emissions. (Figure S11 in the revised manuscript)

Jiang, Z., Worden, J. R., Worden, H., Deeter, M., Jones, D. B. A., Arellano, A. F., and Henze, D. K.: A 15-year record of CO emissions constrained by MOPITT CO observations, Atmospheric Chemistry And Physics, 17, 4565-4583, 10.5194/acp-17-4565-2017, 2017.

Miyazaki, K., Eskes, H., Sudo, K., Boersma, K. F., Bowman, K., and Kanaya, Y.: Decadal changes in global surface NOx emissions from multi-constituent satellite data assimilation, Atmospheric Chemistry and Physics, 17, 807-837, 2017.

Peng, Z., Lei, L., Liu, Z., Su, J., Ding, A., Ban, J., Chen, D., Kou, X., and Chu, K.: The impact of multi-species surface chemical observation assimilation on air quality forecasts in China, Atmospheric Chemistry and Physics, 18, 10.5194/acp-18-17387-2018, 2018.

Schwartz, C. S., Liu, Z., Lin, H.-C., and Cetola, J. D.: Assimilating aerosol observations with a "hybrid" variational-ensemble data assimilation system, Journal Of Geophysical Research-Atmospheres, 119, 4043-4069, 10.1002/2013jd020937, 2014.

3.b. In addition, it is a well established validation procedure and common practice after implementation of a new data assimilation/inversion algorithm, to test the system by "identical twin" experiments and Observation System Experiments (OSEs), where a virtual reality is given by "nature runs", where "artificial observations" serve to estimate known chemical states and emission rates (see textbook by Daley, 1990 for details). There is no hint in the paper, that corresponding activities have been undertaken. These tests do not confirm the correctness of the theoretical approach. This being assumed, they seek a necessary, yet insufficient test being approved. The authors are strongly encouraged to validate and proof their approach by these tests.

Response: Thank you for this suggestion. According to this suggestion, we have added an Observing Systems Simulation Experiment (OSSE) to further validate our system in the revised manuscript. We also show the modifications as follows:

Lines 561-568, page 25.

"… coincides with the research stage. An Observing Systems Simulation Experiment (OSEE) was conducted to evaluate the performance of the RAPAS system, which has been widely used in previous assimilation systems development (Daley, 1997). In the

OSSE experiment, we used the MEIC 2016 inventory as a "true" emission, and reduced the "true" emission by 30% over the mainland of China as a prior emission. The simulations simulated using the "true" emission and sampled according to the locations and times of the real observations were used as artificial observations. The observation errors are the same as those in EMDA. To evaluate the IC improvements …"

Table 3

Exp. Type	Exp. Name	Period	IC of the first DA window	ICs of the subsequent DA window	Emission
Assimilation	OSSE	1-31 December	The same as EMDA	The same as EMDA	The same as EMDA, but with a decrease of 30% for CO, SO_2, NO_x, $PPM_{2.5}$, and PMC

Lines 819-828, pages 42-43.

4.1.6 Evaluation using OSSE

Figure 11 shows the spatial distribution of the error reduction in the posterior emissions of the five species. It can be found that after inversion, in most areas, the emission errors can be reduced by more than 80%, especially in the central and eastern regions with dense observation sites, while in remote areas far away from cities, due to the sparse observation sites, the emission errors are still not well adjusted. Overall, the error reduction rates of CO, SO_2, NO_x, PPM2.5, and PMC are 78.4%, 86.1%, 78.8%, 77.6%, and 72.0%, respectively, indicating that with the ground in-situ observations in China, RAPAS can significantly reduce emission errors, thus has good performance in emission estimates.

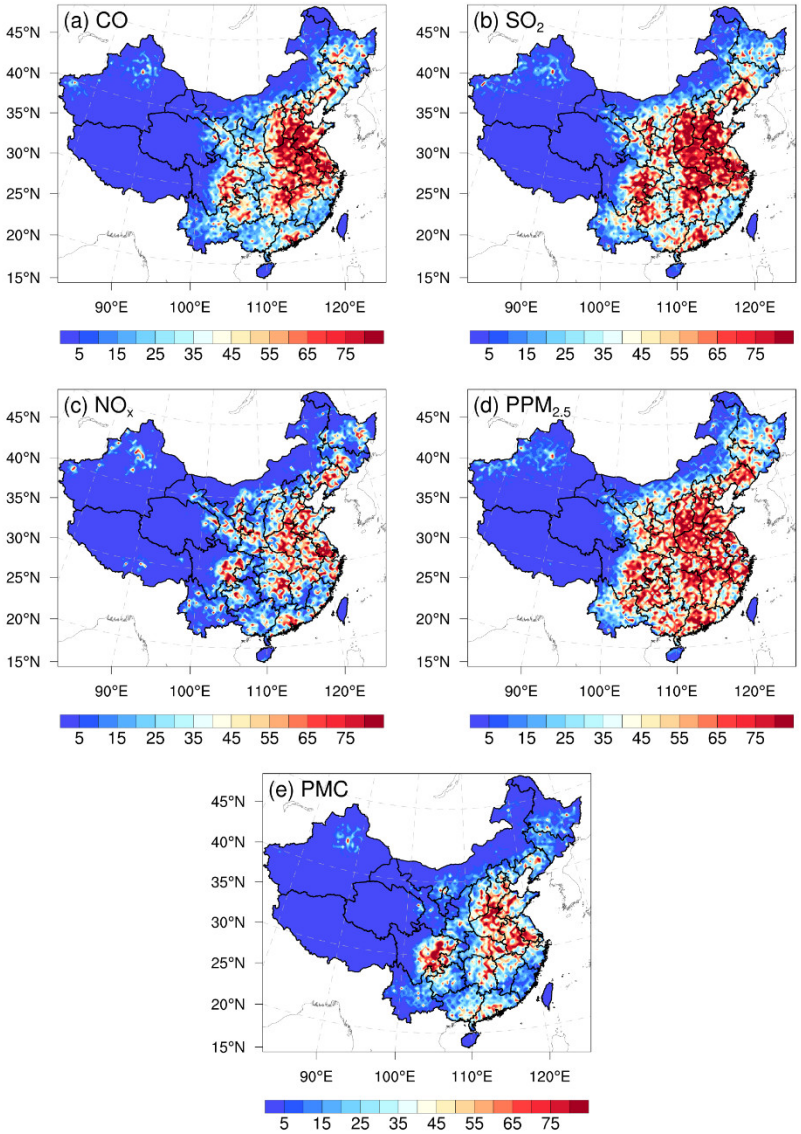

Figure R3 Spatial distribution of the error reduction (%) of posterior emissions in the OSSE. (Figure 11 in the revised manuscript)

4. The presentation of the EnKF is confusing (see equs. (7)-(12)): Below (7) we read "$\boldsymbol{\delta X_i^b}$ represents the randomly perturbed samples that are added to the prior emissions Xb0 to produce ensemble samples of the inputs $\mathbf{X_i^b}$ …..". Evidently these parameters are emissions. In equs (8) and (11), also others in between, these parameters are used as state parameters in traditional ensemble KF notation, revealing in (8) a comparison with observations ($\mathbf{y - H\overline{X^b}}$). I presume observation operator $\mathbf{H}$ does not link

emissions **X** with observations **y**. Rather concentrations are surely meant here (and not the extremely unlikely case of available eddy covariance tower flux measurements). A means to unravel this would be the approach of expanding the state vector of concentrations by emission rates following Wu at al. (2016). See their eq. (5.3). Take also note of their discussion of optimization and control on the finite assimilation window length. What is the assimilation frequency within a DA window of 1 day? Please clarify these points.

Response: Thank you for this comment. The observation operator $\boldsymbol{H}$ 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 (Peng et al., 2017). $\boldsymbol{H\overline{X^b}}$ can be regarded as the simulations at the observation locations. Therefore, $\mathbf{y} - \mathbf{H}\overline{\mathbf{X^b}}$ represents the deviations between the observations and the simulations. Wu at al. (2016) expanded the state variables to realize the joint optimization of concentrations and emissions, which is similar to "one-step" method. The relationship between emissions and observations was also expressed through the observation operator $\boldsymbol{H}$ (Wu et al., 2016, Section 3). The assimilation frequency in one assimilation window is once. Following descriptions are added in the revised manuscripts.

See lines 400-402, page 19.

"$\boldsymbol{H}$ is the observation operator that maps simulated concentrations from model space to observation space; $\boldsymbol{y - H\overline{X^b}}$ reflects the differences between the simulated and observed concentrations;"

See lines 433-434, page 20.

"Due to the "super-observation" approach, only one assimilation is needed in one assimilation window."

Peng, Z., Liu, Z., Chen, D., and Ban, J.: Improving PM_{2. 5} forecast over China by the joint adjustment of initial conditions and source emissions with an ensemble Kalman filter, Atmospheric Chemistry and Physics, 17, 4837-4855, 10.5194/acp-17-4837-2017, 2017.

5. Is there a test set demonstrating prognostic improvements? And if so, which forecast time is used. I was unable to identify such an experiment, which demonstrates a type of benefits from emission optimization.

Response: Thank you for this comment. In previous, many studied have been conducted to illustrate the benefits of emission optimization for air quality forecasting (ref.). In fact, improving air quality forecasts was one of our goals in developing this system, and during the development of this system, we indeed conducted experiments to test its ability to improve forecasts. We found that, compared the forecasts (72 hours) with original emission inventory, overall, the RMSEs of forecasts with the optimized emissions could decrease by 45.2%, 57.0%, 43.8%, 34.9% and 41.3% and the CORRs could increase by 82.0%, 140.0%, 19.9% 40.5% and 59.4% for CO, SO_2, NO_2, $PM_{2.5}$ and PM_{10}, respectively (Figure R4). Because the purpose of this paper is to introduce this system, we do not present in the paper the impact of emission optimization using this system on air quality forecasts..

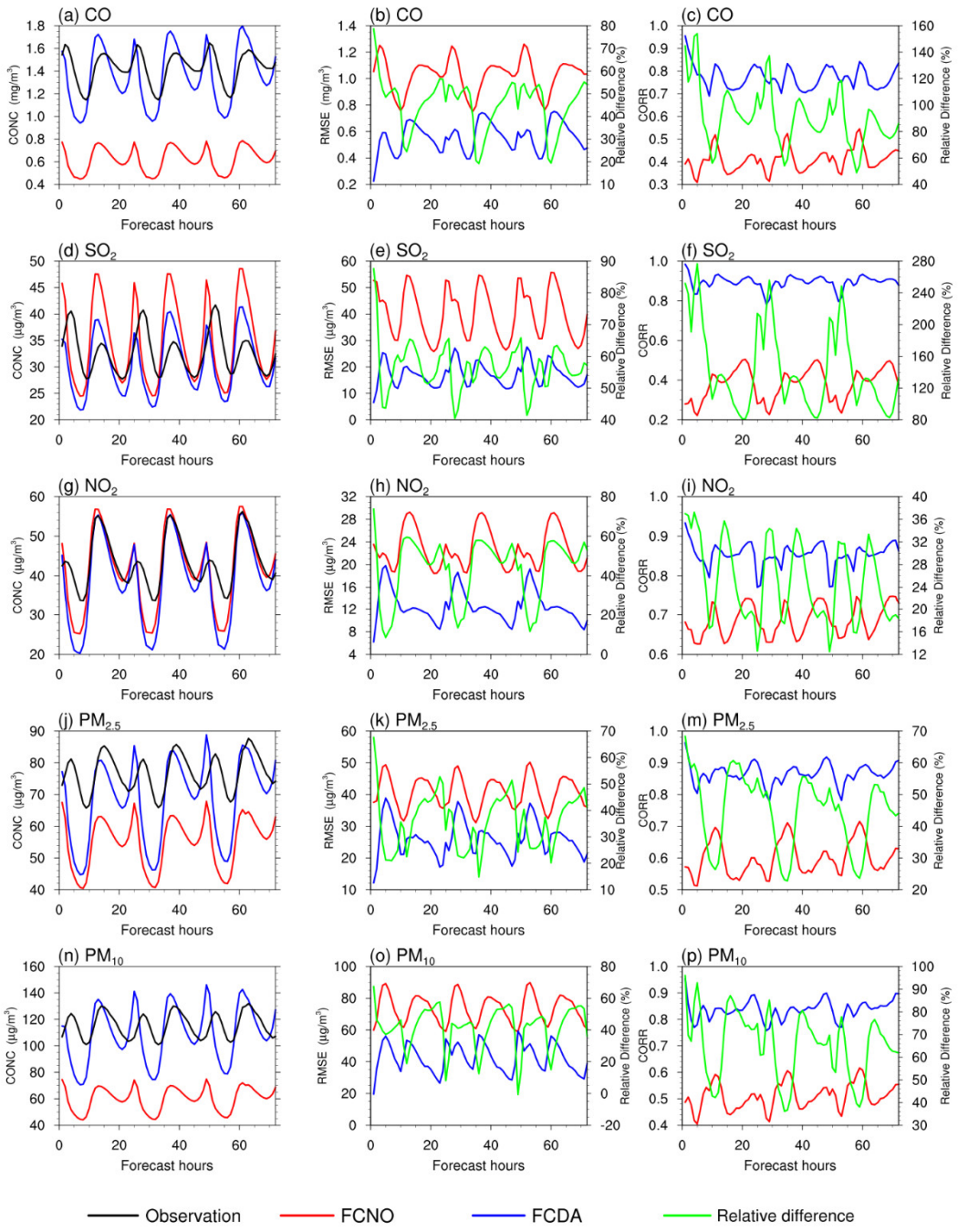

Figure R4 Mean forecasts and observations, RMSE (mg/m^3 for CO and μg/m^3 for other species) and CORR of the 72-hr forecasts, aggregated over 28 forecasts and verified against all surface monitoring data. The forecasts are performed at each 0000 UTC and the emissions of FCNO and FCDA are from the MEIC and a posteriori of the previous window, respectively.

6. L 203: The grid size of 36 km is very coarse. In fact to be judged in relation to your following statement:

L 878-883 “In additionally, 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, in which the O3 concentration and NOx (VOC) emissions are positively correlated in theNOx (VOC)-limited region and negatively correlated in the VOC (NOx)-limited region (Tang et al., 2011).” NOx chemistry is linked to VOCs and CO via O3. Due to the lack of VOC observations, the evolution of O3 (build up) is mandatory, as otherwise NO2 evolution is not properly analysable. This clearly indicates that the model set-up/coarseness is unable to simulate the core reactivity of tropos. chemistry properly and thus is unsuited to infer emission rates. It is suggested to rerun the experiment with asufficiently highly resolved nest on a densely observed area..

Response: Thanks for this suggestion. Indeed, the unoptimized VOC emissions due to the lack of observations will perturb the O3-NOx-VOC relationship, and then affect the optimization of NOx emissions. The purpose of those sentences is to show that due to the highly nonlinear relationship of NOx-O3-VOC, we cannot constrain NOx or VOC emissions by assimilating O_3 observations, not that the model cannot well simulate the tropospheric chemical reactions. CMAQ is a widely used and validated regional air quality model (Nolte et al., 2015), with the popular spatial resolution about 36 km at the regional scale (Chen et al., 2021; Holloway et al., 2012; Moniruzzaman et al., 2020; Sharma et al., 2016). Many studies used the global model (e.g., GEOS-Chem) to simulate O3 (Zhang et al., 2011; Jourdain et al., 2010; Parrington et al., 2008) or constrain NOx emissions (Jourdain et al., 2010; Vinken et al., 2014) with a resolution of tens to hundreds of kilometers at the regional scale, but they can still better capture the temporal and spatial changes of NOx and O3. There are reasons to believe that CMAQ has the ability to simulate O3 well at a higher resolution (36 km) at the regional scale. For example, Sharma et al. (2016) used CMAQ model with a resolution of 36 km to simulate ozone in India, and further investigated the effect of different precursor

species and sources on O3. The evaluations showed that the simulation was in good agreement with the observation. Using the same resolution, Mueller and Mallard (2011) used CMAQ model quantified the contribution of natural sources to ozone over the continental United States. Model evaluations showed that overall bias in daily maximum 8-h ozone was less than3 ppbv and the mean fractional bias was only 6%. In addition, we also evaluated the O3 simulation with posterior emissions inferred in this study and found that, compared with observation, the simulated O3 is still improved due to improved NOx estimation, with a decrease of RMSE from 35.8 to 33.3 μg m^{-3} and an increase of correlation coefficient from 0.43 to 0.53. Gan et al. (2016) found that there was no significant difference for both simulated pollutant concentrations and radiation variables at 36 and 12 km resolution in CMAQ model. We also further conducted a nested emission inversion on a densely observed area (the Yangtze River Delta, China) with a grid spacing of 12 km (Feng et al., 2022). The study period is the same as this study. Results showed that the biases (-2.6 vs. -1.0 μg m^{-3}) , RMSEs (11.0 vs. 9.4 μg m^{-3}) and correlation coefficients (0.86 vs. 0.89) of simulated NO2 with two posterior emissions were comparable against the observation in the Yangtze River Delta. Moreover, the NOx emissions in the Yangtze River Delta retrieved at two resolutions are almost the same (14.7 kt/day vs. 13.4 kt/day), with a difference of 8.8%. These statistics indicate that the emissions can be adjusted effectively by RAPAS. We have added following discusses about the possible effect of coarse model settings on NOx emission inversion.

Lines 243-245, pages 9-10.

"This spatial resolution has been widely adopted in regional simulations and can provide good simulations of the spatiotemporal variations of air pollutants (Mueller and Mallard, 2011; Sharma et al. 2016)."

Lines 1141-1146, page 60.

"The optimization of the initial fields or emissions of NO_2 may also change the O_3-NO_x-VOC relationship. Assuming that NO_2 is underestimated, the NO_2 concentration

increases after assimilation, but the VOC concentration remains unchanged, then in the NOx (VOC)-limited region, the subsequent generation of O_3 will increase (decrease); Conversely, the ozone concentration errors caused by assimilating NO_2 will also affect the subsequent NOx emission inversion."

Lines 1150-1157, page 60.

"To evaluate the influence of O_3-NO_x-VOC relationship change and model resolution on inversion, we also further conducted a nested emission inversion on a densely observed area (the Yangtze River Delta, China) with a grid spacing of 12 km (Feng et al., 2022). The study period is the same as this study. Results showed that the NOx emissions in the Yangtze River Delta retrieved at two resolutions are almost the same (14.7 kt/day vs. 13.4 kt/day), with a difference of 8.8%, indicating that the emissions can be adjusted effectively by RAPAS."

Chen, Y., Shen, H., Kaiser, J., Hu, Y., Capps, S. L., Zhao, S., Hakami, A., Shih, J. S., Pavur, G. K., Turner, M. D., Henze, D. K., Resler, J., Nenes, A., Napelenok, S. L., Bash, J. O., Fahey, K. M., Carmichael, G. R., Chai, T., Clarisse, L., Coheur, P. F., Van Damme, M., and Russell, A. G.: High-resolution hybrid inversion of IASI ammonia columns to constrain US ammonia emissions using the CMAQ adjoint model, Atmos. Chem. Phys., 21, 2067-2082, 2021.

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.

Gan, C.-M., Hogrefe, C., Mathur, R., Pleim, J., Xing, J., Wong, D., Gilliam, R., Pouliot, G., and Wei, C.: Assessment of the effects of horizontal grid resolution on long-term air quality trends using coupled WRF-CMAQ simulations, Atmospheric Environment, 132, 207-216, 2016.

Holloway, T., Voigt, C., Morton, J., Spak, S. N., Rutter, A. P., and Schauer, J. J.: An assessment of atmospheric mercury in the Community Multiscale Air Quality (CMAQ) model at an urban site and a rural site in the Great Lakes Region of North America, Atmos. Chem. Phys., 12, 7117-7133, 2012.

Jourdain, L., Kulawik, S. S., Worden, H. M., Pickering, K. E., Worden, J., and Thompson, A. M.: Lightning NO_x emissions over the USA constrained by TES ozone observations and the GEOS-Chem model, Atmos.

Chem. Phys., 10, 107-119, 2010.

Moniruzzaman, C. G., Bowden, J., and Arunachalam, S.: Aircraft landing and takeoff emission impacts on surface O3 and PM2.5 through aerosol direct feedback effects estimated by the coupled WRF-CMAQ model, Atmospheric Environment, 243, 117859, 2020.

Mueller, S. F. and Mallard, J. W.: Contributions of Natural Emissions to Ozone and PM2.5 as Simulated by the Community Multiscale Air Quality (CMAQ) Model, Environmental Science & Technology, 45, 4817-4823, 2011.

Nolte, C. G., Appel, K. W., Kelly, J. T., Bhave, P. V., Fahey, K. M., Collett Jr, J. L., Zhang, L., and Young, J. O.: Evaluation of the Community Multiscale Air Quality (CMAQ) model v5.0 against size-resolved measurements of inorganic particle composition across sites in North America, Geosci. Model Dev., 8, 2877-2892, 2015.

Parrington, M., Jones, D. B. A., Bowman, K. W., Horowitz, L. W., Thompson, A. M., Tarasick, D. W., and Witte, J. C.: Estimating the summertime tropospheric ozone distribution over North America through assimilation of observations from the Tropospheric Emission Spectrometer, Journal of Geophysical Research: Atmospheres, 113, 2008.

Sharma, S., Chatani, S., Mahtta, R., Goel, A., and Kumar, A.: Sensitivity analysis of ground level ozone in India using WRF-CMAQ models, Atmospheric Environment, 131, 29-40, 2016.

Vinken, G. C. M., Boersma, K. F., van Donkelaar, A., and Zhang, L.: Constraints on ship NO_x emissions in Europe using GEOS-Chem and OMI satellite NO2 observations, Atmos. Chem. Phys., 14, 1353-1369, 2014.

Zhang, L., Jacob, D. J., Downey, N. V., Wood, D. A., Blewitt, D., Carouge, C. C., van Donkelaar, A., Jones, D. B. A., Murray, L. T., and Wang, Y.: Improved estimate of the policy-relevant background ozone in the United States using the GEOS-Chem global model with 1/2° × 2/3° horizontal resolution over North America, Atmospheric Environment, 45, 6769-6776, 2011.

Technical corrections:

7. L. 59-63: What is meant in one case: Emission optimization with or without prior IC by 3D-var?

Response: Thanks. We have rephrased the sentence (See lines 61-69, page 3) as follows:

"Results showed that the simulated concentrations of CO, NO2, SO2, PM2.5 and PM10 with the prior inventory have large systematic biases, with relative biases in the range of -48.2-54.2%. In the IA subsystem, after 3DVAR, the root mean squared error (RMSE) of the simulated concentrations decreased by 50.0-73.2%, and the correlation coefficient (CORR) increased to 0.78-0.92 for the five species. In the EI subsystem, after emission inversions, the RMSE of the simulated concentrations decreased by 40.1-56.3%, and the CORR increased to 0.69-0.87."

8. L. 68-69: Selection of the Dec 2016 case study analysis: in view of typically episodic nature of mineral dust: does this high quantitative increment make sense for a validation of a novel algorithm? Why not another episode?

Response: Thanks for this comment. China has implemented strict pollution control measures in recent years to combat air pollution, especially in the cold season, and December is one of the most polluted months in China. Therefore, we believe that it is more meaningful to invert this month's emission changes for the pollution control. In addition, in China, sandstorms mainly occur in spring, and the month we choose is basically unaffected by sandstorms. However, because northern China is basically arid and semi-arid regions, the wind-blown dust exists throughout the year. According to the suggestion of another reviewer, we have also performed another inversion in July. The increase of PMC emissions (1178%) constrained by observations in July is comparable with that in December. The following discussions are added in the revised manuscript. See lines 549-551, page 24.

"December is one of the months with most severe air pollution, while July is one of the least polluted months in China. Therefore, this study mainly tested the performance of the RAPAS system in these two months. For December, the IA …"

See lines 556-559, pages 24-25.

"For July, the system also operated in the same way as for December. It needs to be noted that due to the stronger atmospheric oxidation, the lifetime of NO2 in July is

significantly shorter than that in December, thus we adopted a smaller localization scale for NO2 (80 km)."

See lines 868-873, page 45.

"The evaluation results for July show that the emission uncertainty can still be significantly reduced, and the performance of the system in July is comparable to that in December (Table S2). Additionally, the seasonal variation of emissions can be well reflected (Figures S4 and S5), which means that our system can perform well at different times of the year."

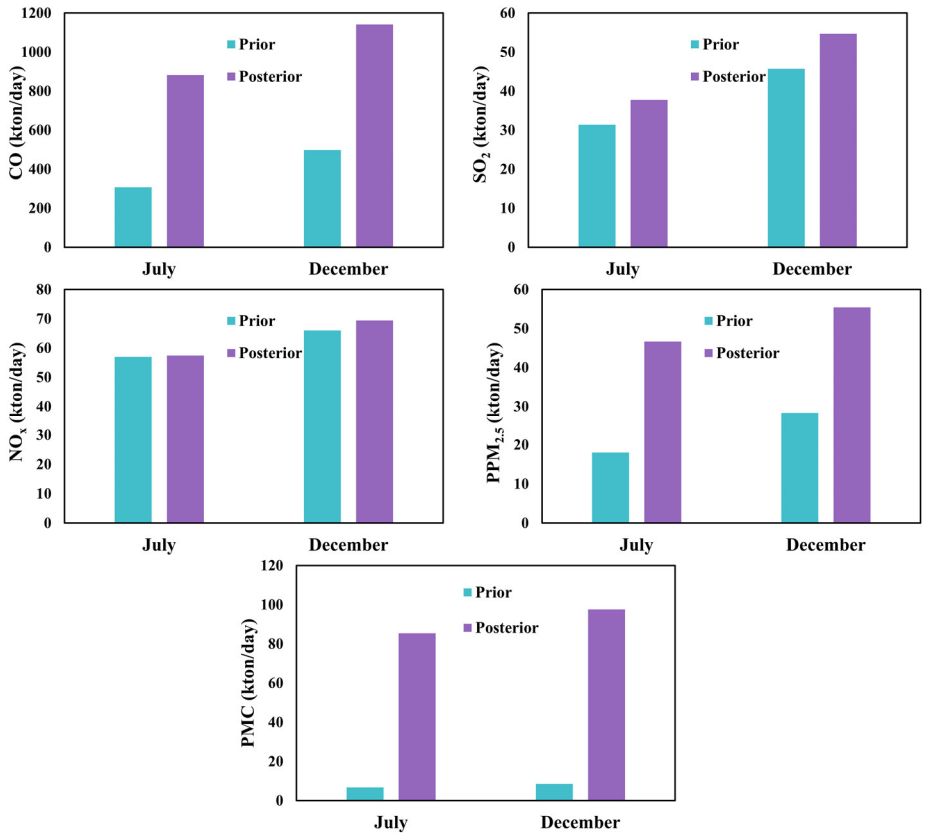

Figure R5 The prior and posterior emissions (kton/day) in July and December 2016 over the whole mainland China. (Figure S5 in the revised manuscript)

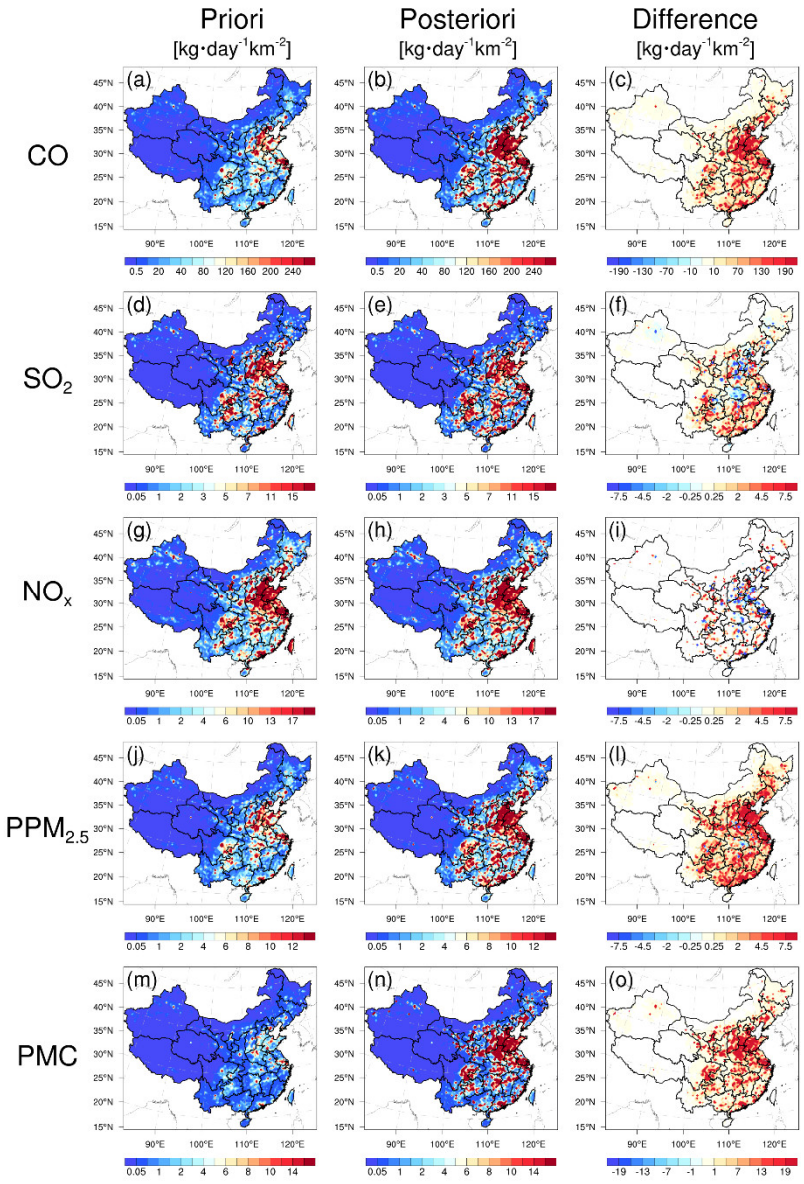

Figure R6 Spatial distribution of the time-averaged prior emissions (left column, MEIC 2016), posterior emissions (middle column), and differences (right column, posterior minus prior) in July 2016. (Figure S4 in the revised manuscript)

Table R1. Statistics comparing the pollution concentrations from the simulations with prior (CEP) and posterior (VEP) emissions against observations for July. CO unit: mg m^{-3}; others units: μg m^{-3}. (Table S2 in the revised manuscript)

Species	Mean Obs.	Mean Sim.		BIAS		RMSE		CORR	
		CEP8	VEP8	CEP8	VEP8	CEP8	VEP8	CEP8	VEP8
CO	0.79	0.33	0.63	-0.46	-0.16	0.58	0.35	0.25	0.65
SO_2	12.9	15.6	9.6	2.7	-3.3	19.7	7.2	0.12	0.71
NO_2	20.0	23.0	16.5	3.0	-3.5	22.0	6.8	0.47	0.81
$PM_{2.5}$	29.2	21.9	23.3	-7.3	-6.0	21.1	13.9	0.51	0.76
PMC	53.6	27.6	42.3	-26.0	-11.3	42.3	30.9	0.38	0.61

* BIAS, mean bias; RMSE, root mean square error; CORR, correlation coefficient

Beirle, S., Platt, U., Wenig, M., and Wagner, T.: Weekly cycle of NO2 by GOME measurements: a signature of anthropogenic sources, Atmospheric Chemistry and Physics, 3, 2225-2232, 2003.

9. L 122-124: Unclear: “Barbu et al. (2009) updated sulfur oxide (SOx) emissions with SO2 gas and sulfate aerosol observations and showed that forecasts were improved overall but degraded when derived only from SO2 or sulfate observations.” Better formulate logics: ... do you mean alone?

Thanks for this comment. We have rephrased that sentence to make it clear. See lines 134-137, page 4.

“Barbu et al. (2009) updated sulfur oxide (SO_x) emissions with SO2 and sulfate aerosol observations and found that simultaneous assimilation of both species had better performance than assimilating one of them alone.”

10. L 176-177: “It runs only once and provides a “perfect” chemical ICs for the subsequent EI subsystem.” Justify "perfect". Justify analysis of unobserved height

levels?

Response: Thanks for this comment. Even with data assimilation, it is still impossible to get a perfect initial field, so we put the quotes around the word of perfect. In the revised manuscript, we have changed "perfect" to "good" in that sentence (See line 42, page 2; line 530, page 23). In addition, for the concentrations at high levels, they could be constrained by the ground observations through the vertical correlation. Figure R7 shows the changes of vertical profiles between the assimilation and control experiments. Overall, it could be found that there are changes of concentrations almost throughout the whole column. However, due to lack of observations of vertical profiles, it is difficult for us to verify the constraints on the concentrations of the upper air. Anyway, the effect of only assimilating ground observations on the concentration in middle and upper layers of the boundary layer is necessarily limited, so there must be a bias. This also needs our attention, and may be partially resolved by further assimilating satellite observations in the future. We have added some discussions about this point.

See lines 666-670, page 30.

"These statistics indicate the initial fields of the ground level have been significantly improved. However, due to the lack of observations, we still do not know the simulation bias in the upper-middle boundary layer. Although concentrations at high altitudes can be constrained by ground-based observations through vertical correlations, the effect is limited, so the bias is still non-negligible."

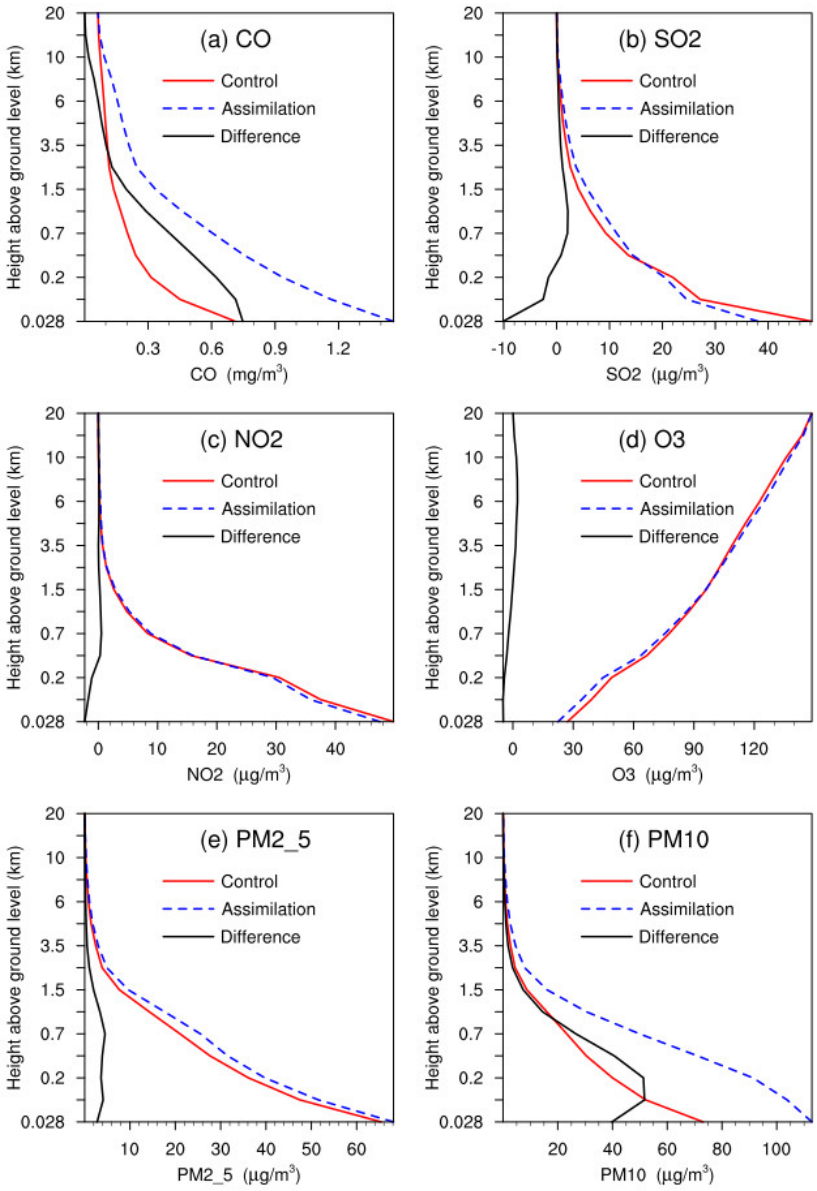

Figure R7 Mean vertical profiles of the background (red) and analysis (blue) fields, and their difference averaged over the China mainland.

11. L. 180-181: "which are then sampled according to the locations and times of the observations" Unclear, more technical details are needed here.

Response: Thanks for this comment. We have changed it to the following sentence. See lines 212-215, page 8.

"In the first step, the prior emissions ($\boldsymbol{X^b}$) are perturbed and put into the CTM model to

simulate chemical concentration ensembles. The simulated concentrations of the lowest model level are then interpolated to the observation space according to the locations and times of the observations using the nearest neighbor interpolation method."

12. L. 184: Do you mean DA window length of 1 day?

Response: Yes, the DA window is set to 1 day. On the one hand, the model needs a longer time to integrate emission information into the concentration ensembles. On the other hand, due to the complexity of hourly emissions, it is very difficult to simulate hourly concentrations that can match the observations well. We have added descriptions about the length of DA window to Figure 1.

13. Figur1 a caption too short. More detail on precise times and data update frequency in the EnKF needed.

Response: Many thanks for this suggestion. The caption of Figure 1 has changed as follows:

"**Figure 1**. The composition and flow chart of RAPAS. The $\mathbf{x_a}$ and $\mathbf{x_b}$ represent the prior and posterior emissions. The 3DVAR assimilation stage lasts 5 days with data input frequency of 6 hours, and the DA window in the EI subsystem is set to 1 day."

14. L 225 "represent the measurement sites" Do you mean "chemistry"?

Response: Yes. We have changed "the measurement sites" to "the air pollution monitoring sites". See line 268, page 12.

15. L 242: "… of surface air pollutant observation operators,…" Explain please.

Response: Thanks for this comment. The surface air pollutant observation operator means the interpolation of simulated concentrations from model space to observation

space according to the locations of observation data. In this study, we need to construct observation operators for different pollutants. It has been explained in lines 297-319, page 14.

16. L 312 “the horizontal length scales decrease with increasing heights,” Please explain why, and not the opposite.

Response: Thanks. We made a mistake here. As shown in Figure 3, for gaseous pollutants and most individual aerosol components, the horizontal length scales increase with height. We have corrected it in the revised manuscript. See lines 354-355, page 17.

17. L 314 “ The ground-level scale generally spread 40-45 km for all control variables on average.” This is little more than a grid cell. How defined?

Response: Thanks for this comment. Although the ground-level scale calculated in Formular 6 generally spread 40-45 km on average, in GSI namelist, two variables, namely scale factor and weight factor, were used for further tuning horizontal scale (e.g., 40-45 km) specified for horizontal smoothing and weight setting. The final background error covariance matrix used in analysis are the content multiplied by these factors. Therefore, the actual increment usually spreads multiple grids.

17. L 315-318: More detailed explanation required: Do you mean stack overshooting vs. surface emissions? A result of NMC? Is boundary height and related mixing considered?

Response: Yes, this result was calculated using the “NMC” method. The boundary height and related mixing were considered because these processes are included in the CMAQ model. The vertical length scale of most species increases first and then decreases with height, which is related to the vertical mixing and stack emissions at

about 200 m height. Ma et al (2018) also found high correlation occurred in the middle boundary layer other than the upper or lower levels.

The sentence (See lines 358-360, page 17) has been rephrased as follows:

"The vertical length scale of most species increases first and then decreases with height, which may be related to the vertical mixing (Kahnert, 2008) and stack emissions at about 200 m height."

Ma, C., Wang, T., Zang, Z., and Li, Z.: Comparisons of Three-Dimensional Variational Data Assimilation and Model Output Statistics in Improving Atmospheric Chemistry Forecasts, Advances in Atmospheric Sciences, 35, 813-825, 2018.

18. L 323-324: Whitaker and Hamil is a meteorological application without emission optimisation. For square root filtering technique fundamentals please refer to much earlier textbook literature, eg. Bierman: 1977, or Maybeck 1979

Response: Thanks! We have updated the references. See lines 368-369, page 17.

19. L 336 variable localization: meaning?

Response: The variable localization was defined in Miyazaki et al. (2012) and Ma et al. (2019). It means that the emission of one species is only constrained with its corresponding air pollutant observation. We have rephrased that sentence (See lines 383-388, page 18) as follows:

"We used variable localization to update the analysis, which means that the covariance among different state variables was not considered, and the emission of one species was only constrained with its corresponding air pollutant observation. This method has been widely used in chemical data assimilation systems to avoid spurious correlations among species. (Ma et al., 2019; Miyazaki et al., 2012b)."

Ma, C., Wang, T., Mizzi, A. P., Anderson, J. L., Zhuang, B., Xie, M., and Wu, R.:

Multiconstituent Data Assimilation With WRF-Chem/DART: Potential for Adjusting Anthropogenic Emissions and Improving Air Quality Forecasts Over Eastern China, 124, 7393-7412, 2019.

Miyazaki, K., Eskes, H. J., Sudo, K., Takigawa, M., van Weele, M., and Boersma, K. F.: Simultaneous assimilation of satellite NO2, O-3, CO, and HNO3 data for the analysis of tropospheric chemical composition and emissions, Atmospheric Chemistry and Physics, 12, 9545-9579, 2012.

20. L 336-337: Do you mean also at the same location?

Response: We reinterpreted that sentence. Please see Response 19.

21. Eq (12) Matrix in a denominator? Please explain

Response: Yes, the denominator is a matrix. We have revised Eq (12) to make it clear.

$$\widetilde{\boldsymbol{K}} = \left(1 + \sqrt{\boldsymbol{R}/(\boldsymbol{H}\boldsymbol{P}^{\boldsymbol{b}}\boldsymbol{H}^{T} + \boldsymbol{R})}\right)^{-1} \mathbf{K} \qquad (12)$$

22. L. 384-386: This is too much a try-and-error like discussion. Emissions are not constant or diurnially constant over these times: weather changes, working days, holidays, ... Please be more precise about your reasoning

Response: Thanks for your comment. That sentence (see lines 438-442, page 20) has been changed as follows:

"Kang et al. (2012) conducted OSSEs and demonstrated that due to the errors of transport and increase the spurious correlation, a longer DA window (e.g., 3 weeks) would cause the analysis system to blur out the essential emission information far away from the observation."

23. L425 arithmetically

Response: Thank you. We have corrected it. See lines 488, page 22.

"… and the ensemble-estimated error covariance matrix is arithmetically likely to be underestimated."

24. L 446 are distributed

Response: We have changed "distribute" to "are distributed". See line 510, page 22.

25. L 449 unrepresentative

Response: We have changed "nonspatially representative" to "unrepresentative". See line 513, page 21.

26. L 452-453 Explain relations, formulae better

Response: We have rewritten that formulae and added some explains in the revised manuscript. See lines 516-519, page 23.

"… a function of $max(|O(t) - O(t \pm 1)|) \leq f(t)$, where $O(t)$ and $O(t \pm 1)$ represent observations at time t and $t \pm 1$, respectively, and $f(t) = T_a + T_b \times O_t$. That means that both concentration differences between time *t* and time *t+1* and *t-1* should be less than $f(t)$."

27. L461 25/336 is a fairly low fraction. Please justify.

Response: For better constraints on emissions, we need to select relatively more observations for assimilation. The 25 independent observations are relatively evenly distributed in the central and eastern regions where emissions are relatively strong. We believe that it is enough for the evaluation of the posterior emissions.

28. L 534- 540 For emission inversion the boundary layer height and its prevalent stability is more decisive than any other parameter. Can you please make efforts to validate this by radiosonde data, if available?

Response: Thank you for this suggestion. We have obtained 92 sites (http://weather.uwyo.edu/upperair/sounding.html, last access: 10 March 2022) of sounding data (including virtual potential temperature, geopotential height, wind speed, wind direction, etc) from the University of Wyoming (http://weather.uwyo.edu/upperair/sounding.html) for the validations of the simulated planetary boundary layer height (PBLH), which are in 12 hours interval (00 and 12 UTC). The observed PBLH was calculated with these sounding data using the bulk Richardson number method (Grachev et al., 2013; Richardson et al., 2013). Overall, the WRF simulations show an overall systematic underestimation of PBLH, with bias of -41.1 m (-15.4%). Similar underestimation was also found in Banks et al. (2015) when evaluated eight different schemes in WRF. Lower PBLH corresponds to worse diffusion and transport of air pollutants, which results in higher concentrations in the atmosphere, and accordingly, less emissions were inferred to compensate this underestimation. It should be noted that the low vertical resolution of most sounding data in China may cause a certain bias in the calculated PBLH. We have added evaluations about PBLH in the revised paper.

lines 622-639, page 28.

"To quantitatively evaluate the performance of the WRF simulations, the mean bias (BIAS), root mean square error (RMSE), and correlation coefficient (CORR) were calculated against the surface meteorological observations measured at 400 stations and the planetary boundary layer height (PBLH) calculated using the sounding data at 92 sites. The surface observations were obtained from the National Climate Data Center (NCDC) integrated surface database (http://www.ncdc.noaa.gov/oa/ncdc.html, last access: 25 October 2021), and the sounding data were obtained from the website of the University of Wyoming (http://weather.uwyo.edu/upperair/sounding.html, last access: 10 March 2022). The sounding data are in 12 hours interval. The observed PBLH were

calculated using the sound data through the bulk Richardson number method (Richardson et al., 2013). The spatial distribution of the meteorological stations is shown in Figure 2. The simulated temperature at 2 m (T2), relative humidity at 2 m (RH2), wind speed at 10 m (WS10), and PBLH from 26 November to 31 December 2016 are evaluated against the observations. Table 4 summarizes the statistical results of the evaluations of the simulated meteorological parameters. Overall, the T2, RH2 and PBLH are slightly underestimated, with biases of -0.1 ℃, -3.8% and -41.1 m, respectively. The CORRs are approximately 0.98 for T2, 0.94 for RH2 and 0.90 for PBLH, showing good consistency…"

Banks, R. F., Tiana-Alsina, J., Rocadenbosch, F., and Baldasano, J. M.: Performance Evaluation of the Boundary-Layer Height from Lidar and the Weather Research and Forecasting Model at an Urban Coastal Site in the North-East Iberian Peninsula, Boundary-Layer Meteorology, 157, 265-292, 2015.

Grachev, A. A., Andreas, E. L., Fairall, C. W., Guest, P. S., and Persson, P. O. G.: The Critical Richardson Number and Limits of Applicability of Local Similarity Theory in the Stable Boundary Layer, Boundary-Layer Meteorology, 147, 51-82, 2013.

Richardson, H., Basu, S., and Holtslag, A. A. M.: Improving Stable Boundary-Layer Height Estimation Using a Stability-Dependent Critical Bulk Richardson Number, Boundary-Layer Meteorology, 148, 93-109, 2013.

29. "These statistics indicate that the initial fields can be adjusted effectively by our IA subsystem." I think this claim is not sustained, despite the fact that emissions are not yet corrected: What about height levels above? How are they validated? How can be made sure that CB5 chemsitry fields are in the proper chemical balance of the system (notably NO_x, O3, CO, VOCs)?

Response: Yes, this conclusion is indeed not rigorous enough. The evaluations only show that the initial fields of the ground level have been significantly improved. Due to lack of observations, we do not evaluate the concentrations at high levels. Although the concentrations at high levels could be constrained by the ground observations

through the vertical correlation, the effect is limited, so there must be a bias. We modified that sentence in the revised paper, see lines 665-670, page 30. The changes are also listed as follows.

"These statistics indicate the initial fields of the ground level have been significantly improved. However, due to the lack of observations, we still do not know the simulation bias in the upper-middle boundary layer. Although concentrations at high altitudes can be constrained by ground-based observations through vertical correlations, the effect is limited, so the bias is still non-negligible."

30. "L 619: This applies more to the NOx-O3-CO/VOC chemistry rather than PM2.5.

Response: Thank you for this comment. Yes, O_3 has complex precursors, and strong nonlinear relationship with its precursors. Although there are nonlinear processes during the formation of secondary particulate matter, overall, the more PM2.5 precursors are emitted, the more secondary particulate matters are formed, resulting in higher $PM_{2.5}$ concentrations. We have revised that sentences as follows:.

"… It may also be related to the complex precursors and complex homogeneous and heterogeneous chemical reactions and transformation processes of secondary $PM_{2.5}$, and the fact that …"

31. L 662: Where did you get the \sigma from? For KF an analysis error covariance matrix A should be available for this.

Response: Thank you for this comment. In the EnKF-based inverse estimation scheme, the uncertainty is represented by the spread of the ensemble samples (Tang et al., 2013). The posterior and prior uncertainties are the standard deviations of the prior and posterior perturbations of $\boldsymbol{X_i^b}$ and $\boldsymbol{X_i^a}$. $\boldsymbol{X_i^b}$ was perturbed from the prior emissions $\boldsymbol{X_0^b}$ by adding a randomly perturbed item of $\boldsymbol{\delta X_i^b}$, which was drawn from Gaussian distributions with a mean of zero and the standard deviation of the prior emission

uncertainty in each grid. After constrained using observations, the perturbed emissions of $\boldsymbol{X_i^b}$ is changed to $\boldsymbol{X_i^a}$ according to Eq. 2 ~ 5.

$$\boldsymbol{X_i^b} = \boldsymbol{X_0^b} + \boldsymbol{\delta X_i^b} \quad , i = 1, 2, \dots, N \tag{1}$$

$$\boldsymbol{X_i^a} = \boldsymbol{X^a} + (\boldsymbol{X_i^b} - \boldsymbol{X^b}) - \boldsymbol{\widetilde{K}H}(\boldsymbol{X_i^b} - \boldsymbol{X^b}) \tag{2}$$

$$\boldsymbol{\widetilde{K}} = \left(1 + \sqrt{\boldsymbol{R}/(\boldsymbol{HP^bH^T} + \boldsymbol{R})}\right)^{-1} \mathbf{K} \tag{3}$$

$$\mathbf{K} = \boldsymbol{P^bH^T}(\boldsymbol{HP^bH^T} + \boldsymbol{R})^{-1} \tag{4}$$

$$\boldsymbol{P^b} = \frac{1}{N-1}\sum_{i=1}^{N}(\boldsymbol{X_i^b} - \boldsymbol{\overline{X}^b})\,(\boldsymbol{X_i^b} - \boldsymbol{\overline{X}^b})^T \tag{5}$$

We have added this calculation method of the posterior and prior uncertainties in the supplemental material.

"… where $\sigma_{posterior}$ and σ_{prior} are the posterior and prior uncertainties, respectively, which were calculated using the standard deviations of the prior and posterior perturbations (Text S3)."

Tang, X., Zhu, J., Wang, Z. F., Wang, M., Gbaguidi, A., Li, J., Shao, M., Tang, G. Q., and Ji, D. S.: Inversion of CO emissions over Beijing and its surrounding areas with ensemble Kalman filter, Atmospheric Environment, 81, 676-686, 2013.

32. L 701-702: With 1000% excess is linearisation still acceptable? A scrutinized analysis is appropriate.

Response: Thank you for this comment. If we assume that the 1000% increment of PMC is all from natural dust, that means the contribution of natural dust accounts for 90% of total PMC emissions, which is reasonable in China. Qiu et al. (2016) conducted source apportionment of PM10 and PM2.5 in different functional areas of Lanzhou, China, and the results showed that the contribution of wind-blown dust accounts for more than 90% of total PMC emissions. Another study in Changsha in central China showed that the contribution of wind-blown dust account for about 75% of the primary

PM10 concentrations (Li et al., 2010).

We also realize that this problem in the inversion results. Therefore, we used a larger prior uncertainty in the first three windows, a "two-step" inversion scheme and cyclic iteration to quickly converge emissions. Figure 10 shows the time series of the relative differences between the prior and posterior emissions in each window. In addition to the relatively large adjustment range of emissions in the first three windows, the adjustment range of the PMC is basically within the uncertainty range (e.g., 40%), indicating that with this scheme, EnKF still has a good performance in emission inversion. For PMC, because the posterior emission includes the contribution of natural dust, which is not considered in the original emission, there is a large increment in the first three windows, which is consistent with Ma et al (2019).

We have added related discussions in the revised paper, see lines 861-865, page 44. The changes are also listed as follows:

"Overall, PM10 emissions (PPM2.5+PMC) increased by 318%. If we assume that all the increment in PM_{10} emissions is all from natural dust, that means the contribution of natural dust accounts for 75% of total PM_{10} emissions, which is consistent with the source apportionment of PM_{10} of 75% in Changsha in Central China (Li et al., 2010)."

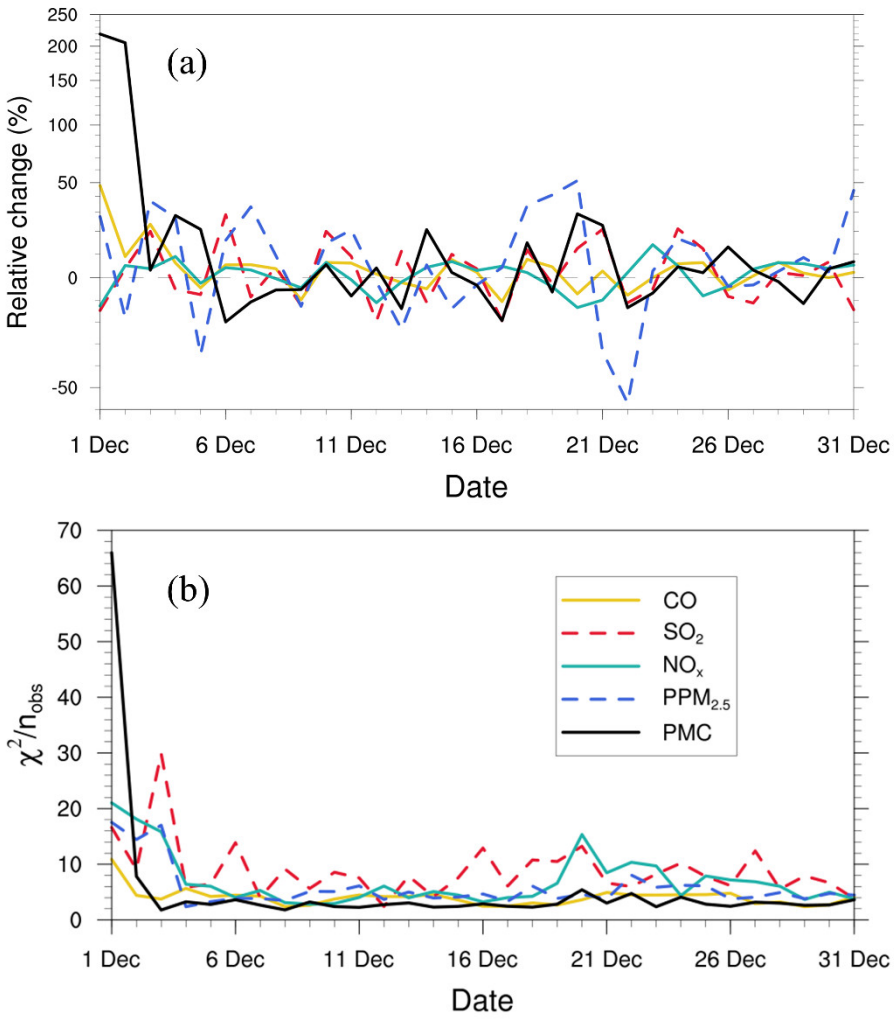

Figure R8 Relative changes (a) in a posteriori emission estimates of CO, SO_2, NO_x, $PPM_{2.5}$ and PMC, and χ^2 statistics (b) of these state vectors in each window. (Figure 10 in the revised manuscript)

Li, J.-d., Deng, Q.-h., Lu, C., and Huang, B.-l.: Chemical compositions and source apportionment of atmospheric PM10 in suburban area of Changsha, China, Journal of Central South University of Technology, 17, 509-515, 2010.

Ma, C., Wang, T., Mizzi, A. P., Anderson, J. L., Zhuang, B., Xie, M., and Wu, R.: Multiconstituent Data Assimilation With WRF-Chem/DART: Potential for Adjusting Anthropogenic Emissions and Improving Air Quality Forecasts Over Eastern China, 124, 7393-7412, 10.1029/2019jd030421, 2019.

Qiu, X., Duan, L., Gao, J., Wang, S., Chai, F., Hu, J., Zhang, J., and Yun, Y.: Chemical composition and source apportionment of PM10 and PM2.5 in different functional areas of Lanzhou, China, Journal of Environmental Sciences, 40, 75-83, 2016

33. L 705 “In addition, without dust may be another reason, since no wind blowing dust scheme was applied in this study as mentioned above.” Check English meaning.

Response: Thanks! We have revised that sentence. . See lines 858-860, page 44.

“In addition, the absence of natural dust is another reason, as the wind-blown dust scheme was not applied in this study.”

34. L727-728 : Please demonstrate this statement.

Response: Thanks for this suggestion. In the second step of the “two-step” scheme, the optimized emissions are entered again into the CTM model to generate the initial fields of the next DA window. In other words, the unresolved posterior emission error will be fed back to the initial field of the next window through simulation in the current assimilation window of one day, and then fed back to the emission again in the next window for timely optimization. Previous studies (Byun, 2002; de Almeida Albuquerque et al., 2018) have demonstrated that a spin-up time of 1-2 days can provide realistic initial conditions for the CMAQ runs, even if the simulation starts from the clean atmospheric conditions, which shows that the emissions can be well mixed in the atmosphere within the window time of one day in this study. Additionally, with this method, the system can maintain mass conservation (Zhang et al., 2015), thus it can update emissions more consistently and stably.

The sentence (See lines 1048-1053, page 456) has been rephrased as follows:

“As mentioned previously, in the “two-step” scheme, the unresolved posterior emission error will be fed back to the initial field of the next window through sufficient mixed simulation within one day for timely optimization. Meanwhile, the system always maintains the mass balance of pollutants. In this way, the system updates emissions more consistently and stably.”

Byun, D. W.: A study of photochemical processes of the Houston-Galveston

metropolitan airshed with EPA CMAQ.

de Almeida Albuquerque, T. T., Andrade, M. d. F., Ynoue, R. Y., Moreira, D. M., Andreao, W. L., dos Santos, F. S., and Sperandio Nascimento, E. G.: WRF-SMOKE-CMAQ modeling system for air quality evaluation in SAo Paulo megacity with a 2008 experimental campaign data, Environmental Science and Pollution Research, 25, 36555-36569, 2018.

Zhang, S., Zheng, X., Chen, J. M., Chen, Z., Dan, B., Yi, X., Wang, L., and Wu, G.: A global carbon assimilation system using a modified ensemble Kalman filter, Geosci. Model Dev., 8, 805-816, 2015.

35. L 746-747: This is well known, but demonstrate that this flaw cannot happen. Please my mathematical means.

Response: Thanks for this comment. During the heavy pollution period, the severe pollution enhances the feedback between the boundary layer and air pollution (Huang et al., 2020). However, WRF-CMAQ is an off-line model, which can not consider the concentration rise caused by the feedback process. Therefore, during the inversion, more emissions will be inferred to compensate this underestimation. This process is a flaw of the off-line atmospheric chemical transport model. In the “two-step” inversion, when such serious feedback leads to overestimation of emissions, the unresolved posterior emission error will be fed back to the initial field. In the next window, the system will quickly adjust the overestimated emissions to compensate for overestimated simulations. The system corrects the emissions by means of error transfer, and thus avoid the impact of error accumulation. The purpose of the “two-step” inversion strategy is not to avoid this flaw, but to make the unsolved emission inversion error be corrected quickly and stably.

Huang, X., Ding, A., Wang, Z., Ding, K., Gao, J., Chai, F., and Fu, C.: Amplified transboundary transport of haze by aerosol-boundary layer interaction in China, Nature Geoscience, 13, 428-434, 2020.

36. Fig 13 emission changes

Response: Thanks for this suggestion. We have changed “emissions changes” to “emission changes”. See line 985, page 53.