Response to Referee #1

We thank the reviewer for constructive comments that help improve the manuscript. Our responses to the comments are in blue.

The paper by Liang et al. compares regional inverse estimates of methane (CH_4) surface fluxes in East-Asia for the year 2019. The driver data are GOSAT and TROPOMI satellite observations of the column-average mole fractions XCH_4. Liang et al. describe the methodology based on the GEOS-CHEM transport model and a regularized inversion. They compare inversions of GOSAT and TROPOMI XCH_4 and find good agreement for some, substantial discrepancies for other regions. Comparisons to independent data sets serve as guidance to explain the regional differences.

Scope: I am a bit puzzled of what the overall goal of the study is. Is it a budget report of East Asian methane emissions or is it an evaluation of GOSAT and TROPOMI biases? The former would require more complete error analyses, more than a year of data, and more extensive discussions of previous work. For the latter, I would argue that the manuscript lacks completeness in terms of discussing error sources (see comment below). I recommend making the overall goal of the study clearer and revising the paper in the view of that goal.

We now clarify our objective in the manuscript (abstract, introduction, and conclusion). We focus on understanding the uncertainty in posterior methane fluxes arising from using different satellite data, which is one of many uncertainty sources for an inverse analysis. We do not intend to provide a comprehensive budget report for East Asia. Our analysis is also more than just evaluation of GOSAT and TROPOMI retrieval biases, as the discrepancy of posterior methane fluxes is related to not only retrieval biases but also other factors such as data coverage. We now state in the text that "*the main objective is to assess the consistency of methane fluxes inferred from the two sets of satellite data that differ in their data coverage and regional accuracy, adding information to the uncertainty characterization of satellite-based methane emission accounting.*". We emphasize that "*the analyses are conducted with identically configured inversions to isolate the effects of observation data*".

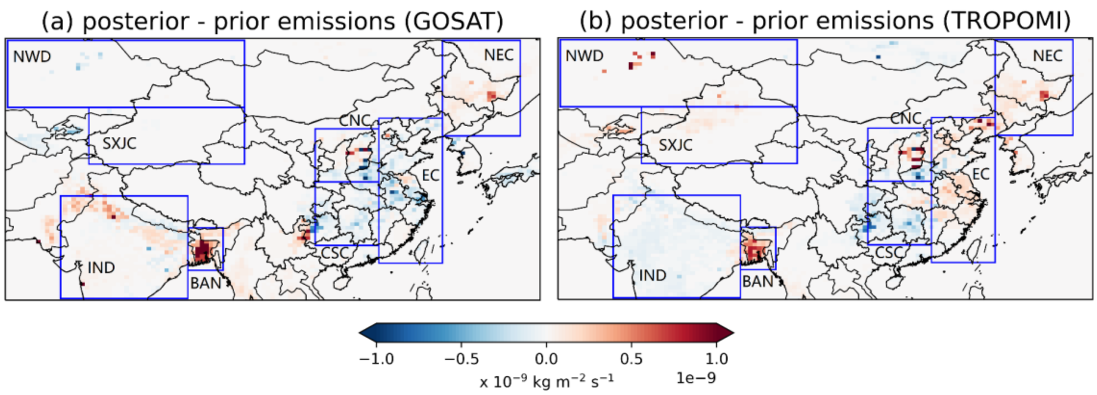
Proxy-CH4: Generally, the main (and, I believe, conceptually limiting) error source of the proxy method (GOSAT) must be discussed more thoroughly. It is the errors of the CO_2 fields that are used to construct XCH_4 from the raw CH_4/CO_2 ratio. Any (e.g. regionally correlated) errors in the prescribed CO_2 fields (typically taken from models) will map into respective errors in XCH_4. In fact, others [Schepers et al., JGR, 2012, https://doi.org/10.1029/2012JD017549] have compared proxy and full physics methods in the early days of the GOSAT mission. They found that, in a case study for India, erroneous CarbonTracker CO_2 fields caused biases in proxy XCH_4 data [Fig. 9 and 10 and related discussion in Schepers et al.]. The paper must examine and discuss this source of error to balance the discussion of scattering induced errors of the full-physics method (TROPOMI). To the best of my knowledge, the current version of the UoL proxy algorithm uses a model ensemble for CO_2-rescaling. One could try to estimate the error by looking at the spread of these (and potentially other) models in the investigated regions.

We add discussion on biases of proxy XCH_4 data in IND due to errors in the CO_2 field in Section 4.3.1. We cite Schepers et al. (2012) and Parker et al. (2015) in the discussion. We also mention broadly in the introduction section that the proxy method is subject to errors in specified CO_2 columns. We have not investigated the spread of CO_2 fields used in the UoL product as suggested, as our objective is not to assess error sources of a retrieval product. We aim to assess the impact of different retrievals on the inversed methane fluxes. Both retrieval biases and data coverage affect the inversion results. We show that retrieval biases between GOSAT and TROPOMI are relatively small over India, but they have large differences in data coverage and density.

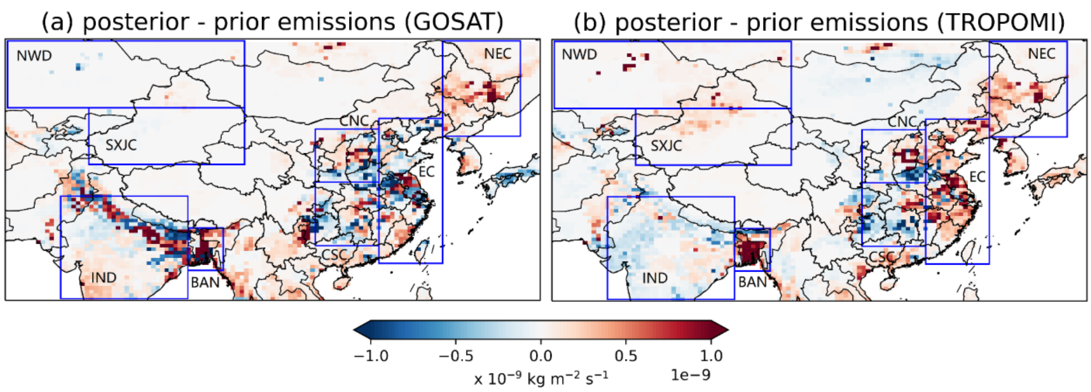
Setup of the inverse problem: I wonder about the setup of the inverse problem. If I get it right, the parameter vector contains 600 spatial elements which represent spatially distributed annual surface fluxes. I find this a mismatch of spatial and temporal scales. While the inversion is free to optimize a lot of spatial detail, any sub-annual temporal variability of fluxes is imposed. Given further, that the measurement vector contains daily XCH_4 data, I would argue that the temporal resolution of the inversion is at odds. The authors should discuss this aspect and provide sensitivity studies showing that their choice does not induce biases (e.g. by imposed seasonality).

Further, the authors have chosen to represent the prior covariance in relative terms (50%) with respect to the prior. This choice imposes that the spatial structure of posterior fluxes will be very similar to the one of the prior fluxes (simply because changing a small flux by 50% (or likewise) remains a small flux). This is clearly visible when comparing Fig. 2 and 4 (even though the log-scale in Fig. 2 needs some defiant eyeballing). The authors should clearly state the consequences of this assumption.

We perform additional sensitivity inversions to address the concerns raised by the reviewer. We first conduct a seasonal inversion, which shows similar correction pattern as the annual inversion (including the discrepancies between the GOSAT and TROPOMI inversions over East China and northern India, and the agreement over Northeast China, Central South China, and Bangladesh). The seasonal inversion in general infers smaller corrections than does the annual inversion (given the same error assumptions) because of less observations during a shorter period.



We then conduct an inversion with augmented prior errors of 100%, which again leads to similar correction patterns as the base inversion (agreement in NEC, CSC, BAN and disagreement in EC, IND, NWD, and SXJC). Results of this inversion tend to be noisier and generally have greater magnitudes because of less prior constraints, but the effect is overall small (by ~ 6 Tg a^{-1} over the whole domain).



These sensitivity inversions show that the agreements and discrepancies in posterior methane emissions between GOSAT and TROPOMI inversions are robust against perturbations of inversion setups. These results are now presented in Section 4.1 and supplementary information.

Inverse method: Equation 1 is the cost function of the inverse method. It is the classic regularization setup with a prior mismatch and a least squares measurement term where one term is scaled by a regularization parameter which the authors determine according to Figure S3. If I understand correctly, the condition on the selected regularization parameter is that the scaled least-squares term and the prior term impose equal cost. Why would one set such a condition when aiming at evaluating the information content of different data sets? In my understanding, this particular condition implies that whatever your measurement data are (be it dense or sparse, accurate or not), you force the inversion to deliver roughly the same degrees of freedom (for a given prior constraint). Figure 7 appears to confirm this conclusion: while GOSAT and TROPOMI have vastly different data density, the information content of the inversion is roughly the same. In consequence, the presented findings on degrees of freedom would not in any way represent the “natural” information content of the data but they are driven by design of the inverse method.

Generally, I would think that an L-curve method should work better for getting a regularization parameter that actually represents the information content of the data [see the cover (or chapter 4.6) of the book by Per Christian Hansen cited in the manuscript].

We now add more clarification in Section 3.2. The regularization γ is necessary here because error correlations are omitted in specified S_O (which is assumed to be diagonal for computational reasons). The extent of error correlations is different for different satellite data because of varied data density.

We have tried using L-curve method to determine a regularization parameter but found no apparent inflection points in the value range 0 ~ 1. We then determine γ following Lu et al. (2021) (section 2.4), which are also used by Qu et al. (2021) and Chen et al. (2022). The theoretical basis of the Lu et al. (2021) method is that if inversion results are consistent with specified errors ($\mathbf{S}_A$ and $\mathbf{S}_O$), $(\boldsymbol{x}-\boldsymbol{x}_{\mathbf{A}})^T\mathbf{S}_{\mathbf{A}}^{-1}(\boldsymbol{x}-\boldsymbol{x}_{\boldsymbol{A}})$ should follow a chi-square distribution with n degrees of freedom and $(\boldsymbol{y}-\mathbf{F}(\boldsymbol{x}))\mathbf{S}_{\mathbf{O}}^{-1}(\boldsymbol{y}-\mathbf{F}(\boldsymbol{x}))$ a chi-square distribution with m degrees of freedom.

Discussion: The posterior error bars of the satellite inversions (e.g. line 226f) are very small. I assume that they only represent the propagated measurement errors according to equation (4) (and line 185f) and that model transport errors, representativeness errors, more systematic measurement errors are neglected. When comparing the satellite-derived emissions to other studies (line 230ff), the reported error bars should be representative of the full error budget.

Our purpose here is not to compare with results from other studies or report a methane budget with comprehensively quantified uncertainties. We aim to assess the impact of different satellite observations on inversed methane fluxes. As we use identical inversion setups for the two inversions, systematic error sources mentioned by the reviewer would have similar effects on the two inversions, and therefore do not affect their comparisons. Errors derived from Eq. (4) is what we need to determine whether the difference in Fig. 3 is statistically significant, as it expresses random errors of the posterior estimates. We now clarify in the caption of Fig. 3.

References

Chen, Z., Jacob, D. J., Nesser, H., Sulprizio, M. P., Lorente, A., Varon, D. J., Lu, X., Shen, L., Qu, Z., Penn, E., and Yu, X.: Methane emissions from China: a high-resolution inversion of TROPOMI satellite observations, Atmos. Chem. Phys., 22, 10809-10826, https://doi.org/10.5194/acp-22-10809-2022, 2022.

Lu, X., Jacob, D. J., Zhang, Y., Maasakkers, J. D., Sulprizio, M. P., Shen, L., Qu, Z., Scarpelli, T. R., Nesser, H., Yantosca, R. M., Sheng, J., Andrews, A., Parker, R. J., Boesch, H., Bloom, A. A., and Ma, S.: Global methane budget and trend, 2010–2017: complementarity of inverse analyses using in situ

(GLOBALVIEWplus CH4 ObsPack) and satellite (GOSAT) observations, Atmos. Chem. Phys., 21, 4637-4657, https://doi.org/10.5194/acp-21-4637-2021, 2021.

Parker, R. J., Boesch, H., Byckling, K., Webb, A. J., Palmer, P. I., Feng, L., Bergamaschi, P., Chevallier, F., Notholt, J., Deutscher, N., Warneke, T., Hase, F., Sussmann, R., Kawakami, S., Kivi, R., Griffith, D. W. T., and Velazco, V.: Assessing 5 years of GOSAT Proxy XCH_4 data and associated uncertainties, Atmos. Meas. Tech., 8, 4785-4801, https://doi.org/10.5194/amt-8-4785-2015, 2015.

Qu, Z., Jacob, D. J., Shen, L., Lu, X., Zhang, Y., Scarpelli, T. R., Nesser, H., Sulprizio, M. P., Maasakkers, J. D., Bloom, A. A., Worden, J. R., Parker, R. J., and Delgado, A. L.: Global distribution of methane emissions: a comparative inverse analysis of observations from the TROPOMI and GOSAT satellite instruments, Atmos. Chem. Phys., 21, 14159-14175, https://doi.org/10.5194/acp-21-14159-2021, 2021.

Schepers, D., Guerlet, S., Butz, A., Landgraf, J., Frankenberg, C., Hasekamp, O., Blavier, J.-F., Deutscher, N. M., Griffith, D. W. T., Hase, F., Kyro, E., Morino, I., Sherlock, V., Sussmann, R., and Aben, I.: Methane retrievals from Greenhouse Gases Observing Satellite (GOSAT) shortwave infrared measurements: Performance comparison of proxy and physics retrieval algorithms, J. Geophys. Res. Atmos., 117, https://doi.org/10.1029/2012JD017549, 2012.

Response to Referee #2

We thank the reviewer for constructive comments that help improve the manuscript. Our responses to the comments are in blue.

Inverse methane emissions over East Asia are estimated and compared in this research for the year 2019 using data from two satellite observations, GOSAT and TROPOMI. Based on the GEOS-Chem transport model and analytical Bayesian inversion, Liang et al. developed the regional inversion framework. Comparisons at the regional level reveal consistency overall, but significant variation in some areas. The authors analyze the observations and independent measurements in further detail and argue that the variations can be explained by the data coverage and various retrieval techniques. Some arguments, however, need more analysis or are not convincing enough. Only after the authors address the following concerns can I recommend the paper to be published.

General comments

1, The overall goal of this paper is not very clear. Several key points are mixed up in this manuscript, but the authors do not break it down into individual points for this study. If the authors intend to assess the impact of the various satellite retrievals, I would suggest using one satellite but two retrieval products (for GOSAT: "proxy" v.s. "full-physics"; TROPOMI: official dataset v.s. WMFD). If the authors want to show the robustness of the inversion system they built, I suggest discussing error sources in detail and showing intermitted results.

We now clarify our objective in the manuscript (abstract, introduction, and conclusion). We now state in the text that "*the main objective is to assess the consistency of methane fluxes inferred from the two sets of satellite data that differ in their data coverage and regional accuracy, adding information to the uncertainty characterization of satellite-based methane emission accounting.*". Observations from different satellites are being applied in varied inversion studies to quantify regional emissions, but it is yet unclear if or to what extent different retrievals result in consistent methane emission estimates. It is often difficult to isolate the effect based on these studies because of confounding effects from differences in inversion configurations. We conduct the analyses with "*identically configured inversions to isolate the effects of observation data*".

We use the GOSAT UoL proxy retrieval and the TROPOMI full-physics retrieval in our evaluation. The assessment of inversion using these products are relevant because they have been widely used in inversion studies on regional and global scales. They are sufficiently different (different instruments, retrieval algorithms, data coverage, and regional accuracy), so our results bracket the uncertainty that may arise from using different satellite observation data. Evaluation of different retrievals from one satellite, as suggested by the reviewer, is also useful, but that will be a completely different study focusing more on retrieval algorithms.

To avoid misinterpretation, we now emphasize that "*there are other operational and science retrieval products available from both GOSAT and TROPOMI measurements*" and "*our analyses and conclusions are specific to the retrieval products used here*".

2, In both the abstract and discussion, the authors mention the large discrepancy over certain areas in East Asia is caused by retrievals. The cost imposed by the least-squares term and the prior term, however, appears to be equal. Thus, the changes in posterior emissions are mainly driven by the inverse system but not observations.

As expected from a Bayesian inversion, posterior emission estimates are a combination of observations and prior estimates and are affected by the inversion configurations. We isolate the effects of observation data on posterior emissions by applying different satellite data to identically configured inversions. We now clarify this point in various places throughout the manuscript. We also add results from sensitivity inversions with perturbed inversion configurations, which shows that the difference between results of the two inversions is robust against these perturbations. This is now shown in Section 4.1 and supplementary information.

3, The comparisons between simulations (with a priori and a posterior) and other independent measurements also indicate that increasing the emission intensity is ineffective to improve the result (see specific comment 6) in background areas. Does it imply that the model is unable to adequately capture the variations in these areas or that certain sources are missing from the same grid cell? The common problem in inverse modeling is the missing sources in a priori emission inventory. Again, if the authors aim at evaluating the emissions in China, please add more discussion on this aspect.

Table 1 shows that the inversion reduces biases substantially at background sites at WLG and PDI, which indicates better performance in the background area. The inversions are unable to improve performance for temporal variability (R^2) at the surface sites. This is largely due to the fact that optimization of methane emissions is done on an annual basis and other factors include model transport errors and observation representativeness. We now clarify in a footnote of Table 1.

Specific comments

1, Line 50-55: Please add more information about other methods to derive methane emissions as well as other satellites that are currently in service for methane monitoring (Sentinel-2, GHG-sat, etc.). The introduction here can be more comprehensive.

Sentinel-2 and GHGSat are point-source imagers that are not suitable for global and regional emission quantification. We describe and clarify in the introduction section that we are interested in area flux mappers such as GOSAT and TROPOMI.

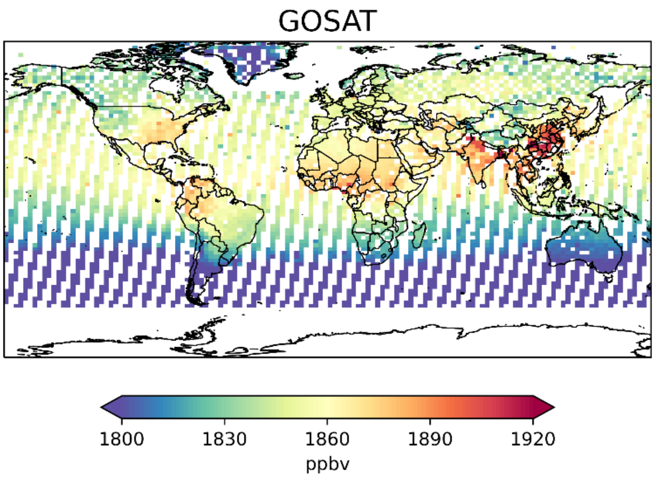
2, Line 95-100: As far as I know, the new version of TROPOMI has already been reprocessed. And they provide the data over the ocean (glint-mode). If the authors downloaded the official operational product, I strongly suggest using the reprocessing data. The operational product comes in a variety of versions, each of which contains various errors and biases that might cause inconsistency in error analysis. Please check/specify if the data in 2019 comes from the same version.

Additionally, retrieving data over the ocean (typically retrieved under sun-glint conditions) differs from retrieving data over land. It might cause discontinuity from land to ocean. Do authors check if there are any corresponding biases in GOSAT data?

All data in 2019 comes from the SRON S5P-RemoTeC scientific TROPOMI CH_4 dataset, version 14 (Lorente et al., 2021). We downloaded data from an ftp site: https://ftp.sron.nl/open-ccess-data-/TROPOMI/tropomi/ch4/14_14_Lorente_et_al_2020_AMTD/.

According to the SRON RemoTeC-S5P scientific XCH4 data product Product User Guide, the algorithm in the scientific product by Lorente et al. (2021) is implemented later in the official operational product of version 2.02.00. However, the published operational data are only available from July 2021 to November 2021, which does not cover the study period. The more recent TROPOMI data v 2.03.01 (over land and ocean) are only available from November 2021 to July 2022. The 2019 official operational product has not been reprocessed so far with the new algorithm. (Page 5 of S5P-MPC-SRON-PRF-CH4_v2.4.1_2.2_20220720 (copernicus.eu))

GOSAT glint-mode data are continuous from land to ocean, as shown below (presented on the 2° × 2.5° grid). A similar conclusion is reached by Qu et al. (2022) who used GOSAT observations over both land and ocean for a global study.

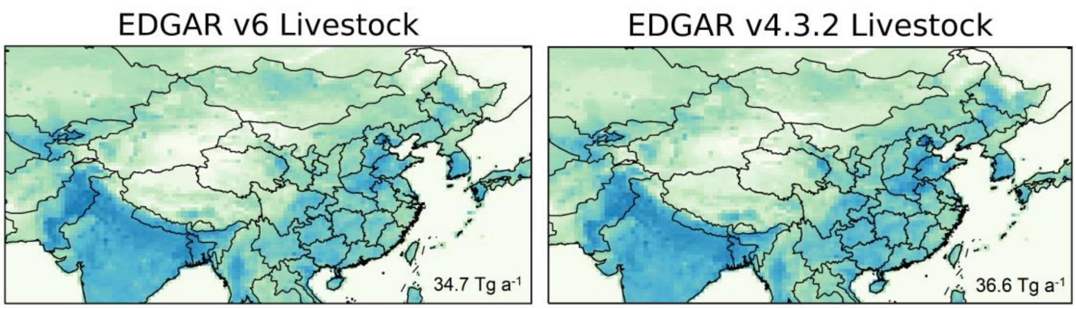


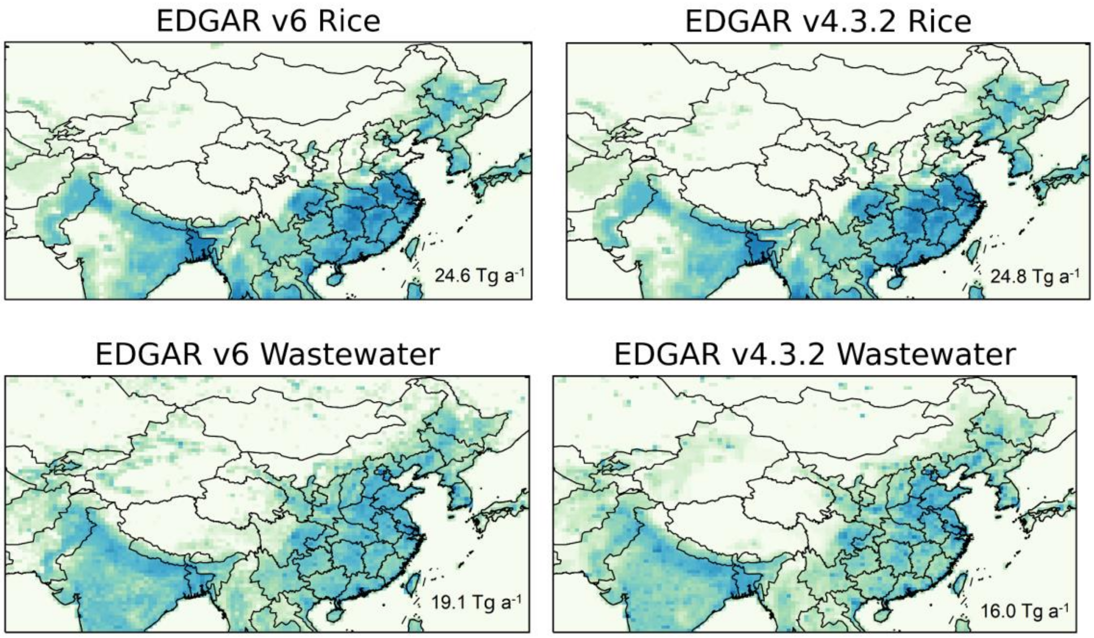
3, The diverging colormap of Figure 2 causes confusion. It is better to use a monotonically increasing colormap.

The figure has been updated following the suggestion.

4, Line 160. The anthropogenic emission from EDGAR v4.3.2 is relatively out of date, and which has also been found that the emissions have been overestimated in many areas. Why do authors not use the later version (latest: EDGAR v6)? At least, the authors should mention/estimate known biases in EDGAR v4.3.2.

We find the difference between different versions of EDGAR is insignificant in terms of the emission magnitude and spatial distribution over East Asia. We do not expect this newer bottom-up inventory will change the findings and conclusion of our study. We now present this comparison in Section 3.1 and a supplementary figure (see also below).





5, About Figure 4, either in GOSAT or TROPOMI inversion, the spatial differences show a strong spatial correlation between a priori and a posterior (a v.s. c and b v.s. d). Is it caused by the assumption of a priori covariance?

It is unclear to us what the reviewer meant here. The differences between the prior simulation and GOSAT observations are well corrected. The differences between the prior simulation and TROPOMI observations are also reduced but not fully corrected, reflecting constraints from prior estimations.

6, About Table 1, there are no improvements in the values of R^2. The low R^2 may imply the model lack repetitiveness in some places (considering they are background stations). Additionally, after being constrained by satellite measurements, the negative biases with the a priori inventory simply turn to positive biases, demonstrating that adjusting the emission intensity does not improve the outcomes of simulations.

The objective of Table 1 is to evaluate the discrepancies from the two inversions with independent observations, which is mainly shown by data at AMY, XH, and HF. Nevertheless, Table 1 does show substantial reduction in biases in background sites WLG and PDI by both inversions. The large negative bias in the prior simulation at WLG (the background site upwind East China where XH and HF are located) also indicates that small prior biases for XH and HF are due to the wrong reason (underestimated background concentrations). We now add clarification along with Table 1. We discuss in Section 4.3.3 reasons for poor performance at LLN.

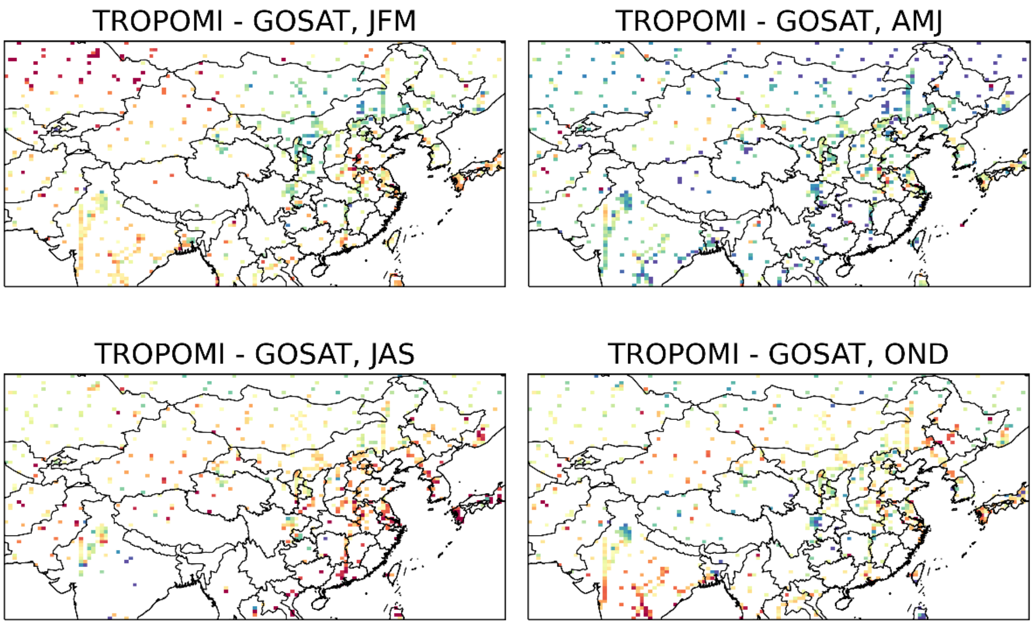
Many factors may limit improvement of R^2 for independent observations at a single station. The main reason is that this inversion also does not seek to optimize temporal variations in emissions. Other factors include model transport errors and observation representativeness. We now clarify in Table 1.

7, Line 310, section 4.3.1. Figure 3(a) show higher emission corrections than (b) while Figure 6(a) displays a small variation in concentration in IND. However, the variation of XCH_4 in EC demonstrates the consistency of Figures 3 and 6. Any explanations for this?

This is extensively discussed in Section 4.3.2. Differences in data coverage are also an important factor for the discrepancy in inferred emissions.

8, section 4.3.1. How do the sampling biases in different seasons and regions affect the comparisons between GOSAT and TROPOMI?

Figure 6b shows regional biases as a function of seasons. Spatial distributions of the comparison are also shown by season below. There are almost no coincident GOSAT and TROPOMI observations in Southeast Asia, South India, and Bangladesh during July and September. We assess the impact by performing additional inversions that optimizes emissions by season instead of annually. We find similar correction patterns as our main results. This result is now presented in Section 4.1 and in supplementary figures.



9, Line 390, Section 4.3.3. The authors argue that the lack of observations over the ocean leads to unrealistic enhancement of XCH_4. However, there are no sources over the ocean, the strong enhancement at the southeast corner is more likely caused by the model's erroneous processes of the transport. For example, the boundary condition of the regional model is updated by the output of the global model, which may contain bugs/errors.

The ocean data are useful for the inversion to reduce biases in the boundary conditions given by the global model. This mechanism is included in our inversion and is described in Section 3.2.

Reference

Lorente, A., Borsdorff, T., Butz, A., Hasekamp, O. P., aan de Brugh, J., Schneider, A., Wu, L., Hase, F., Kivi, R., Wunch, D., Pollard, D. F., Shiomi, K., Deutscher, N. M., Velazco, V. A., Roehl, C. M., Wennberg, P. O., Warneke, T., and Landgraf, J.: Methane retrieved from TROPOMI: improvement of the data product and validation of the first 2 years of measurements, Atmos. Meas. Tech., 14, 665-684, https://doi.org/10.5194/amt-14-665-2021, 2021.

Qu, Z., Jacob, D. J., Zhang, Y., Shen, L., Varon, D. J., Lu, X., Scarpelli, T., Bloom, A., Worden, J., and Parker, R. J.: Attribution of the 2020 surge in atmospheric methane by inverse analysis of GOSAT observations, Environmental Research Letters, 17, 094003, https://doi.org/10.1088/1748-9326/ac8754, 2022.