Editor
Atmospheric Chemistry and Physics

25 March 2023

Dear Dr. Stiller,

Thank you for handing our manuscript. In this submission, we include substantial revisions to address the reviewer's concern. Below I summarize the main points and a point-wise response is also included.

First, we add a series of sensitivity inversions to test the robustness of our main findings. There are in total 4 sensitivity inversions configured differently from the main inversion, including (1) a seasonal inversion to address the reviewer's concern on "sub-annual variability" and (2) an inversion whose prior errors are specified for greater flexibility to adjust at locations with small prior methane emissions, which is to address the reviewer's concern on "strongly constrained sub-regional patterns". These sensitivity tests all find large discrepancies between the GOSAT and TROPOMI inversions in East China and northern India, indicating that our main finding is robust. The results are presented in Section 4.1 and Figure S7 and S8. Results of individual sensitivity inversions is also discussed wherever relevant.

Second, we develop and implement a new method to specify the $\mathbf{S_O}$ matrix, which does not involve the regularization parameter γ used previously. This is to address the reviewer's concern that the empirical procedure to choose the γ value leads to "the total information content for the whole region to be roughly the same". This new method should provide better interpretability than the traditional γ method. The method is described in the new Section 3.3 and its mathematical derivation in the Appendix A. Previous inversion result is included into the set of sensitivity tests.

Third, we investigate the reason for the discrepancy found over the Northern India in detail. We add more details to our explanation that this discrepancy is related to differences in data coverage between GOSAT and TROPOMI. Our additional analyses reveal that the discrepancy is also related to the large model-observation mismatch in the downwind Bangladesh region. In the absence of observations over the Gangetic Plain, the GOSAT inversion attributes this mismatch partly to northern India, while the TROPOMI inversion attributes it entirely to local Bangladesh emissions. This result is now added in Section 4.3.2.

Finally, we add more discussion on the retrieval errors, particularly for the CO_2 proxy retrieval (Figure S1, Section 2.1, and Section 4.3.1).

We hope that our revised manuscript will be suitable for publication in *Atmospheric Chemistry and Physics*. Thank you for your consideration!

Sincerely,

Yuzhong Zhang
zhangyuzhong@westlake.edu.cn
Westlake University
Hangzhou, Zhejiang, China

We thank the reviewer for constructive comments that help improve the manuscript. Below are our point-to-point responses shown in blue.

(1) without assessing the full error budget:

The discussion of the retrieval errors, in particular the imposed CO2-rescaling of proxy-CH4, remains short and qualitative (despite the comment raised in round 1). In their reply, the authors write that they do not assess these retrieval errors essentially because it is out-of-scope (although there is a section "4.3.1 Regional retrieval bias"). Since (parts of) the differences between the GOSAT and TROPOMI-inversions could originate from regional biases in the respective satellite data and not just differences in data density, I would argue that it is important to make an attempt to quantify such biases. For GOSAT proxy-CH4, this could be as easy as looking at the spread of models that go into the CO2-rescaling of the UoL algorithm (as suggested in round 1). For TROPOMI CH4, looking at the size of the (albedo-driven) bias correction could inform on which regions are more difficult than others.

We now show the modeled XCO_2 used by the UoL algorithm and their ranges from the three models in Figure S1. We add in Section 2.1 and Section 4.3.1 discussion on the uncertainty of the CO_2 proxy retrieval due to the specified XCO_2 field. We have shown the TROPOMI bias correction in Figure S2 and discussed its impact on the inversion in Section 4.3.1 and Figure S11 and S12.

As pointed out by the reviewer, the above information provides some hints about which regions are likely more difficult for retrievals. However, large discrepancy in emission estimation may emerge in regions that do not stand out in this type of retrieval error analysis, for example, East China (because there may be error sources that have not yet been identified). Our approach here is to empirically evaluate the retrieval and the corresponding emission estimates against independent observations, using the transport model as a platform for the comparison (Section 4.2).

For better clarity, we now modify the title of Section 4.3.1 to "Regional differences in XCH_4 retrievals", to make it clear that our focus is to explain the differences of inversion results rather than analyzing the sources of errors/biases for different retrievals.

(2) with imposing sub-annual variability:

In response to the comment on "any sub-annual temporal variability of fluxes" being imposed, the authors conducted a sensitivity run where they optimized for seasonal fluxes instead of the annual total. The reply contains a figure (without much explanations) that shows that the results for the seasonal and annual inversions differ substantially in magnitude and for some regions in the sign of the fluxes. Assuming that the figures show the annual fluxes for both inversions – which is the only thing that makes sense – I would argue that attributing the differences to "less observations" is not valid and that this needs further investigations.

We now clarify in the caption that the figure shows annual-averaged emissions from the seasonal inversion (Figure S7). We remove the statement that the smaller correction inferred from the seasonal inversion is due to less observations.

We now present the results from a series of sensitivity inversions including this seasonal inversion (Section 4.1, Figure S7 and S8). We show that our main findings on the regional emissions are robust against varied inversion configurations. Despite some differences in the posterior solutions, all these sensitivity inversions including this seasonal inversion find that the large discrepancy between the GOSAT and TROPOMI inversions occurs in East China and northern India, which is consistent with the main inversion.

We also examine whether the seasonal inversion better captures the temporal variability in observations from surface sites than does the annual inversion (Section 4.2 and Figure S9). We find the improvement is overall small.

(3) with strongly constraining the sub-regional patterns through a relative prior covariance matrix:
In response to the comment on using a "prior covariance in relative terms" (i.e. define as a percentag of the fluxes), the authors did a sensitivity run where they scaled the prior covariance to represent 100% instead of 50% flux errors. Obviously, this does not respond to the concern: if the covariance is defined in relative terms wrt. the fluxes (be it 50 or 100%), the spatial pattern is imposed because small fluxes will correspond to small variances. In other words, while calling the inversion "high resolution", the choice of the prior covariance prevents reshuffling of fluxes in the spatial domain. While this is a common choice, the paper needs to address this aspect (and point the reader to it) when discussing the sub-regional emission patterns (which are essentially just scalings of the imposed spatial prior patterns). One consequence is that flux areas that are not in the prior cannot be detected.

To address the reviewer's concern, we now add an additional sensitivity inversion, in which the prior error for emissions is specified as 50% of prior emissions or 1×10^{-10} kg m^{-2} s^{-1} whichever is larger (as a reference a methane flux of 1×10^{-10} kg m^{-2} s^{-1} ranks roughly 30th percentile in the prior inventory). This configuration allows the inversion to adjust more freely at locations with small or even no prior emissions. It is however more susceptible to noises in observations. Nevertheless, as discussed above, we find that our main findings of regional emissions are unchanged with this sensitivity test (Figure S7c and Figure S8).

(4) with constraining the total information content for the whole region to be roughly the same for the two datasets:
I find my concern still valid that, by design of the inverse method, the GOSAT and TROPOMI inversions will deliver roughly the same degrees of freedom for the entire domain (DOFs 70 and 71, Fig. 7). This is because the regularization parameter is determined by making the cost of the data and prior terms equal. While any operator is free to design such an inversion scheme, the resulting DOFs are not indicative of the actual information content of the datasets. Conclusions such as

L407: “However, over the entire East Asia domain, TROPOMI and GOSAT achieves almost the same (70) DOFS, with similar spatial patterns of averaging kernel sensitivities (Figure 7). Although the number of TROPOMI observations is much larger, strong error correlations in densely distributed data reduce the efficacy of individual observations, as shown by the difference in the regularization parameter determined for TROPOMI (γ=0.09) and GOSAT (γ=0.6) observations.”

are wrong since the total DOF is just limited by the selection method of the regularization parameter. I might be mistaken here, but the authors’ reply did not provide any explanations.

To address the issue raised by the reviewer, we now develop a new method to specify the $\mathbf{S_O}$ matrix, which does not require the γ parameter which is somewhat ambiguous and difficult to interpret. We now add Section 3.3. to describe the new method and an appendix to present the mathematical derivation. In this new method, we first fully specify non-diagonal elements of $\mathbf{S_O}$ by accounting for spatial and temporal correlations, and then find an approximation to $\mathbf{S_O^{-1}}$ that is computationally tractable (direct inverse of $\mathbf{S_O}$ is difficult to compute because of its large dimension). We now use this new method in the main inversion, which should provide better interpretability than the traditional γ method. The new method finds a higher total DOF from the TROPOMI inversion (74) than the GOSAT inversion (46). We also include a sensitivity inversion that use the old γ method for comparison (Figure S7d and S8).

The DOFS analysis is mainly used to explain the discrepancy found in the northern India between the GOSAT and TROPOMI inversions. We now add more details to this discussion (Section 4.3.2) by showing that differences in data coverage over northern India affects how the inversion attributes downwind methane column mismatches.

(5) Other comments

a) I am still puzzled by the small error bars e.g. on the order of 2-3% for the IND region (L234). The manuscripts states L235: “errors reported for regional estimates are 1-sigma standard deviations derived from posterior error covariance matrices”. I assume that equation (4) is used, i.e. the posterior error covariance matrix comprised of smoothing error and data error. The native dimensions of the posterior covariance is 604 x 604, since the inversion is run on 600 spatial clusters plus 4 boundary elements. The discussion, however, reports fluxes and error bars for aggregated regions. How is the aggregation of the covariances carried out? Are the covariances (off-diagonal elements) duely taken into account when aggregating?

   Yes, off-diagonal elements are taken into account. The variation of an aggregated quantity is computed as $\boldsymbol{w}^{\boldsymbol{T}}\hat{\mathbf{S}}\boldsymbol{w}$ where $\boldsymbol{w}$ is the aggregation vector such that the regional aggregation is given by $\boldsymbol{w}^{\boldsymbol{T}}\hat{\boldsymbol{x}}$. We now add a brief description of how regional results are computed in Section 3.2.

b) For the validation study, it would be good to have a figure showing the timeseries of validation and model data in addition to the statistics in the table. Given that the inversion method imposes sub-annual flux variability, I would expect quite poor agreement which might just not jump into the eye in the statistics. The poor R2 for some of the stations might be a hint.

We add the timeseries in Figure S9 following the suggestion. Figure S9 also include results from the seasonal inversions in addition to the main inversions. This shows that seasonal inversions only improve R^2 slightly at most sites, indicating that these poor R^2 is not mainly due to seasonal variations in emissions. The improvement is relatively larger at HF where the influence of rice emissions is strong. The relevant discussion is added to Section 4.2.