# Reply to reviewer #2

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 first in detail the general comments. We also discuss the specific comments, in as far as our implementation and choices required further explanation. Points not mentioned were implemented without issue.

## General comments

1. *While the paper is generally well written, it is quite long and at times a little too verbose. This is partly because some of the issues are technical and, for the most part, are rightly discussed in depth. I've made a few suggestions where the text could be clarified or shortened below. However, I would encourage the author team to take another look over the manuscript as a whole and try to re-structure and re-focus parts of the text and shorten some bits.*

Reviewer #1 similarly commented on the wordy nature of parts of the manuscripts. We have sharpened and shortened the manuscript, partly based on the reviewers' comments and partly based on our own insights. These adjustments mainly concerned the abstract, the discussion and the conclusions.

2. *Wouldn't the site-specific mole fractions from the 3D model be biased high because of the presence of "local pollution" in the model? As I understand it, NOAA conditionally sample data to select background mole fractions where possible, and then additional filtering is applied to remove obviously "polluted" samples. In contrast, the 3D model, particularly at the coarse resolution used (and I'm assuming for monthly mean output?), will almost always be influenced by nearby sources at most NOAA sites. The authors should at least mention this. Ideally, local pollution could be removed in the model mole fractions.*

For our analysis, TM5 was not sampled from monthly mean fields (!), but samples were taken by interpolation between the 3-hourly fields in time, and linearly in space. This means that as long as meteorology was simulated correctly, our model-sampled observations should be of air from clean air sectors, similar to the NOAA sampling strategy.
However, it's true that our TM5 simulations were done at the coarse resolution of 6x4 degrees, so that some additional sample pollution could occur due to colocation of sources with sample sites in the same grid box (though in principal spatial interpolation is implemented to correct for this).
For this reason, we performed a quality check of the model-sampled observations. For the check, we removed the trend and seasonal cycle from the $CH_4$ and MCF observational records, and we investigated the spread in the residuals per site as a proxy for pollution. We quantified this in the annual mean residual standard deviation (RSD), which will be higher in the presence of frequent pollution events. For both MCF and $CH_4$, we find that the RSD in the model-sampled and real-world records agree well at most sites. At a few sites the RSD of model-sampled observations is slightly higher than that of real-world observations (indicating over-polluted sampling in the model), but at others it is the other way around. As such, we did not find any evidence of systematically higher pollution of model-sampled observations, relative to real-world observations.

As this is an important issue, we now added a brief comment on the quality check to the manuscript :

*We checked that the TM5-derived observational timeseries were not systematically more polluted than the real-world NOAA-GMD observations. For this we detrended and deseasonalized the $CH_4$ and MCF timeseries per surface site, and quantified the spread in the residuals. At most sites, we found no offset between residual spread found in the TM5-derived versus the real-world timeseries. At a small number of sites, TM5-derived timeseries showed more spread in residuals, while at others the spread was less. Therefore, we find no evidence for systematic biases in TM5-sampled observations..* **P22 Lines 27-32**

3. *In Section 2.4, I think it might be worth clarifying how the "combined biases" run was done. I assume this was from a single model run, outputting all time-dependent quantities, rather than summing the individual biases?*

This is explained in Section 2.4. Indeed, to obtain the results for combined biases, we ran the two-box inversion with all four biases corrected for simultaneously. We feel we already cover this subject by treating it in its separate section. For example:

*Through comparison of the outcome of the standard inversion and an inversion with one or more biases implemented simultaneously, we can evaluate the individual and cumulative impact of the biases on derived OH and $CH_4$ emissions.* **P11 Line 20**

*For this purpose, Table 3 presents five metrics for each of the two inversions, as well as for inversions where we implemented the bias corrections one-by-one (taking standard settings for the other parameters).* **P19 Lines 1-3**

4. *Am I wrong to be left with the impression that, in order to avoid these biases, rather than throwing away 2-box models entirely, we could just reparameterize them based on the outputs of 3D models (as was done in the inversions in this paper)? There are still good reasons why we might want to do this. For example, Rigby et al. (2017) used an MCMC approach which required many thousands of model evaluations. This is challenging with a 3D model, but fairly trivial with a box model. Therefore, if we could use a small number of 3D model runs to derive better parameters, we could still use the advantages of a Monte Carlo inverse method (e.g. non-Gaussian distributions, non-linear models, etc).*

This is a fair point that was not sufficiently acknowledged in the original manuscript. Indeed, a two-box inversion allows incorporation of results from multiple 3D transport models, and this is an important advantage it holds over any one 3D transport model. Moreover, computational efficiency is a great advantage for many reasons. Therefore, we have added some additional discussion to emphasize this potential use of our analysis:

*The identified two-box model biases contribute to the already significant uncertainty in derived OH, and properly accounting for them can be a piece in the puzzle of improving constraints on OH. Moving forward, a likely next step is to incorporate more tracers in an effort to further tighten constraints on OH. In such a scenario, the tracer-dependent nature of the biases will likely increase the bias impact, and a proper 3D model analysis for each tracer becomes even more important. Already, efforts have been made to do so (Liang et al., 2017), and in this study we provide further suggestions for such an approach. A distinct advantage in this approach is that information from multiple 3D transport models can be used to tune the two-box inversion, making the inversion outcome less reliant on transport parametrizations of any single 3D transport model. Additionally, computational efficiency of simple models allows for complex statistical* *inversion frameworks, incorporating, for example, hierarchical uncertainties (Rigby et al., 2017).* ***P23 Lines 27-34***

However, as a counter-point there are also significant downsides to such an approach. For example, the inversion becomes sensitive to the settings used in the 3D transport model, e.g. source-sink fields. Sensitivity analyses with different OH fields and different source fields should be made in all the different model configurations. Additionally, if a two-box inversion suggests an adjustment to certain state parameters (e.g. emissions or OH), then it should be tested again whether the 3D model-derived bias corrections are sensitive to these adjustments.

Then there are issues that are difficult to resolve at all in a two-box model. For example, the latitudinal gradient of MCF minimizes in the tropics, post-2000. Thus, IH exchange of MCF is mostly driven by the slight IH asymmetry in this minimum, rather than by the overall IH gradient, which is the parameter that is optimized in a two-box model. Very likely for this reason, we find that the derived IH transport rate for MCF is sensitive to the demarcation of the two hemispheres. This uncertainty does not reflect any real uncertainty in the 3D transport model, but is rather an artefact resulting from the two-box parametrization. We find it hard to resolve this issue.

So while our analysis is indeed an example of how an approach suggested by the reviewer could be implemented, there are many challenges that remain before such an analysis actually resolves all of the problems we have identified. We encourage such work, of course, but in some ways a full 3D model inversion might be easier and more complete, and better represent reality.

## Specific comments

*P1 L2 and several other places ("In two recent studies: : :"). A box model was also used in McNorton et al., 2016. However, I believe this was a 1-box model. Do you also want to discuss this? Or limit your discussion strictly to two-box models?*

We have added more emphasis in the introduction to the fact that we really mean to focus on two-box models. While there might be some overlap, most issues we've identified are specific to the two-box model and would be different in a one-box. Due to the usefulness of the IH gradient we really do think that future work regarding the problem of OH will also involve models of (at least) two boxes. We do now refer to McNorton et al. (2016) in more general terms, as the work is relevant to our study.

*P1 L20 – L25: Given that the main results on OH and CH4 anomalies aren't too dissimilar from the two-box model studies compared to the other uncertainties (Figure 6), is it fair to say that it is "crucial" to use a 3D model? (Also see general comment 4).*

The small impact on the final conclusions reflects the many uncertainties of the problem. Even if we know the four parameters we've derived exactly, as is assumed in our two-box inversion, we still find a very uncertain solution. In large part due to this large uncertainty of the final solution we find agreement with existing literature.

However, moving forward, the objective of future research will be to reduce the uncertainties on derived OH. Given that the final uncertainty results from uncertainty in many parameters, reducing the uncertainty on any one parameter will not solve the problem (as is reflected by our results). However, if incorporation of information from a 3D transport model allows us to reduce uncertainty in a few of these parameters, then that does seem like a crucial first step, even if the immediate impact is not directly noticeable.

This is also related to the reviewer's later comment on observational uncertainties. Again, even with lower observational uncertainties, the problem might still be strongly underdetermined. However, piece-by-piece the combination of these kinds of improvements should put us on the right track to converging constraints on OH. The fact that the large bias corrections and the large differences in observational uncertainties do not significantly affect the final solution, is a testament to how hugely uncertain the problem was to begin with.

*P2 L10: What does "explicitly contain much information on a species' distribution" mean? Does this mean that there is no information on longitudinal gradients in mole fractions, etc. If so, what is this trying to imply?*

We mean that there is little spatial information included in a one- or two- box model, e.g. the tropical maximum of OH is not captured. Many of the biases we derive are driven by non-linearities between a species' distribution and varying source-sink fields (see e.g. Supplement 3). Use of "explicitly" acknowledges that it might be included implicitly through parametrizations derived from a 3D transport model.

*P2 L25 (e.g. Montzka et al., 2000; Bousquet et al (2005)). I suggest referencing some of the earlier papers here (e.g. Lovelock, Prinn)*

We already referred to these studies in the next sentence.

*P6 L2: suggest change to "leave more freedom with respect to the timing of emissions". However, I'm not sure what this assertion is based on? Is this justified? If so, how?*

In Rigby et al. (2017), the emission model allows the emissions to be shifted between decades. In our emission model, the shifting occurs between years. Though there are different constraints on the shifting in our model, in practice our model still results in more freedom with respect to the timing of emissions.

Our emission model results in uncertainties that roughly agree with those reported in McCullogh and Midgley (2001). These uncertainties are very high in the years where production was phased out (e.g. a 2-sigma range of 15.0 to 65.1 Gg/yr in 1997), and so we sought to reflect these uncertainties in our model. Emission timing, to us, seems a very uncertain uncertainty. In the absence of conclusive evidence that the Rigby et al. (2017) emission model is a better reflection of the actual uncertainties, we will retain our current approach.

*P12 L20: What does "the treatment of this data in TM5" mean?*

This phrase relates to how meteorological fields (wind, temperature …) result in transport of tracer mass in TM5. An example would be the parametrization of convection, which can have a large influence on interhemispheric exchange (e.g. Tsuruta et al., 2016), and the pre-processing of the meteorological data to create mass-conserving transport in TM5 (Bregman et al., 2003). The point is that there has to be a trend in some meteorological parameters for the final tracer transport in TM5 to exhibit a trend, but that not necessarily every 3D transport model will show a similar trend in the end-product, due to different sensitivities to meteorological parameters. We deem this issue sufficiently covered in the manuscript.

*P13 L21 – L26: This seems relatively important. Worth showing as dotted lines in Figure 2?*

It is important. Also resulting from a comment by Reviewer #1, it seemed useful to visualize the effect of various sensitivity tests on the IH transport rate (hemispheric demarcation; nudging; annually repeating meteorological fields). For this purpose, we have included the information in additional supplemental figures, so as not to overcrowd the figure in the main text (Supplement 4).

*P15 L9 – P16 L8: Is this long discussion of (random?) uncertainties really necessary? To me it seems that this is a side issue that distracts from the main message (the biases). Perhaps cut this, or shorten substantially.*

We think this is actually a very important issue. The gap between observational uncertainties used in Rigby et al. (2017) and Turner et al. (2017), and that reported by NOAA (at least for $CH_4$) is a factor 7. Given that the shared conclusion of the two studies is that the problem of constraining OH is strongly underdetermined, it seems crucial to resolve which observational uncertainty is correct.

From the TM5 simulations, we find much lower uncertainties than Rigby et al. (2017) and Turner et al. (2017), more in line with the observational uncertainty reported by NOAA, and also derived by ourselves. These lower observational uncertainties can be an important step in reducing the final uncertainty on derived OH variations (see also comment above).

*P18 L1: This is an interesting discussion. However, I'm interested in the relative magnitudes. Would treating the stratospheric loss as a function of the strat-trop gradient get us most of the way to removing the "bias" (or would the addition of the stratospheric box be a pretty good way forward)? Or do you really need resolved stratospheric transport in detail to address most of this issue?*

*P22 L11: Is the use of "persistent" correct here? Aren't you referring to a transient effect. Also, similarly to comment on P18 above, do we know how much of this drop might be expected with a model that attempts to resolve the stratosphere (e.g. Rigby et al., 2017).*

If we take stratospheric loss proportional to the tropospheric abundance, we find a reduction of 68% in the loss rate. If we take it proportional to the stratosphere-troposphere gradient, we find a reduction of 63%. We have added these numbers to the main text. Thus, the slow-down is really related to transport, and can hardly be corrected for by using a stratospheric box.

As for the use of persistent, we intend to say that it's multi-annual and not random year-to-year variability. Indeed, if at any point the emissions stop decreasing, then the loss rate will recover, so that it is indeed a potentially transient effect. We have clarified this in the text.

*S1: Wouldn't it be safer to omit months where there was a significant fraction of missing data, rather than interpolate from nearby stations?*

Omitting months from individual stations results in a changing surface network. This is very undesirable, as it results in large jumps in mixing ratios, which reflect the large systematic uncertainties in the global mean mixing ratios we derive. However, as long as these uncertainties are systematic, they have no impact on the derived growth rates, which provide the main constraint in our inversions.

While interpolation has its own uncertainties, it does circumvent the more significant uncertainties that a changing station network would cause. In general, we found that the site-to-site ratios we used are quite constant through time, as all site pairs are background sites located relatively close to each other (e.g. SPO/CGO and BRW/ALT). Finally, the (small) uncertainties we derived from the TM5 simulation show that the technique works remarkably well.

**References**

Bregman, B., Segers, A., Krol, M., E Meijer & van Velthoven, P. On the use of mass-conserving wind fields in chemistry-transport models. *Atmos Chem Phys 3, 447–457, 2003.*

McCulloch, A. and Midgley, P. M.: The history of methyl chloroform emissions: 1951–2000, *Atmospheric Environment, 35, 5311–5319, 2001.*

McNorton, J., Chipperfield, M. P., Gloor, M., Wilson, C., Feng, W., Hayman, G. D., Rigby, M., Krummel, P. B., O'Doherty, S., Prinn, R. G., et al.: Role of OH variability in the stalling of the global atmospheric CH4 growth rate from 1999 to 2006, *Atmospheric Chemistry and Physics, 16, 7943–7956, 2016.*

Rigby, M., Montzka, S. A., Prinn, R. G., White, J. W. C., Young, D., O'Doherty, S., Lunt, M. F., Ganesan, A. L., Manning, A. J., Simmonds, P. G., et al.: Role of atmospheric oxidation in recent methane growth, *Proceedings of the National Academy of Sciences, 114, 5373–5377, 2017.*

Tsuruta, A., Aalto, T., Backman, L., Hakkarainen, J., van der Laan-Luijkx, I. T., Krol, M. C., Spahni, R., Houweling, S., Gomez-Pelaez, A. J., van der Schoot, M., et al.: Development of CarbonTracker Europe-CH4—Part 2: Global Methane emission estimates and their evaluation for 2000–2012, *Geosci. Model Dev. Discuss, 2016.*

Turner, A. J., Frankenberg, C., Wennberg, P. O., and Jacob, D. J.: Ambiguity in the causes for decadal trends in atmospheric methane and hydroxyl, *Proceedings of the National Academy of Sciences, p. 201616020, 2017.*