We thank the reviewer for the insightful comments. Most comments concern the configuration of the inversions. Often, we agree with these comments. Based on the lessons we have learned during our work on the inversions presented in this manuscript, we would design a new set of inversions differently. In this response we would like to make two points:

1. We did run test inversions and we did consider all relevant previous work when we designed our inversions, and so we designed these inversions as best as we could, with the information available.

2. We have learned valuable lessons from our work that we consider worth sharing with the scientific community.

Below, we discuss the reviewer's comments in more detail.

1) The inversion set-up. Unlike previous studies (e.g. Bousquet et al. (2005)) that together optimized MCF destructions by OH and the ocean sink, this study only optimized OH and applied the first-order ocean flux. Thus the inversion results may be largely impacted by the uncertainties in the prescribed ocean flux.

We agree with the reviewer that the ocean sink can play an important role in the MCF budget, as becomes apparent also in our manuscript. Our treatment of the ocean sink is based on pragmatic considerations. During test inversions we already diagnosed possible convergence problems. The inverse system is very delicate and to derive a physical realistic solution we had to carefully balance all the terms in the cost function. Therefore, in the design of this study, we considered the expected difficulty of co-optimizing an ocean sink to be more important than the possible advantage of including it.

In our manuscript, we do consider the impact of this choice. Partly, because we show that using an ocean sink different from the one in our standard inversion can have a significant impact on the latitudinal distribution of MCF, for example in Supplement S3. We also discuss the potential impact of an uncertain ocean sink on our results in Lines 420-427 of the discussion. To summarize the results, the impact on the derived interannual variability of OH will likely be small, but the trends and latitudinal distribution of OH could be affected more significantly. Another reason to not optimize the ocean sink is that we do not think that the NOAA surface network can separate destruction by OH from a Wennberg-like ocean flux, since both have a similar latitudinal distribution.

The reviewer refers to Bousquet et al. (2005) for an example where co-optimization of an ocean flux does work out. The inverse system employed in that study differs significantly from the one we use, and consequently it has different strengths and limitations. Whereas we calculate a comprehensive cost function gradient in each iteration with the adjoint model, they approximated the sensitivity of monthly mean MCF observations to a limited number of state variables in a set of forward simulations. This makes their inverse estimation computationally more efficient, because once the source functions are calculated, the 3D transport model doesn't need to be run anymore. On the other hand, we were able to optimize grid level emissions of MCF, as well as OH in 45 latitudinal bands, with individual surface observations. The source-function approach necessitates a large-region approach with monthly mean observations to remain computationally viable. In fact, one recommendation in Bousquet et al. (2005) was to move away from this large-region approach, which we have done.

In hindsight, we are not sure which approach works best. The increased complexity of the inverse problem that we tried to solve is likely an important reason that we ran into convergence problems. The added value of individual observations over monthly mean observations is questionable, and the source-

function approach is an elegant solution to the computational expense of a multi-decadal inversion. However, these are lessons we have learned that we want to communicate to the science community.

We extensively discuss the convergence problems in our manuscript and we have tried to assess the possible impact on our results, for example by reducing the problem to a ten-year inversion (Supplement S4). Based on all of our findings, we remain confident that the timeseries of OH variations we derive in the REF inversion is credible and consistent with the MCF observational record.

Before inversion, the prior emissions were pre-optimized to fit the global mean MCF mole fractions. It is not clear why the emission needs to be pre-optimized. Will the pre-optimization reduce the OH variation estimated by the inversions since the MCF emissions already fit the observations?

The study assumed a 50% error for MCF emissions. The MCF emissions become small after 1997. In Turner et al. (2017), the error for the MCF emission in the northern hemisphere is set to no less than 1.5 Gg/y. Will the assumption that prior MCF emission error proportional to emissions lead to underestimation of the prior error?

Firstly, we would like to correct an error in the manuscript. The assumed gridbox error is 50% over 1998-2005, but increases with 15% per year afterwards, up to a maximum of 200% after 2015. This is partly to reflect growing uncertainties in emission inventories and partly because emissions become exceedingly small in later years, as is indicated by the reviewer. Our minimum prior MCF emissions are around 2 Gg/year (e.g. Fig. 3), so the minimum emission uncertainty is 4 Gg/year, on the high end of what has been assumed in previous work. We have now outlined our prior settings correctly in the manuscript (Lines 120-123).

Initially, we ran inversions without pre-fitted emissions and assumed minimum emissions of 5 Gg/year. The result was a large overestimation of MCF mole fractions in later years. Therefore, many expensive iterations were required to even get close to observed mole fractions. We use a 3D transport model because we are interested in the subtle spatial gradients between surface sites, the temporal variations therein, and if and how these gradients inform on the separation between MCF emissions and OH. Pre-fitting the global MCF emissions to global mean MCF mole fractions turned out an efficient ways to reduce computational costs.

The impact of uncertain MCF emissions on derived OH variations is somewhat similar to that of the ocean sink discussed above: more likely to impact an OH trend than interannual variability of OH. However, the difference with the ocean sink is that we have found evidence that the inversion can spatially separate OH from MCF emissions to some degree. For example, in the REF inversion, we derive OH and MCF emission variations that are uncorrelated, signaling different degrees of freedom in the optimization.

The optimized OH is not well presented in the results. Only the interannual variations weighted by temperature were shown in Figure 1 and Figure 2. For the posterior OH, what are the global tropospheric mean OH concentrations and the corresponding CH4 lifetime? Are the latitudinal distributions consistent with previous studies? What is the N/S ratio? In the inversion TM5OH, the prior OH field shows higher concentrations over the northern hemisphere. Is the inversion using two different prior OH distribution estimated similar posterior latitudinal OH distributions? I think these values are also worth discussing.

We agree that there are many aspects of the derived OH distributions that are interesting. This is why we decided to provide the derived distributions in various formats in the Supplemental Data and urge any follow-up study to carefully consider which of these metrics best fits their purpose. We now provide the tropospheric lifetimes of MCF and CH_4 with respect to oxidation by OH in the manuscript and a

timeseries of these quantities in Supplement 1. We find that the derived lifetimes agree well with literature estimates.

The N/S ratio and global mean OH concentration are a reference quantities used often in previous literature. However, in the context of our study, we do not consider them informative and possibly misleading. We have found some adjustment to the N/S ratio and large adjustments in the global mean OH concentration, but this is a side-effect of the very large displacement of extra-tropical OH towards the tropics. This displacement of OH we deem unrealistic in amplitude and we at least partly attribute it to an erroneous ocean sink. Consequently, from the results presented in this study we are not confident that we can draw any conclusions regarding a N/S ratio or the global mean OH concentration and we are therefore hesitant to provide updated estimates.

We have changed Supplemental Figure S8, so that it now includes a comparison of our prior and posterior latitudinal distributions of OH with literature estimates, mostly from full-chemistry simulations. Here the difference between the TM5-OH and the Spivakovsky OH distributions can be clearly observed. Additionally, it is clear that the adjustments made to the prior distribution of OH are similar for the three different inverse set-ups in the ten-year inversions, although differences that were present in the prior OH distributions (e.g. the double peak in TM5-OH) are not adjusted.

The inversion results show small OH interannual variability and no significant OH trend. This is different from previous top-down studies using the two-box model inversion (Rigby et al., 2017; Turner et al., 2017). The author shows that the interannual variation can be supported by the negative correlation with the ENSO cycle. The large negative OH anomaly during 1998 (El Niño year) has been proven by several previous studies (Bousquet et al., 2006; Butler et al., 2005; Nguyen et al., 2020; Zhao et al., 2020). However, here the inversion shows a large positive anomaly in 1998. Besides, how to explain is a large positive anomaly in 2012?

Firstly, while we do not find evidence for a trend in OH, we also do not find evidence for the absence of a trend. This strongly relates to the uncertain ocean sink and MCF emissions, which we discuss above and in Lines 420-427 of the manuscript. Similarly, neither of the two-box model studies mentioned was able to exclude constant OH throughout their inversion period, so we consider our results to be consistent with these studies.

Top-down modeling studies of MCF that constrain OH have traditionally not provided strong evidence for physical mechanisms that could drive the derived OH variations. Here, we show that one strong driver of atmospheric variability (ENSO) correlates well with our derived OH variations. We consider this an improvement on previous work.

The positive OH anomaly in 1998 is likely due to spin-up effects. We mention this in the manuscript the first time we introduce global OH variations:

We have shown the entire twenty-year inversion period, which will include a spin-up and spin-downperiod of 1-2 years (indicated by the gray bars). For example, even though our initial MCF mole fraction fields are realistic, the strong positive oxidation anomaly in 1998 might be linked to errors in the initial field. Lines 182-184

We now also exclude 1998 and 2017 from our comparison with ENSO and indicate spin-up and -down periods in Figures 1, 2 and 8 as gray shaded areas.

We do not know what could drive the 2012 positive anomaly, but we also do not consider OH chemistry to be well-enough understood to exclude an anomaly of a few percent.

Other comments: L75: What is meant by "the most promising period in its measured history"?

Our message was that we investigate the period where MCF emissions are low, but MCF mole fractions (and thus loss to OH) are still relatively high. We have changed the phrasing:

The objective of this study is to investigate information on large-scale [OH] variations contained in measurements of the most promising tracer identified to date, MCF, during the period that follows its drop in emissions, using the most comprehensive tools available to us, in the form of a state-of-the-art inverse system built around a 3D transport model. L80

L110: Are the interannual variations of stratospheric photolysis considered over the inversion time period?

Interannual variations in the stratospheric photolysis rates are not considered. However, due to a changing stratospheric burden and distribution of MCF, and due to interannual variations in stratosphere-troposphere exchange, the absolute photolysis sink will vary. While in reality stratospheric photolysis rates can also vary interannually, we consider this effect second-order to other budget terms, similar to previous MCF inversions.

L118: The 10% error in latitudinal OH may be underestimated. Usually, the error for global annual mean OH is given by 10%. But for the monthly mean OH averaged for latitude, the error can be much larger (e.g. Naik et al., 2013; Zhao et al., 2019).

It is difficult to define a prior error on OH: it does not really represent the error on monthly latitudinal averages, because we have included strong temporal and spatial correlations. It is possible to argue that we should have chosen a larger error, but in the current set-up the inversion still finds adjustments to the prior OH distribution of up to 50%. This indicates that the prior and observational errors should be interpreted as balancing terms of the different components in the cost function, rather than as quantities that can be interpreted in an absolute sense.

Given the large amplitude of adjustments that we find, we consider that the cost function is well-balanced between observations and prior, or if it is unbalanced that we have attached too much weight to the observations (i.e. we are overfitting). Therefore, we do not think a larger error on OH would be beneficial to the inversion: if anything we would consider increasing the observational error.

L105-115: How you get the initial conditions for MCF? Are the initial conditions also optimized?

Initial fields were obtained from a 1988-1998 forward simulation that used initial fields, sources and sinks from the TransCom-CH4 protocol. As discussed in Patra et al. (2011), MCF fields in these simulations match observations quite well. The initial field was not optimized. Instead, the first 1.5 years of the inversion are considered as spin-up period during which optimized emissions and OH might correct for errors in the initial MCF distribution.

L128-L133: Are the MCF emissions also pre-optimized in POP and TM5OH?

Yes, and additionally we use the same global emission totals for all inversions. For example:

In our second inversion, referred to as POP, we redistributed the same annual total MCF emissions as in the REF inversion proportional to population density.. L134-145

L175: The results during the spin-up and spin-down period are not significant, I suggest the author remove the results of the corresponding period.

We have chosen to still include the spin-up and spin-down period in the revised manuscript. Instead, we have demarcated these periods as shaded areas in Fig. 1, 2, 8. The reason is that we do not know exactly how to quantify the length of these periods and so we consider it impossible to exclude a sufficiently long spin-up/down period without risking that we exclude potentially interesting information. We leave it now to the discretion of the reader to interpret and to use our timeseries correctly. We hope that in the revised manuscript it is clear that for example the positive 1998 OH anomaly is more likely to be a spin-up artifact than a real OH variation.

L196: The top-down estimated emissions and OH variations also depend on the variations of observed atmospheric MCF concentrations and the reaction rates (temperature) of MCF with OH. It is not clear for me why the small correlation coefficient between OH and MCF variations can reflect the OH and MCF are independently derived.

We can consider the extreme of a one-box model inversion of MCF that optimizes OH and emissions. In this set-up, OH and MCF emissions have opposite but indistinguishable effects on the modeled quantity: global mean MCF. Therefore, if simulated MCF is too low this can be compensated partly by a decrease in OH and partly by an increase in emissions. Depending on the error settings, the result will be a strong negative correlation between the posterior adjustments to emissions and OH.

In our 3D inversion, observational constraints from spatial MCF gradients are included and so there is some skill to separate OH from emissions, since their respective spatial distributions are very different. However, the surface network is sparse and since MCF has a long lifetime relative to most atmospheric transport timescales the distinction between emissions and OH is not sharp. Therefore, it is difficult to predict a priori if and how well this separation will work.

This is why we report the correlation between derived adjustments to emissions and OH. If the correlation would be strongly negative, then this is evidence that emissions and OH are used to correct for the same model-measurement differences, as expected in the one-box model. That this is not the case provides some evidence that the OH and emission adjustments address different aspects of the MCF gradients.

Figure 4: The MCF emissions are pre-optimized to reproduce the global mean MCF mole fractions, why there are still very large mismatches between the model simulated and observed MCF mixing ratios (dash line)?

For pre-optimization of emissions we needed to convert observed, global mean mole fractions to an atmospheric burden. For this we needed to define a global total atmospheric mass, which is a parametrized quantity, since emissions initially mainly spread through the troposphere and only later to the stratosphere. Apparently, we choose a somewhat too low atmospheric mass, resulting in too-low mole fractions in the REF and POP inversions.

We could have adjusted the emissions upwards, but we actually considered the poor match a good test of the inverse system. It shows that we can fit global mean mole fractions of MCF, when the prior mismatch is within reasonable bounds (which was not the case in the test inversion where we floored the emissions at 5 Gg/year; see above). Therefore, we decided that these pre-optimized emissions are a good starting point for the optimizations.

In Figure 3, Figure 4, and Figure 5, the line color corresponding to each inversion experiment is different. I suggest using the same color for each experiment in different figures.

We agree and have adjusted the colors to be consistent.

L254-L255: Is the bias in OH vertical distributions also contribute to the model-measurement mismatch?

This is a good point, we make mention of it now in the manuscript:

Estimates of the total atmospheric oxidizing capacity are more likely to be affected by a systematic underestimate of vertical MCF gradients. This underestimate could be driven by an underestimate of the vertical OH gradients or by too-fast vertical mixing in TM5. L277-279

L277: Why MEI should lead by one year? Is this mean that the OH should show a negative anomaly one year after the El Niño? This is not consistent with the explanation in L283-L286 and previous studies (Nguyen et al., 2020; Zhao et al., 2020), which show negative OH anomaly during El Niño years.

We agree that this lag was an unnecessary addition. We now only consider a zero-lag correlation for the 1999-2016 period (i.e. excluding one year spin-up and spin-down) and have slightly rephrased this section.

References

1. Bousquet et al. (2005), Two decades of OH variability as inferred by an inversion of atmospheric transport and chemistry of methyl chloroform, *Atmos. Chem. Phys., 5, 2635–2656*
2. Patra et al. (2011), TransCom model simulations of CH_4 and related species: linking transport, surface flux and chemical loss with CH_4 variability in the troposphere and lower stratosphere, *Atmos. Chem. Phys., 11, 12813–12837*