Reply to reviewer #1

We thank the reviewer for the helpful comments, and are confident that the revisions encouraged by the reviewer have resulted in an improved manuscript. Below we address the main points put forward by the reviewer. We found no issue with the minor comments, and have implemented all of them in the revised manuscript.

1. *Use of tense in "Abstract" and "Summaries and Conclusions" are not consistent. I have noticed that the authors swap back and forth between present tense and past tense in these two sections. The common practice would be use consistently one tense through abstract and conclusions sections.*

We have corrected this.

2. *Personally, I found the current version of Abstract not fully capturing the essence of the findings from this work. While the results presented in the main body of the text are important additions of the existing literature, the abstract only includes rather general and vague discussions and no clear identifications of what are the crucial parameters that needed to be considered if one is to adopt the two-box model approach, etc. A reorganization of the abstract, with clear emphasis on what are the key findings of this, in the context of previous literature, would be helpful to readers.*

We have restructured the abstract, to better reflect the key findings of the paper.

3. *P12, last paragraph and Figure 2. The positive trends in the IH exchange rate for CH4 and SF6 are very puzzling. Based on the results presented, it is not convincing to arrive at the conclusion that these trends are due to acceleration of IH transport of air mass or a shift in the pattern of IH transport. Are you sure this isn't due to a model spin-up related issue? If it is indeed the change in the IH transport rate, there must be a way to quantify this using the proper diagnostics from the TM5 model or set up sensitivity runs to tackle the problem. At this point, it sounds very hand-waving to arrive at the conclusion that it is due to changes in IH transport with no real analysis backing up the conclusion.*

As mentioned in the manuscript, we ran a simulation with annually repeating meteorology. As expected, for CH_4, the IH exchange coefficient we derived from this simulation was near-constant, with no spin-up effects evident. Moreover, a simulation in which the meteorological fields of 1992 were repeated resulted in slower exchange than a simulation with 2012 fields. The only difference between those simulations and the simulations that give a trend in IH exchange are the meteorological fields that were used. This provides a solid basis for our conclusion that the meteorological fields (+ their treatment in TM5) are driving the trend. We discuss this issue now in more detail in Supplement 4, along with other sensitivities of the exchange rate.

4. *P15, 1st paragraph. As discussed in section 3.3.1 last paragraph in Liang et al. (2017), if one were to use the surface measurements to perform two-box model calculation, the NH surface to SH surface IH transport time is needed in the gradient-to-emissions calculation to be consistent. On the hand, if the entire tropospheric air mass is considered, the IH transport timescale is significantly reduced (as demonstrated in this paper) due to the nature of cross-hemispheric transport in the troposphere. More details can be found in Liang et al, section 3.3.1. The authors seem to have the two methods (or concepts) mixed when discussing this. It would be good to add a discussion on*

these two different conceptual models if the authors wish to compare the results presented in this work with those discussed in Liang et al. to avoid confusion.

The difference in approach is whether the exchange time is defined with respect to the surface gradient (as in Liang et al., 2017) or with respect to the tropospheric gradient (as in this study). As noted in Liang et al. (2017), our approach will likely result in faster IH exchange, and it is conceivable that there will also be time-dependent differences between the two approaches.

We cannot define our parametrization with respect to the surface gradient, because the 3D model budget would not be closed. Therefore, we cannot explicitly quantify how our results would be different if exchange is taken proportional to the surface gradient. However, it is possible that the large variations we find in the exchange rate of MCF could also be there for an exchange rate defined with respect to the surface gradient. Therefore, we deem it important to point out the difference in results, even if the two methods are not identical. If 3D transport models can provide us with inter-annual variability in IH exchange, then we should try to use this information.

To explain this issue more clearly, we added:

Additionally, for the parametrization, Liang et al. (2017) used hemispheric mean mixing ratios derived from the surface network, whereas we based mixing ratios on the full hemispheric troposphere in TM5. **P22 Lines 15-17**
Some of the differences may be explained by the definition of hemispheric mean mixing ratio (surface-based versus full troposphere), but further reconciliation of the two approaches in future research is necessary. **P22 Lines 22-24**

5. *P22, L26-27. "In the end, conclusions from our study and those drawn by Rigby et al. (2017) and Turner et al. (2017) remain qualitatively similar". Isn't this a much more important conclusion than the way it is presented here in this paper? Despite all the other details discussed, e.g. box-model simplifications, bias corrections, etc., this paper confirms the findings from Rigby et al. (2017) and Turner et al. (2017).*

We indeed confirm the results of previous studies, and we added a line to the Conclusions to acknowledge this more fully (see below). However, this could also be expected: due to the undetermined nature of the problem, improvements in any one parameter will not significantly improve constraints on OH. Despite this insensitivity, improving constraints on parameters one-by-one seems like an obvious way forward. Therefore, one of our conclusions is that we can use information from a 3D model to improve constraints on four parameters in the inversion. Even if this does not directly translate in improved constraints in the end-products (emissions, OH), it can be a useful first step. We think that for future research this is an equally important result, compared to agreement between two/three very underdetermined two-box model inversions.

This indicates that significant uncertainties in parameters unrelated to the identified biases remain, and these uncertainties attenuate the impact of the biases on an inversion. As such, we confirm in large part the conclusions drawn by Rigby et al. (2017) and Turner et al. (2017) regarding the underdetermined state of the problem. **P23 Lines 21-23**

In addition, I found Sections 4 and 5 somewhat wordy. While lots of details are discussed, it is hard to drawn the main conclusions, e.g. what are the important details that one needs to consider when conducting box-

model-based calculations. Some reorganization of the discussion and conclusions and emphasis on the key factors/uncertainties/parametrizations can be helpful.

Reviewer #2 commented similarly on the length of some parts of the manuscripts. Per your suggestions, we have restructured parts of the discussion and conclusion and sharpened our phrasing.