Response to Anonymous Referee #1 of Lund et al. 2022

We thank the referee for the review of our manuscript and suggestions for improvements. Responses to individual comments are given below.

More substantial changes in response to both referees include:

- Table S1: simplified to show the relative differences discussed in the main text instead of absolute numbers that were not really used.
- Updated Figures 1 and SI1 to include NH3 and VOCs.
- Updated Figure 3 and 4 to include more regionally explicit information.
- Trend calculations refined to reduce influence of individual years.
- Moved radiative forcing section to after evaluation of AOD against observations.
- Larger modifications to the description of simulated AOD trends and RF following comment from referee #1 on inclusion of regional information.

General comments

This paper reports the effect of three different inventories on aerosol optical depth and radiative forcing simulated with one model. Two of the inventories are different versions of CEDS. Both magnitude and trend of aerosol optical depth are compared with MODIS.

I find the contribution of this paper to scientific understanding to be rather modest. The use of models to understand how anthropogenic emissions affect the atmosphere's aerosol content and climate is of course a worthwhile pursuit. But an interesting work would include careful diagnosis of the causes of difference, or their implications for radiative forcing and climate response in different regions. This work doesn't provide that.

We thank the reviewer for the feedback and the suggestion. However, we do not agree that a diagnosis of the underlying causes between inventories is the only relevant question to be answered. It remains a fact that the inventories, as they are provided, are used in many applications, as documented in our introduction. Given this situation, a quantification of the uncertainty this introduces to current projections is of high importance for the community. We have now made this framing, and key question, clearer in both the title and introduction of our manuscript.

We do, however, agree with the reviewers that more regional detail beyond the full 2D maps can be of interest for the reader. As we're sure the reviewer is aware, we are not able to diagnose regional climate responses beyond radiative forcing within the modeling framework used in our study. Such quantifications typically require more resource demanding simulations with coupled models. We therefore see our study as an important first step that, given the differences we find, can motivate and support spending the time on further work.

To better highlight the regionality, we have produced two new figures where we quantify and show the regional mean AOD and trend, and the regional mean RF (see end of this document). Here we use a set of regions broadly covering the globe and largely in agreement with the study by Lund et al. (2019;2020), as well as the coarser set of IPCC AR6 region definitions. The text has been modified accordingly, as well as in response to the comments below (see respective responses). We think this has significantly improved the manuscript and provides information of broader relevance to the more of the community.

Authors argue that the later version of CEDS was not included in AR6 and therefore it is worthwhile to analyze implications of the new inventory. That may be true, but the paper simply reports averages and shows spatial distributions. It doesn't provide much understanding of how or why the new inventory is different or whether it is more or less suitable to represent anthropogenic influence.

As described above (and below), we have added more detail about the regional changes and attempt to link them better to the underlying emission changes (and reasons, where such information is available). We also connect these regions better to the description of observed AOD trends, although the latter is of course influenced by more than anthropogenic emission and these factors can sometimes obscure the detection of anthropogenic emission-driven trends. Moreover, our existing comparison of the simulated AOD with MODIS over the Asian region where key emission biases are known to exist (changing estimates of Asian aerosol emissions and evolving inventories was a challenge for the assessment of climate change throughout the sixth assessment report of the IPCC) does confirm that our model is better able to represent the trends (and hence anthropogenic influence) with the updated emissions.

To the extent it's possible to derive from the limited documentation, we give a summary of the underlying updates to the inventories, with associated references. See also further responses to the comments regarding inventories below.

Specific comments

Emission inventories do not affect actual aerosol influence (as is suggested in the title), rather simulated influence. The purpose of a model is to attempt to reflect the real evolution. Certainly if one changes any flux (emissions) or any input then it changes the simulation, but this shouldn't be a surprise if one's model is working properly. Perhaps some understanding could be gained by exploring whether the model outputs (AOD, RFari, RFaci) scale with the inventory changes. This sort of analysis is hinted at, e.g. in lines 207-208, but for a helpful contribution to the community, much more analysis would be presented.

We see the point about the wording not being accurately representative for what we do here. The title has been modified to make it more precise in wording, now reading:

"Implications of differences between recent anthropogenic aerosol emission inventories on diagnosed AOD and radiative forcing from 1990 to 2019"

We have also made updates to introduction to better reflect this same point, as well as to clarify our aims and motivation.

Yes, changing any one flux or input changes the simulation. This is true for any modeling experiment. However; one needs to do the actual simulation to understand exactly how the change manifests. It is not given that the model will perform better with changed data, as there may be compensating issues affecting biases compared to observations, or competing effects when there are concurrent differences across emitted species. Furthermore, while some atmospheric species are more linearly connected to emissions, others are affected by complex atmospheric chemistry. The effect on atmospheric composition of equal changes in emissions can also be regionally dependent, as is the resulting radiative forcing.

We believe that the inclusion of more regionally explicit information as suggested by the reviewer, helps to highlight the importance and implications of our work. Throughout we have also made modifications to the text to better link the results with the input data, e.g. linking regional burden and AOD of individual species better with the underlying emission changes to see if they scale or not.

Unfortunately, we do not have the RF per component available, so linking forcing to inventory changes is not straightforward. We have, however, looked into the RF per dAOD, a measure sometimes used for the intercomparison across different models, for both global and regional means. Across different regions, this measure of course varies substantially due to the different aerosol composition, but also due to background climate and underlying surface characteristics. Within a given region, we also, in some cases, find marked differences in the normalized RF between the experiments using different emission inventories. For instance, the normalized RF averaged over Russia is consistently lower (by up to 20%) with CEDS21 and ECLv6 emissions than with CEDS, while for other regions both higher and lower normalized RF values are found depending on whether RFari and RFaci is considered. This likely reflects the complexity of the response when multiple emitted species are changed at the same time and sometimes with different signs. In some cases, these normalized numbers are also difficult to interpret given their sensitivity to the denominator. In particular where the AOD change is very small, the numbers can become very high. However, the diversity across region and inventory adds to the importance of conducting detailed simulations and not try to derive the response directly from the inventory change.

Emissions are a component of the physical system that are affected by processes. This paper compares different compilations of or assumptions about those processes. But ascribing these differences to the compilation label itself, eg 'CEDS', 'ECLIPSE' is overly simplistic. What assumptions have the inventory developers made that cause these differences?

The objective of this paper is to document results from simulations using some of the central recent inventories provided to the modeling community. The names of inventories are used to label the experiments in the analysis, and we describe and quantify the differences between these experiments using different input data. To try to clarify, we added a table describing the experiments and their names.

The modeling community is dependent on the provision of up-to-date emission inventories. However, their development is comprehensive and complex work, and the documentation can be limited. Hence, it is often difficult for non-experts to know the numerous underlying assumptions and data or understand changes. We would maintain that not being able to attribute differences in emission trends or magnitudes uniquely to specific changes in assumptions or bottom-up statistical data does not take away from the importance of quantifying the impact of these differences on the diagnosed quantities critical for assessing the air quality and climate implications of anthropogenic aerosol. In fact, rather the opposite, as it is not possible to assess the consequent changes in aerosol composition and distribution (nor RF) directly from the inventory differences, e.g. via scaling by emission changes. This requires the type of detailed modeling we have performed.

A comprehensive documentation of the assumptions that inventory developers have made or which data they have updated is beyond the scope of this study. We certainly think that such a study would be very helpful for the modeling community but would argue that this is a task for the inventory developers. In fact, a recent opinion piece by Smith et al. highlights the importance of development of consistent and transparent inventories for pollutants, which is not the case today.

However, some first order information about the link between emission changes and underlying updates can be derived from the limited literature that is available. We have expanded the paragraph describing the emission inventories. Where possible and most relevant, we also try to link the observed difference trends more clearly with underlying emission changes. By providing more regional information, in response to the comment above, this is better facilitated. Finally, we try to better highlight our aims and motivation in the introduction section, and point to why such simulations are needed (i.e. that the information cannot be simply obtain from scaling by emissions) in the results description.

Aerosol trends are discussed in lines 234-249 and 342-364. This sort of analysis could aid in identifying, explaining and quantifying differences in the input (emissions) and response (AOD, RFari). But the analysis presented here is rather broad ('weaker in magnitude', 'consistent with…') What are the implications for RFari and RFaci, since a global average is given for these measures relative to 1990?

We have quantified regional trends in AOD as well as regional mean RFs in the revised manuscript, where possible describing the connection with the underlying emission trend better.

The presentation also ascribes some masking of trends to interannual variability, especially among natural (sea-salt) or biomass burning emissions. This is a well-known issue in comparing model results with observations. It would seem that model evaluators should have some set of best practices about how to account for this effect after many years of such studies, such as using running means. Otherwise, the persistent inability to draw conclusions will render all such studies only marginally useful. Do authors have thoughts on this?

We certainty agree that model comparisons with observations of various spatiotemporal resolutions and type is not straightforward, which is also well known in the community. Many approaches exist (we for instance use a linear least square fitting and 10-year boxcar average), but if there are opportunities to unite the community better around a common, optimal approach, that seems like an effort that is worthwhile to pursue in further work.

In the revised manuscript, in response also to comments by referee #2, we have changed the calculations of trends to minimize the influence of single years or start-end yeas (mainly due to the influence of biomass burning, since our dust and sea salt fluxes are kept constant in the fix met runs). We now calculate trends with one and one year removed and show the average of this set of coefficients. We note that the difference is small and the change in method does not affect our overall findings. We have also made some modifications to the paragraph on sea salt following referee #2.

It may be worthwhile to define how well one needs to know the forcing since emissions and other inputs are always uncertain. Then one would have more confidence in stating a 'strong effect' as is done in the title.

This is an interesting question and one we don't have a clear answer to. However, we agree that the use of "strong" in the title can be perceived as subjective without a definition such as suggested by the referee. In light of this and the comment above, we have modified the title.

The paper acknowledges analysis by other researchers on the same topic, e.g. Lund et al 2018, Mortier et al 2020, Quaas et al 2022. However, other than broadly comparing findings, this work doesn't indicate what new insights it has offered – what has been done here that wasn't done before, and if another future paper is done with similar approach, what questions should it attempt to answer? There seems to be a limited review of prior work, especially considering that Asian emission inventories have been evaluated against observations, and those emissions are also stated to play an important role in this work.

We have modified the introduction section to clarify our motivation and aims. Upon submission this was, to our knowledge, a first paper comparing these three inventories which makes it difficult to compare directly with other literature. Furthermore, our objective/scope is not to provide a review of emission inventories or aerosol trends, but rather to evaluate our modeling tool, assess whether the availability of new inventories improves the model performance or highlights other issues, and understand to what extent the known biases in emissions have affected assessed aerosol-induced climate effects (or may even continue to do so given that studies are still published using the older generation emissions data). We have tried to make this clearer in the introduction and conclusion. Furthermore, we have added more references to aerosol uncertainty and climate effects and work on Asian emissions.

Technical comment

I found the paper well written and I did not note technical corrections to make it clearer or more accurate.

Thank you.

Draft new figures 3 and 4 (larger size, better resolution in revised manuscript):