Reply to referee#3

The authors would like to thank anonymous referee #3 for the careful review and the helpful comments.

In the following, the reviewer's comments will be in **bold** font, and the responses will be in plain font, with suggested new text in *italics*.

The authors used four inverse models to estimate European emissions of HFC-134a, HFC-125 and SF6 for the year 2011. All systems used measurements from Jungfraujoch, Mace Head, and Monte Cimone. The paper is well written and provides interesting insights. I think the main problem of the paper was that the differences in the choices, such as spatial correlations of the prior and background treatment, had a quite substantial impact on differences among the models. What was the reason that those were not controlled? If they were better controlled, maybe we could have had more insights on which models are doing better and what we might do to improve the emissions estimation through inverse modeling. Below are some other comments and questions I had and I would recommend publication after they are addressed.

The main motivation was to document the uncertainty associated with the different choices that have been made in recent halocarbon inversion studies. There is no doubt that differences would have been substantially smaller with a more strongly controlled setup. It was not our intention to assess the quality of the transport simulations of FLEXPART as compared to NAME, which it would ultimately come down to if all other choices were identical.

For Figure 1, is this the sensitivity created using FLEXPART or NAME? I would also assume that the sensitivity is quite different depending on the month. Which month is this? And is this the monthly mean?

We apologize that the figure caption was not sufficiently clear (as also noted by another reviewer). It will be changed to

Annual mean surface sensitivity in units of [ppb/(kg m-2 s-1)] for (a) the original 0.1°x0.1°grid and (b) for the reduced grid of the FLEXPART-based model system EMPA.

It was a little unclear why NAME needed such a high release height at Jungfraujoch. If the point of the paper is to better understand the differences among the four inversion systems, I find it puzzling that the authors would modify to make the model footprint sensitivities comparable to each other.

As also mentioned in our reply to reviewer #2, the measurements from Jungfraujoch have not been used in previous inversion studies based on the NAME model, because the results had not been satisfactory for this site and no independent analysis on the optimal release height had been conducted before, in contrast to FLEXPART. The approach chosen here was pragmatic, so as to not disadvantage the NAME model, and allowed us to include the results of NAME despite of these difficulties. A thorough investigation of the reasons for the differences between FLEXPART and NAME for Jungfraujoch would be desirable, but was not feasible within the scope of this project.

I had a hard time understanding how the emissions were created following the country outlines. What was the means used to split the EDGAR grid to country outlines? Also, because the prior emissions are so different, I find it more informative if the Fig. 6 was not comparing between prior and posterior but **EDGAR and posterior.**

We will add the information that the original resolution of EDGAR was 0.1° x 0.1°. In the case of the UKMO system, EDGAR emissions were first regridded to a fine grid, and the country outlines were then followed as closely as possible. Each grid cell was assigned to the country with the largest share. Except for small countries, the error introduced by this procedure with cells at the borders shared by

more than one country is small. We extended the sentence explaining the grid for the UKMO system with

follows the country outlines ***as closely as possible given the resolution of a fine grid uncerlying the reduced inversion grid***.

EDGAR is the prior. Note that the prior emissions are not different, only their spatial representation. Comparing the results with EDGAR at the original resolution would require redistributing the emissions estimated on the reduced grid to the original fine grid. This would be doable technically, but it would give a wrong impression of high resolution of the inverted emissions. We strongly prefer the present representation of the results.

Why did EMPA2 use the uncertainty set uniformly to 137%? This seemed a little strange and was curious for the reason behind this specific value.

The 137% was a result of the requirement, that the total uncertainty of a domain covering most of Europe was 20%. We will add this information to the text.

One of the explanations for why UK's estimated emissions are much higher than what is reported to UNFCCC, the authors mention the use of an assumed high loss rate of HFC-134a from car air conditioning systems in the UK. Why is this only in the UK and how different is the loss rate among the countries? Is a similar explanation possible for overestimation and/or underestimation for different species?

The UK inventory is conservative (overestimates), as it assumes that there is a 100% replacement of air conditioning fluid in all mobile air conditioning systems each year. Each country makes their own choice in this aspect provided it is backed by expert knowledge. It is not clear what every country across Europe does in this respect. This particular situation is specific for HFC-134a but other issues will undoubtedly impact the emissions of different countries for different gases.

Backwards mode time differ substantially among the models and I would have expected UKMO to have larger difference between prior and posterior away from the measurement sites, compared to the other model systems that have shorter time span. Why is it that UKMO shows almost no difference between the two farther away from the measurement sites?

A backward simulation over 5 days captures most of the sensitivities of the measurements to emissions within Europe because the sensitivity decreases very rapidly as time and distance from the measurement increases. Extending to 10 days (NILU) or even 19 days (UKMO) changes little. The fact that UKMO adjusts relatively little at larger distances from the sites must be due to the specific choices of a priori versus observation uncertainties.

Minor comments

1. Sometimes authors state the country by name and sometimes by the ISO2 convention country code. It is a little confusing to me and so I would suggest to be consistent and I would appreciate if there was a table listing the country names with ISO2 code if the authors want to use the codes.

Instead of adding another table we added the country names in the caption of Figure 11:

CH=Switzerland, DE=Germany, IT=Italy, FR=France, ES=Spain, PT=Portugal, UK=United Kingdom, IR=Ireland, BE=Belgium, NL=Netherlands, LU=Luxemburg, AT=Austria, DK=Denmark, SW=Sweden, FI=Finland, PO=Poland, CZ=Czech Republic, SV=Slovakia, NO=Norway.

2. P. 11 l. 4 "An important question is the context … is the question" -> delete the second "the question" in the sentence to make it "… Paris Agreement is, how suitable is…"

Thank you, done

3. I am not quite sure what 0.1°x0.1°min means in Table 1.

It means that the minimum size of a grid cell is 0.1°x0.1°, but larger when cells are aggregated for the reduced grid. We will change "min." to "minimum".

4. “reduced acc. to” -> “reduced according to” in Table 1 for UKMO

Done

5. State vector length is mentioned in Table 1 but was not explained in the text at all. Can this be clarified in terms of how this is used in the equation and why the equations look so different depending on the system?

We don't quite agree: The state vector ***x*** was introduced in the context of equation 1 as follows: "***x*** *is the state vector which includes the gridded emissions and possibly other elements such as background mole fractions, and n is the number of state vector elements to be estimated/optimized by the inversion*." The state vector was also referred to at other locations, e.g. in the sentence "*In the EMPA system, a single element per observation site is added to the state vector to represent the background at time step k.*"

To better link the information in Table 1 with the text we will add the following line after the first sentence mentioned above:

An overview of the number and type of state vector elements used in each system is provided in Table 1.

Note that the equations presented in the manuscript do not depend on the specific choices of state vector elements.

6. How is the EDGAR prior uncertainty determined in Figure 13? I find that to be a little misleading, since I do not think EDGAR provides such a value.

EDGAR indeed does not report uncertainties. These uncertainties denote the range of uncertainties in the prior that is introduced by the different gridding methods. This information will be added to the caption.

7. Figure 14 is very difficult to see – maybe a different color scheme would work better.

We played a lot with different color schemes and found that using a single color in different shadings works best. The main issue is that the different grid shapes already introduce a lot of variation, such that the figures become too complex and less easily readable when using a color scheme composed of multiple colors.