General comments:

The manuscript presents a comprehensive study aimed at estimating air pollutant emissions in China through the assimilation of surface observations. The authors found that the emission reduction efforts during the 2018-2020 Action Plan generally exceeded those of the 2013-2017 Action Plan. They also conducted comparisons with various bottom-up emission inventories, and provided detailed explanations for differences and uncertainties. These findings are relevant and potentially important.

However, while reading the manuscript, I encountered several unanswered questions, mainly related to the settings and parameters of the estimation technique, as well as potential uncertainties and biases in the inferred emission estimates. In particular, I have doubts about the credibility of the NMVOC emission inversion. I believe that further analysis and discussion addressing the major and specific issues outlined below are necessary to substantiate the authors' claims and make the manuscript suitable for publication in ESSD.

Major comments:

1. The authors compare the posterior results with other sources and frequently employ terms like "underestimate" and "overestimate" without explicitly specifying what is considered an under- or overestimate relative to a reference. For instance, in line 47, the use of these terms lacks clarity. More critically, the terms "underestimate" and "overestimate" imply that the posterior is inherently closer to the truth than the other sources, assuming that the other sources are less accurate. This assumption is not self-evident. In inversion, adjusting emissions to match observations does not conclusively prove that the posterior emissions are improved, nor does it inherently indicate biases in other bottom-up inventories.

To claim that HTAP and other sources are less accurate and to justify the terms "overestimate" and "underestimate," the authors need to provide a more convincing argument. Simply relying on posterior simulations is not sufficient to demonstrate the improvement in posterior emissions and the existence of biases in other bottom-up inventories. If a more convincing argument cannot be made, the authors should consider using more neutral terms to avoid implying a hierarchy of accuracy among different emission sources.

An example of inconsistency can be found in Section 4.3, where the authors, in comparing their emission inventory with others, occasionally use alternative inventories as a basis to highlight the agreement and reduced uncertainty of their inventory compared to bottom-up inventories. At other times, however, they claim that these alternative inventories exhibit significant uncertainties. This contradiction raises concerns about the clarity and consistency of the manuscript. It is essential that the authors provide a more coherent explanation or rationale for the varying assessments of uncertainty in other inventories.

2. In particular, the authors' comparison of natural and anthropogenic species emissions (such as PM10 and NMVOC) reveals a significant issue. Natural sources inherently exhibit considerable uncertainty, and in many regions, natural sources contribute significantly more than anthropogenic sources. Therefore, using the uncertainty in natural sources as a basis does not necessarily indicate large uncertainties in anthropogenic sources within the bottom-up inventories. An inconsistency arises in Line 790, where the authors' explanation appears contradictory. They simultaneously assume minimal variations in natural sources and cite literature indicating an increasing trend in natural sources. Additionally, the manuscript attributes emission changes to anthropogenic sources while acknowledging substantial uncertainty in natural sources. If it is acknowledged that natural sources indeed carry significant uncertainty (which is indeed the case), the manuscript should avoid using terms such as "not captured," "overestimated," or "underestimated" concerning the bottom-up inventories. These terms imply a clear attribution of error that may not be justified given the uncertainties associated with natural sources. Clear and consistent handling of uncertainties in both natural and anthropogenic sources is crucial for maintaining the credibility of the manuscript.

3. In Line 272, it is mentioned that VOC emissions are optimized through assimilating ground-level O3 observations. However, several factors need consideration. On one hand, VOC-O3 interactions involve strong nonlinear chemical reactions, and emission adjustments exhibit bidirectionality (Tang et al., 2016). Despite the convergence of simulations and observations, VOC inversion results may deteriorate due to these complexities. On the other hand, the majority of the national monitoring stations are situated in urban areas, whereas VOC primarily originates from suburban or rural regions. I am skeptical about the feasibility of assimilating O3 to constrain VOC emissions. As evident from Figure 3, the posterior simulations do not show a significant improvement in O3. As the authors noted, O3 cannot effectively constrain precursor NOx (L278). Therefore, I recommend deleting the VOC emission inversion.

4. The changes in observation coverage each year can significantly impact emission

estimates. If the authors intend to include the years 2013-2014 in this study, they should compare the impact of site differences on emissions for a more robust analysis. If the authors aim to investigate trends, it is advisable to delete emissions in the 2013-2014 period, as this might otherwise potentially mislead readers, given that the changes during this period do not contribute meaningfully to the study's overall trend analysis.

There also appears to be some discrepancies in the manuscript where emission changes are often stated as occurring from 2015-2017, while the text descriptions indicate the period as 2013-2017, as seen in lines 452, 485, and 561, among others. Furthermore, it is important to note that the changes in emissions observed from 2015-2017 not necessarily reflect the overall reduction rate of the action plan for the entire period of 2013-2017. Additionally, the data from 2015-2017 alone may not be sufficient to conclude that the emission reduction rate during the 2013-2017 period is lower than that during the 2018-2020 action plan.

5. The authors use PM2.5 observations to simultaneously constrain BC, OC, and primary PM2.5. If they do not consider inter-species correlations or use random perturbations, and, for instance, if BC and OC increase while PM2.5 decreases in one ensemble member. How do they constrain emissions when the simulated PM2.5 and observations are the same.

6. The authors simultaneously constrain concentrations and emissions, emphasizing that concentration errors arise from emission uncertainties, implying a shared source of uncertainty (L222). In this context, the question arises whether optimizing concentrations would diminish emission uncertainties, thereby affecting emission estimates.

7. The NOx emission changes optimized by the authors appear to contradict existing research findings and are inconsistent with recent emission reduction policies. Despite citing the study by Zheng et al., (2018), the actual NOx emissions reported by Zheng show a significant decrease. Could this discrepancy be attributed to the bottom-up inventory lacking sufficient statistics on mobile vehicle emissions? Moreover, according to Zheng's study, industrial and power plant emissions collectively contribute to over 50% of total emissions. Hence, the second reason provided by the authors may not be suitable if the industrial and power plant emissions are substantial contributors

8. "Figure 12 shows significant discrepancies between the results of the author's inversion and other bottom-up inventories. According to other literature, it is known that China experienced two peaks in VOC emissions in May and July 2015. The

variation in VOC emissions closely follows the changes in O3 levels, suggesting a strong dependence of VOC on O3 variations. This raises the question of whether non-linear changes are being overlooked.

9. The author has provided a high-resolution, multispecies emission inventory. To facilitate users' understanding of the data's accuracy, could you please provide information on the uncertainties associated with different species, allowing users to assess the error range in the data?

Specific comments

1. Change "Fengwei Plain" to "Fenwei Plain"
2. In L212, the VOC adjustment factor was omitted.
3. Since MOZART data products are no longer updated, are the boundary conditions in this study based on simulations conducted by the author's team? Additionally, maritime shipping emissions have a significant impact on the generation of NO2 and O3 in coastal provinces. Has the model taken into account inputs from maritime emissions? Why was a localized scale of 180 km chosen?
4. In L291, is it necessary to reassemble simulations for each iteration, or is it multiple inversions on the original ensemble? If multiple iterations are performed, does it imply that the posterior approaches the observations more closely with each iteration? Why was the choice made to iterate twice?
5. With such a high grid resolution of 15 km, how does the computational cost for the inversion of multiple years in the ensemble calculations? Additionally, what is the size of the assimilation window?
6. The inflation factor 'r' varies for each window. Is 'r' a matrix or a scalar? If it is a scalar, could the author provide the specific range of 'r'?
7. Table 3 lacks information regarding the year.
8. How were diurnal variations of the emissions specified?
9. How is the optimization of VOC components conducted when VOC consists of multiple components?
10. Region name in Figure 1 refers to specific areas. Consider a different expression to avoid potential ambiguity.
11. Please consider adopting a clearer representation for Figure 11.

12. L483 Change "Fig.3" to "Fig. 4"