General comments

The authors present a data assimilation system for a regional air quality model run over China. The data assimilation composes two algorithms, the 3D-Var and the EnKF. Both are combined in this system to run sequentially to first optimize initial conditions of the chemical fields and then to optimize pollutant emissions. The authors demonstrate improvements in air quality model forecast after running this assimilation system.

Developing such a system is an ambitious project and the authors have clearly undertaken a significant amount of work to complete this. Data assimilation (DA) and emission inversion are very challenging in the context of air quality. In addition, such efforts are very important if the ambition associated with expansions in observation capacity on the ground and in space are to be fully realized. I therefore commend the authors for this work for its relevance and due to the technical challenges faced. That said, there are number of significant problems in the current version of the manuscript that need to be addressed. These problems are not insurmountable, though, and I therefore recommend that the paper undergo major revisions prior to consideration for publication in GMD. I outline my concerns below before going into more detail in the specific comments section.

1. A major problem arises due to the choice of the Ensemble Kalman Filter (EnKF) to perform the emission inversion. A key assumption of the EnKF, and indeed of any sequential DA method, is that the model minus observation errors have mean zero bias, be random, and have a Gaussian distribution. The authors' own results show here that this assumption is invalidated. They report an a posteriori emission scalings of between 5-1045% for each of the pollutants studied – this strongly implies the presence of large biases between the observations and model when run with a priori emissions. Such large biases will affect the optimality of the EnKF solution, and it is essential that the authors consider this in a revised manuscript. For solutions the authors should look to other fields. The problems of bias with the EnKF have been well discussed in its application in the fields of land surface (Brandhorst et al., 2017; De Lannoy et al., 2007) and in oceanographic DA (e.g., Keppenne et al., 2005). Air quality/atmospheric chemistry cannot be exempt from this. Extensive published work exists trying to solve this problem in other fields by exploring various possibilities: bias estimation as an additional state term, removal of observation bias through statistical methods, and parameter estimation. The parameter estimation presented in Brandhorst et al. (2017) gives a clear pathway for the authors to resolve this problem. If emission inversion could be framed as a bias correction method via parameter estimation (citing the correct literature), rather than being an end and means unto itself alone, then I think this would be acceptable. Right now, emission estimation is present in the first sentence of the abstract and appears to be the primary focus, but due to the limitations of the EnKF this emphasis combined with the current approach is problematic.
2. Furthermore, the authors need to make use of the available diagnostics exist to help diagnose the optimality of the EnKF analysis, e.g., the chi-squared diagnostic. Based on reported examples in the literature, my guess is that with each sequential iteration of the EnKF the emission biases are slowly reduced (prior to the reported overestimation) and the EnKF will slowly reach a more optimal state. The authors should plot chi-squared along with the relative change in a posteriori emission estimates for each successive emission inversion cycle with the EnKF. Doing this should highlight whether the optimality improves over time. If it does not then it would imply other fundamental problems with the approach.
3. The combination of the 3D-Var and EnKF together could be more clearly motivated and the choice of combining these two algorithms seems somewhat expedient. There are various self-citations justifying this choice, but the authors should dedicate some text for justifying why the two algorithms combined represent something greater than the sum of their parts. I am not saying a justification cannot be made, but we are missing a fundamental justification of the method choice.
4. As an example of potential problems with their combination, it seemed odd to first do the 3D-Var IA step and then the EnKF EI step afterwards for a mixture of long- and short-lived gases. In fact, if you optimize the concentration of CO, which is long-lived relative to the DA windows and DA cycles, then this should remove some of the bias associated with low a priori emissions of CO. The authors say this removes some of the bias associated with errors from other issues other than emissions, but this raises other issues – see next comment (5). I think this would be fine if optimizing the concentrations for

forecasting purposes is the only goal, but if the emission estimation is a primary or secondary goal (as presented) then I think that this goal is undermined by the approach taken. At the very minimum the authors should include the changes in the emission estimates obtained from and experiment with the EI without first running the 3D-var. If the goal was emission estimation, then it would seem to make more sense to perform a long spin-up simulation prior to running the EI step rather than to partially fix the model concentration field prior to this. The authors cite concerns about other types of model bias as a justification for this step, but I do not find this fully convincing. Surely, if the emissions are biased to the extent shown, this would impact the a priori ICs as much as the model run after the IA step. The large biases discussed in sect. 4.1.2 between the ICDA and ICNO experiments seem to be consistent with this point. Indeed, the authors even find that the spatial patterns in the simulated concentrations using posteriori and a priori emissions are similar to the increments calculated in the 3D-var, which indicates a potentially similar cause in both cases. I think the authors need to make at least another experiment to test the effect of this in the EI step (see specific comment 11).

5. The authors appear to claim that the IA step removes some biases and prevents biased emission estimates. My concern here though is that I see no way for the 3D-Var to distinguish model-observation biases arising due to emissions or biases due to some other source. If this is somehow possible then the authors should explain how. In either case I think the authors should be clearer, either about the limitations of partitioning different types of bias, or on how this part of the algorithm functions. Certainly the similarity in spatial patterns of the 3D-var IA increments and a priori vs a posteriori EI hints that both systems are addressing the same root cause of the biases but with different assumptions about the cause.
6. As a further comment on model error, bias, and emission estimation. I think the authors should be clearer that despite the IA step, from which the model is assumed perfect, inherent errors arising from discretization and model parameterizations will still exist meaning that the model will have other non-resolvable biases. This will naturally mean that other types of model bias will feed into the emission inversion step leading to biased emission estimates. Perhaps the authors could discuss the work on DA methods that consider model error, e.g., weak constraint 4D-var, that could be used to resolve such problems. This would help the discussion on the overestimate of emissions lines 740 onwards.
7. In addition to problems with the DA of long-lived gases in the IA step, I do question the DA of short-lived gases, i.e., NOx. I am aware of unpublished negative results showing that NOx concentration DA has limited benefit for the model forecast skill and in the worst case negatively impacts results by perturbing ozone chemistry in unrealistic ways. Hints of such problems can be found in ECMWF reports (Flemming et al., 2009; Inness et al., 2009, 2015) that clearly show the limited efficacy of NOx concentration assimilation due to the decay of DA increments. The authors should discuss these limitations in a meaningful way.
8. Important details of the work could have been explained in a clearer and more concise way (specific comments to follow), e.g., why do the authors not include the length of DA window and analysis run the figure describing the architecture of their system? These details appear almost at the end of the descriptions, but such information, that can easily be compacted into concise form, could appear earlier either in figure 1 or in a suitable table. See specific comments below. There are also some contradictions in the explanations of the different experiments, which make it awkward to evaluate what has been done.
9. The DA experiments are only run for a single month. Given that this system will presumably be used in an operational capacity throughout the year it would have been nice to see statistics on the results of other experiments at different times of the year. I think this point is especially relevant for the pollutants with a strong seasonality in their emissions, e.g., PM2.5. Why was December 2016 chosen as the period of study? I could not see any discussion on this. Please can the authors add something to explain this.
10. Another point regarding the season chosen and the performance of the EI step. Since the EI system does not include emission estimates of ammonia, and ammonia can be very important to spring time loadings of PM2.5, could its absence from the system lead to compensation effects on the primary PM2.5 emission estimates? I think the authors should discuss this as a potential issue for the wider

application of RAPAS throughout the year. This might help to identify the necessity for future development paths.

Brandhorst, N., Erdal, D., & Neuweiler, I. (2017). Soil moisture prediction with the ensemble Kalman filter: Handling uncertainty of soil hydraulic parameters. Advances in Water Resources, 110, 360-370.

Flemming, J., Inness, a., Flentje, H., Huijnen, V., Moinat, P., Schultz, M. G. and Stein, O.: Coupling global chemistry transport models to ECMWF's integrated forecast system, Geosci. Model Dev. Discuss., 2(2), 763–795, doi:10.5194/gmd-2-253-2009, 2009.

Hamer, P. D., Bowman, K. W., Henze, D. K., Attié, J.-L., and Marécal, V.: The impact of observing characteristics on the ability to predict ozone under varying polluted photochemical regimes, Atmos. Chem. Phys., 15, 10645–10667, https://doi.org/10.5194/acp-15-10645-2015, 2015.

Inness, A., Flemming, J., Suttie, M. and Jones, L.: Chemically Reactive Gases, , (May), 2009.

Inness, A., Blechschmidt, A., Bouarar, I., Chabrillat, S., Crepulja, M., Engelen, R. J., Eskes, H. and Flemming, J.: Data assimilation of satellite-retrieved ozone , carbon monoxide and nitrogen dioxide with ECMWF ' s Composition-IFS, , (2), 5275–5303, doi:10.5194/acp-15-5275-2015, 2015.

Keppenne, C. L., Rienecker, M. M., Kurkowski, N. P., and Adamec, D. A.: Ensemble Kalman filter assimilation of temperature and altimeter data with bias correction and application to seasonal prediction, Nonlin. Processes Geophys., 12, 491–503, https://doi.org/10.5194/npg-12-491-2005, 2005.

De Lannoy, G. J., Reichle, R. H., Houser, P. R., Pauwels, V. R., & Verhoest, N. E. (2007). Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter. Water Resources Research, 43(9).

Specific comments

1. Can the authors briefly mention the ICNO experiment results in the abstract.
2. Sentence running from line 138-line 141. Please can the authors rephrase this? I had difficulty making sense of this sentence.
3. Please reformulate sentence running from line 147 to 150.
4. Line 172 onwards in section 2.1.1. Please can the authors find a way to shorten the summary text to avoid repetitions. There is already a lot of detail here some of which is repeated in the following sections.
5. I found no mention of the temporal variability of the emissions. The EI seems to be run for single days resolving an emission estimate for each day. Are the daily a posteriori emission estimates somehow applied to the hourly emission variabilities each day?
6. If the daily emission estimates from the EI are estimated for a single day and then applied to the following day as the updated a priori, how does this system behave due to day-of-the-week effects? Emissions from traffic and other sources are known to change at the weekend compared to the day, so how does this affect the day-to-day performance of the DA system when transitioning Friday-Saturday or Sunday-Monday? Further to that, how does this affect the hourly emission variability following point 5 above?
7. Figure 1. Please can the authors show the length of DA window in each step of this system.
8. Line 210. Would "interpolated" be more accurate than "compressed"?
9. I think the name of the ICNO experiment should be renamed to NODA or DANO to indicate more clearly that neither the IA or EI steps are performed. As it is, ICNO implies only not IA.
10. The descriptions of EMS1 in Table 3 and on lines 498-501 seem to be in contradiction of one another, i.e., Table 3 says EMS1 is initialized with prior emissions from the previous window (no indication of 3D-Var) and the text from line 498 claims that the ICs are from the 3D-Var experiment. The authors should make this clearer to avoid any confusion.
11. It would have been good to see an EMDA experiment without any 3D-Var to see what affect the 3D-Var has on the emission inversion step. I think this is an important test because of the potential overlap between emission error and initial condition error arising from emission errors.

12. I think Figure 10 should be expanded to show the changes in emissions beyond the borders of China. I am assuming of course that the EI system's state variables include emission terms beyond China's frontiers. If they do not, then I would ask the authors to make this point clearer. But given the large emission region (north India) included in the western part of the modelling domain, it would make sense to show any emission changes in that region.
13. I assume that the authors have performed some offline testing on the number of ensemble members in the EnKF. It would be interesting to hear about these tests and what they showed with regard to selecting 40 over any other number of ensemble members.
14. I think the application of the chi squared metric to the discussion in sections 4.3.3 and 4.3.4 would be very informative of how the optimality of the EnKF is changing in each case.
15. Line 854. When the authors speak about reduced emission uncertainties, I think some care is needed. In fact, the errors in the simulated concentrations are reduced, and from that the a posteriori emissions are assumed to have lower uncertainty. However, this ignores the fact that there are unquantifiable model uncertainties included within the new emission estimates, and so I think it is dangerous to say the emission uncertainty itself is reduced without the support of an independent estimate.
16. Line 872 onwards. This is probably also a symptom of the fact that the EnKF was not specifically developed to estimate state variables with significant bias errors. The incremental improvement described here are exactly symptomatic of this issue, and examples like this are described in the literature on the topic cited above. This is an example of the text that will need to be revised in light of the required shift to speak of parameter estimation to solve for biases being treated by the EnKF.
17. Line 878 to 883. There is an example of ozone observations being successfully used to estimate NOx and VOC emissions within the 4D-var framework (Hamer et al., 2015) that specifically deals with the problem of NOx-limited vs VOC limited conditions. I would suggest to the authors to consider including some text on the potential ways to address this problem. This might be one limitation of the EnKF method compared to the 4D-var?

Proposed Technical Changes (Grammar and Typos)

Line 37. "It is capable of…"

Line 38. "…assimilating spatially…"

Line 44. "…subsystem in each data…"

Line 103. "…and large amounts of …"

Line 126. "…because atmospheric…"

Line 139. "…means that the assimilation windows are independent from each other, generally,…"

Lines 176-177. "…and provides "perfect" chemical ICs…"

Line 199. "…can address the complex…"

Line 204. "…and it covers the whole of mainland of China…"

Line 217. "…using optimized emission from the previous window…"

Line 242. "Additional work includes the…"

Lines 256-257. "Hourly mean surface pollution observations within a 1 hour window of the analysis …"

Line 259. "For gas concentrations that are directly…"

Line 314-315. "The ground-level scale generally spreads 40-45 km…"

Line 320. “In the EnKF…”

Line 401. “…with a lifetime more than 1 day.”

Line 403. “In addition, NO2 is rather reactive…”

Line 424. “During the inversion cycles, …”

Line 439. “…assimilation cycle…”

Line 446. Missing last access date on link.

Line 582. “…are over the rest of the areas…”

I would recommend a proof-read by a native English speaker to try to remove the various remaining grammatical errors.