Reviewer #1 Dr. Julia Marshall

Comment [1-1]: This paper presents an analysis of the global methane budget and trend from 2010-2017 by simultaneously optimizing the source distributions, the OH sink (through hemispheric scaling factors), and linear trends using an analytical inversion approach with the GEOS-Chem model. Overall it is clearly written and structured and the figures are sufficiently clear and complete. From the subject matter it fits well within the scope of ACP.

At first glance this paper seems extremely similar in approach and content to Maasakkers et al. (2019) who used a very similar setup with the same model over an overlapping period (2010-2015) to do basically the same thing. The main difference that I can see is that here surface measurements are also included as a data constraint in order to show their complementarity (and consistency).

Response [1-1]: We thank Dr. Julia Marshall for the positive and valuable comments. All of them have been implemented in the revised manuscript.

As mentioned by the reviewer, the main improvement in our work relative to Maasakkers et al. (2019) is adding the in situ observations in the analytical inversion framework, comparing their ability and result with the satellite-based inversion, and quantifying the maximum information from the joint inversion. Such information is extremely important for a better understanding of the methane budgets and for the design of methane observing systems, yet it has not been addressed in previous studies to the best of our knowledge. In addition, our analytical inversion as done here implements a number of improvements to the Maasakkers et al. (2019) methodology, including in particular (1) separate optimization of subcontinental wetland emissions from other emission sectors to resolve their seasonal and interannual variability; (2) optimization of annual hemispheric OH concentrations rather than mean value of the period. Achieving these improvements increases the number of state vectors (and therefore computational costs) by 60%. We believe this work delivers sufficiently novel and important knowledge to the community.

Comment [1-2]: There's something a bit worrying showing up in Figure 6. Figure 6 seems to show that the both the in-situ-only and GOSAT-only inversions overestimate concentrations in the southern hemisphere and underestimate them in the northern hemisphere (more in the mid-latitudes in NH than in the Arctic). Interestingly, this consistent latitude-dependent bias does not seem to be present in the priors, or at least not as strongly. (Note that the 60-90N and 60-90S curves are more or less on top of each other when compared to the observations for the prior runs.) The fact that they then diverge so systematically after optimisation seems to imply that something is going wrong with the OH hemispheric optimisation **-** or is there another explanation?

Interestingly this pattern appears least distinct when considering the in-situ-only posterior sampled at GOSAT locations, whereas it is most pronounced in the GOSAT-only posterior. Can you explain this? Does this have something to do with the seasonal latitudinal coverage of the GOSAT measurements? In the comparison of the GOSAT-informed concentrations (both with and without the in-situ data) to the ObsPack measurements (panels 6c and 6d, less evident in 6b) there seems to be almost an temporal anti-correlation in the model-data mismatch between the 30-60N stations and the 60-90N stations.

It seems to represent a systematic error in the interhemispheric gradient, which can be explained through either the distribution of the sink, the distribution of the sources, or errors in the transport – or most likely a combination of all three. However as both the sink and the sources are being optimised, it seems surprising that such a zonally dependent offset is emerging. Even if there are transport errors (and there always are), I would expect a solution to emerge that was consistent with the interhemispheric gradient of the measurements. Of course the OH sink is only being optimised as a hemispheric scaling: might this reflect a problem in the spatial or temporal distribution that is being scaled? Still, usually the fluxes will adapt to compensate, provided they have sufficient flexibility. The fact that Zhang et al. (2018) found the inversion results to be not so sensitive to different OH fields suggests that this is not the case.

Some explanation of the source of this systematic error should be included. The only mention of transport errors is the claim that the regularisation factor gamma should help account for error correlations in the observations due to transport and source aggregation errors. Interestingly this does not seem to appear in the very similar simulations from the same group with a similar set-up, as seen in Figure 3d of Maasakkers et al. (2019).

Response [1-2]: Thank you very much for pointing this out. We figure out that the hemispheric bias as shown in the original Figure 6 is because the posterior hemispheric OH scaling factors were not correctly implemented in the posterior model simulation. We have corrected the implementation, rerun the posterior model simulation, and updated Figures 5-7 and Tables 1 and 3 in the text. As shown in the updated Figures 6b and 6g, the latitude-dependent bias between the observed and modeled methane concentration has been corrected for the ingested methane observations, indicating that there is no systematic error in the inversion. The updates do not influence the analyses or conclusions. We apologize for the confusion.

Figure 6f shows that the in-situ-only inversion biases low to GOSAT observations, and Figure 6c shows that the GOSAT-only inversion overestimates in-situ observations in the Southern Hemisphere while underestimates them in the Northern Hemisphere. These discrepancies, as already presented in the original texts and figure, do not reflect systematic error in the inversion, but rather provide insights on the consistency and complementarity between the two observations in the methane inversion, as analyzed in Section 3.5 and in [Response #1-3]. We have revised the text to clarify.

We now state in Section 3.1 "…The in-situ-only inversion effectively corrects this bias and its trend, and also significantly improves the correlations across all platforms. The GOSAT-only inversion performs comparably in correcting the 2010-2017 trend for the independent in-situ data (Fig.6c) and bias for background observations (e.g. aircraft observations in the Southern Hemisphere (Fig.S2)), but there is a low bias at northern mid-latitudes reflecting surface and tower data in North America and Europe. As we will see, the in situ observations are important for optimizing emissions in these regions.

… The GOSAT-only inversion corrects the bias and trend in the prior simulation at all latitudes. The in-situ-only inversion corrects the trends, but biases low to the GOSAT observations by about 10 ppbv with larger bias in the Southern Hemisphere due to the sparsity of in situ observation there. The comparison suggests that in situ and GOSAT observations are largely consistent for informing the global methane change but also have some complementarity for the inversion…."

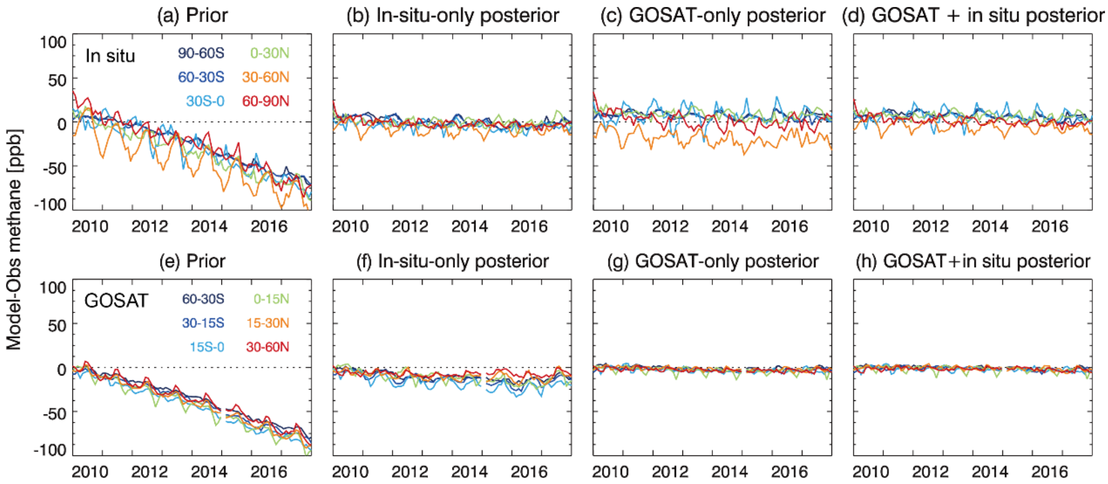

Figure 6. Ability of the inversions to fit the in situ methane observations and GOSAT satellite observations. Panels (a)-(d) show the monthly time series of the differences between observed and simulated in situ methane concentrations averaged over different latitude bands from 2010 to 2017. Panels (e)-(h) are the same as panels (a)-(d) but for GOSAT methane concentrations.

Comment [1-3]: Perhaps the most interesting (while also troubling) result is in Figure 13: the negative correlation between methane lifetime and estimated (anthropogenic) emissions is not in and of itself surprising. What is surprising is the fact that none of the three solutions are in any way consistent with each another. This can be explained by an underestimating of the posterior error covariances, as the authors do in L505-509. The fact that the GOSAT+in situ result does not lie somehow between the GOSAT-only and in-situ-only result is, however, worrying. The authors suggest that this is due to a correction of a bias in the GOSAT-only inversion by ingesting the in-situ measurements. This bias was diagnosed as being in both the OH (too low, because the methane in the SH was overestimated) and the fluxes (too low, because the methane in the NH mid-latitudes was underestimated). From this perspective it makes some sense that it would correct in the direction that it did, but why would it overshoot the in-situ-only solution? Is there some fundamental inconsistency in the two types of measurements (or an error in the model) that makes it impossible to match them both simultaneously?

This result seems to suggest that the measurements themselves are not really consistent with each other, which the paper claimed to set out to test (L91-94). Thus this result seems to contradict the conclusion that "the GOSAT and in situ data are generally consistent and can fit each other independently through our inversions" (L535-536). Even if the concentrations in the different inversion come closer to each other, is the result really consistent if the emissions and the lifetime are so very divergent?

Response [1-3]: The fact that the GOSAT+in situ result does not lie between the GOSAT-only and in-situ-only result (Fig.13) can be inferred from Figures 6c and 6f. Figure 6c suggests that both emissions and OH concentrations are too low in the GOSAT-only inversion, as the reviewer understands, while Figure 6f indicates either underestimation of emissions or overestimation of OH concentrations in the in-situ-only inversion, and the former one is more likely as GOSAT measurements used here are over land which should be more sensitive to emissions than OH loss. The GOSAT + in situ joint inversion thus has to enhance both the

methane emissions and OH concentrations compared to the In-situ-only and GOSAT-only inversions to correct these biases. We have revised the text accordingly in Section 3.5 to clarify this issue.

We agree with the reviewer that Figure 13 indicates that the measurements are not consistent with each other in optimizing the global methane budget, as stated in the original text (L505-506) "Comparison of the posterior PDFs between the GOSAT-only and In situ-only inversions implies that the two are inconsistent, since the 99% probability contour does not overlap (Fig.13)". We have removed "the GOSAT and in situ data are generally consistent and can fit each other independently through our inversions (L535-536)" which caused confusion. We have revised several places to clarify that the observations are consistent in correcting regional methane emissions in the inversion but are less consistent in terms of informing global methane budgets.

In the abstract, we now state "The in-situ-only and GOSAT-only inversions show consistent corrections to regional methane emissions but are less consistent in optimizing the global methane budget."

In Section 3.5, we now state "Comparison of the posterior PDFs between the GOSAT-only and In-situ-only inversions implies that the two are inconsistent in optimizing global methane budgets, since the 99% probability contours do not overlap (Fig.13a). ... Remarkably, the solution from the GOSAT + in situ joint inversion is more in agreement with in situ observations than GOSAT, and does not lie between these two solutions. Inspection of Figure 6c shows that the GOSAT-only inversion is biased low relative to in situ observations at northern mid-latitudes and biased high in the southern hemisphere, implying that both emissions and OH concentrations are too low. On the other hand, Figure 6f indicates either underestimation of emissions or overestimation of OH concentrations in the in-situ-only inversion, and the former one is more likely as GOSAT measurements used here are over land which should be more sensitive to emissions than OH loss. Ingestion of both observations in the GOSAT + in situ inversion thus enhances both the methane emissions and OH concentrations compared to the in-situ-only and GOSAT-only inversion to correct these biases. It also narrows the posterior error of mean anthropogenic emissions and methane lifetime against tropospheric OH by 20% and 50% compared to the GOSAT-only and in-situ-only inversions, respectively (Fig. 13a). Thus we find that the GOSAT and in situ observations are complementary in quantifying the global budget. "

In the conclusion, we now state "We find that the GOSAT-only inversion can generally fit the in situ data and the in-situ-only inversion can generally fit the GOSAT data, indicating consistency between the two data sets. However, …", "The GOSAT-only and in-situ-only inversions also show consistent corrections to regional methane emissions in the US, Europe, and China.", and "GOSAT and in situ observations have complementarity in constraining global emissions."

Comments [1-4]: While trying to understand this rather surprising result I realised that I would like to see some more figures: OH was scaled per hemisphere per year (16 state vector values). A time series of these scaling factors (perhaps as an additional panel or two in Figure 7?) would be interesting to see, rather than just an average lifetime over the whole period (similar to Figure 7d in Maasakkers et al. (2019)). This might also help convince me that scaling OH based on surface-based methane measurements alone makes sense - do the OH scaling factors in this case stay close to one throughout?

Another plot that might help convince the reader of the adequacy of the transport model and the improvement of the sources and sinks would be geographical (zonal + altitude?) plot of the model-data mismatch for aircraft data presented in Figure 5d. Even if it has to go into a supplement, it would be a useful piece of information for the reader to assess if this very surprising result might make sense.

Once these concerns are addressed, I think the paper would be appropriate for publication in ACP.

Response [1-4]: Thank you for the advice, we have added the two figures (Fig.7b and Fig.S2) and revised the text accordingly.

1) We present the posterior methane lifetime (as an indicator of OH scaling factors) in Figure 7b. We now state in Section 3.5 <u>"We also find that the in-situ-only inversion yields a larger interannual variability of posterior OH concentrations and thus methane lifetime than the GOSAT-only inversion (Fig.7b), due to the heterogeneous spatial and temporal distribution of the in situ observations."</u>.

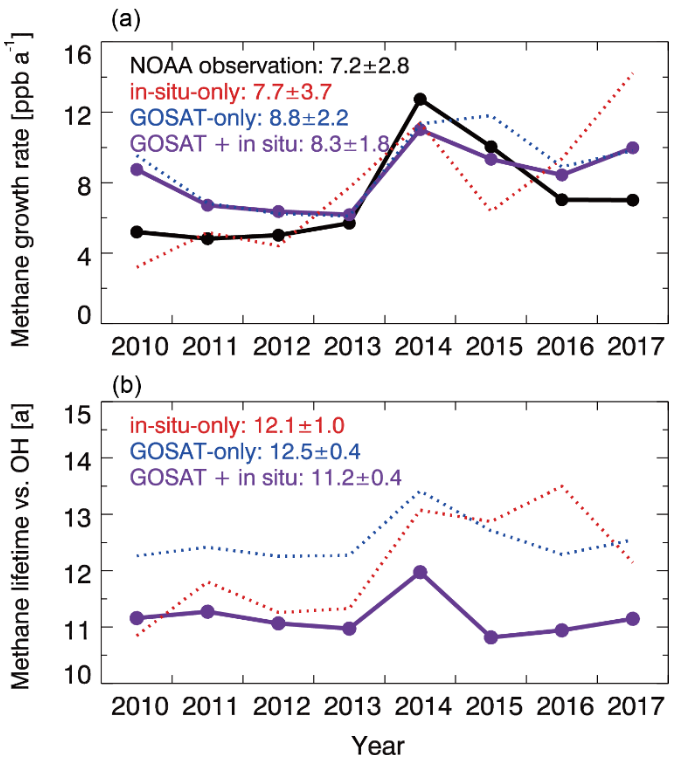

Figure 7. (a) Annual global growth rate of atmospheric methane, 2010-2017. Results from our three different inversions (In-situ-only, GOSAT-only, GOSAT + in situ) are compared to the observed growth rates inferred from the NOAA surface observational network (https://www.esrl.noaa.gov/gmd/ccgg/trends_ch4/, last access: 20 June, 2020). Mean annual growth rates and standard deviations from the different inversions are shown inset. (b). Methane lifetime against oxidation by tropospheric OH, 2010-2017, from the three different inversions. Mean lifetime and standard deviations are shown inset. The methane lifetime in the prior estimate is 10.6 years.

2) We present the model-observation bias for aircraft data for the prior and posterior

simulation in Fig.S2, and state in Section 3.1 <u>"The GOSAT-only inversion performs comparably in correcting the 2010-2017 trend for the independent in-situ data (Fig.6c) and bias for background observations (e.g. aircraft observations in the Southern Hemisphere (Fig.S2))"</u>

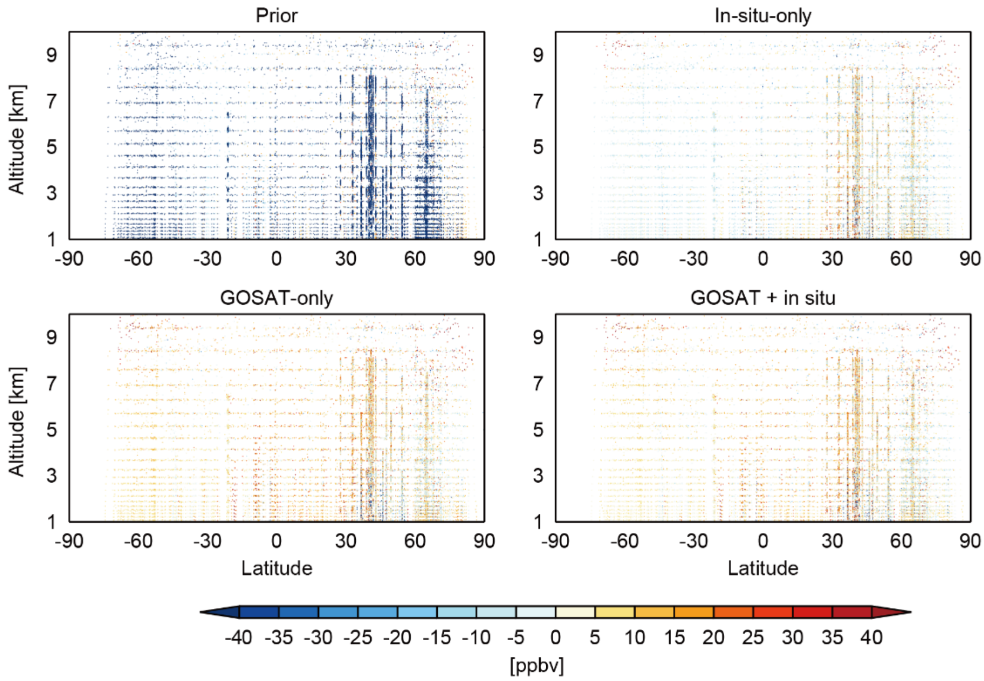

Figure S2. Differences between simulated and observed aircraft methane concentrations from the GLOBALVIEWplus ObsPack data product using GEOS-Chem with prior estimates and with posterior estimates from the in-situ-only, GOSAT-only, and GOSAT + in situ inversions.

Comments [1-5]: Minor comments: I would recommend adding how many independent pieces of information are contained in the GLOBALVIEW measurements alone to the abstract. This information is contained in the paper, but the way the numbers are presented in the abstract (which is as far as some readers get), it rather underplays the observation constraint brought about by the in-situ measurements alone.

Response [1-5]: We have revised accordingly in the abstract <u>"The in-situ-only and the GOSAT-only inversion alone, achieve respectively 113 and 212 independent pieces of information (DOFS) for quantifying mean 2010-2017 anthropogenic emissions on 1009 global model grid elements, and DOFS of 67 and 122 for 2010-2017 emission trends. The joint GOSAT + in situ inversion achieves DOFS of 262 and 161 respectively for mean emissions and trends. The in situ data thus increase the global information content from the GOSAT-only inversion by 20-30%."</u>

Comments [1-6]: One point that should be added into the discussion: When looking at the ability of a measurement system to assess long-term trends it is critical to consider the length of time over which these measurements are available. In this case, the surface-based network still has an advantage, and does not suffer from the same comparability issues that can arise when new sensors/sampling are introduced. This is mentioned briefly in lines 567-568, but they are first

mentioned as a method for satellite validation. Unless this measurements are being made across a profile (such as AirCore or aircraft), I cannot see how this could be the case.
Response [1-6]: We agree. We now rephrased in the Section 3.5 <u>"In situ observations will in any case continue to play a critical role for documenting long-term trends of methane with consistent calibration, …".</u>

Comments [1-7]: In line 475-476 you mention in passing that your optimisation approach can only solve for constant linear trends over the whole inversions period, which may not be appropriate for China. I wonder if it is really appropriate for other regions either? This is a clear drawback to the choice of state vector in your analytical inversion setup, and should be more clearly stated as such. If you want to test if this lack of trend is consistent with the findings of Sheng et al. (2019), showing an increase to 2012 and a decrease afterwards, perhaps you could perform the same inversion but broken up into two chunks: 2010-2012 and 2013-2017. Yes, this would require new transport simulations, but it would be interesting to check the robustness of the other trends as well. However this might be beyond the scope of the current study. (Perhaps something to add to the discussion?)
Response [1-7]: We agree, and indeed separating the inversions into two or more chunks will increase significantly the computational costs. We have clarified this limitation in Section 2.2: <u>"The inclusion of linear trends in state vectors allows us to identify the direction of emission change for each 4° ×5° grid in the 8-year period, but it would not capture high-frequency interannual variability."</u>

Comments [1-8]: I noticed that the panels labelled "China" and "Canada" in Figure 12 are identical. I suspect that they're both showing the results for Canada? In any case, this should be checked carefully and corrected.
Response [1-8]: Thanks for pointing it out. We had corrected the figure before it was posted on ACP Discussion.

Typographical/language remarks:
Comments [1-8]: Co-author Hartmut Boesch's last name is misspelled.
Response[1-8]: Corrected

Comments [1-9]:L127: with largest -> with the largest
Response[1-9]: Corrected

Comments [1-10]:L162: WETCHART -> WETCHARTS
Response[1-10]: Corrected

Comments [1-11]:L169: "full-chemistry" should not be hyphenated here (not a compound adjective before the noun)
Response[1-11]: Corrected

Comments [1-12]:L172: closed -> close
Response[1-12]: It has been rephrased.

Comments [1-13]:L218: challenged -> challenging
Response[1-13]: Corrected

Comments [1-14]:L225: Bayesian -> The Bayesian
Response[1-14]: Corrected

Comments [1-15]:L231: underestimate -> underestimation
Response[1-15]: Corrected

Comments [1-16]:L238: change -> changes
Response[1-16]: Corrected

Comments [1-17]:L266: be somewhat deviated -> deviate somewhat; overfit -> overfitting
Response[1-17]: It has been removed.

Comments [1-18]:L278: overfit -> overfitting
Response[1-18]: Corrected

Comments [1-19]:L284: Analytical solution -> The analytical solution
Response[1-19]: Corrected

Comments [1-20]:L288: I would suggest adding a colon after "analyses"
Response[1-20]: Corrected

Comments [1-21]:L290: capitalisation of "In situ-only" seems odd. Perhaps "in-situ-only" would be better as a compound adjective.
Response[1-21]: Corrected

Comments [1-22]:L339: year -> years
Response[1-22]: Corrected

Comments [1-23]:L345: by year -> by the year
Response[1-23]: Corrected

Comments [1-24]:L349: has insignificant -> has an insignificant
Response[1-24]: Corrected

Comments [1-25]:L364: higher information than in situ observations -> more information than do in situ observations
Response[1-25]: Corrected

Comments [1-26]:L375: I guess that ".," should just be ","?
Response[1-26]: Corrected

Comments [1-27]:L392: In situ observation is -> The in situ observations are
Response[1-27]: Corrected

Comments [1-28]:L418: Thompton -> Thompson
Response[1-28]: Corrected

Comments [1-29]:L453: US -> the US
Response[1-29]: Corrected

Comments [1-30]: Figure 11: I guess this percentage change is over the full period (rather than per year)? This should be clarified in the caption label. It also makes it a bit hard to compare to the text, where % trend per year is given. I assume that this is not a compounding percentage change, but rather the total percentage change divided by the number of years? In any case, this should be clarified.
Response [1-30]: Figure 11 shows the percentage change per year that derive directly from the inversions. We now state in the figure caption "Figure 11. Same as Figure 8 but for optimization of non-wetland (mainly anthropogenic) emission trends (% a^{-1}) in 2010-2017.".

Comments [1-31]:L501-502: This might seem like a small thing, but this is one of the most interesting findings of the paper, and as such should be perfectly clear. I would suggest the following change in phrasing: "are more effective than the satellite observations in independently constraining methane emissions from the sink by OH." -> "are more effective than the satellite observations in constraining methane emissions independently from the OH sink."
Response [1-31]: We have rephrased as suggested.

Comments [1-32]:L553: weak -> a weak
Response[1-32]: Corrected

Comments [1-33]:L560: remove "the"
Response[1-33]: Corrected

Comments [1-34]:L561: and methane lifetime -> and a methane lifetime
Response[1-34]: Corrected