Reply to Review 2.

We thank the reviewer for the thoughtful comments. Reviewer comments are provided below in black with our responses in blue.

The authors utilized the latest version of TROPOMI XCH4 retrievals averaged in 2018-2019 to optimize various sources of CH4 on a global scale. In addition, since OH and CH4 are intertwined, they added OH to the state vectors for adjustment (optOH). Finally, the authors proposed a statistical downscaling method leveraging both prior knowledge from bottom-up emission inventories and the oversampled TROPOMI data to scale the optimized 2x2.5 degree emissions down to 0.1x0.1 degree. This method enabled them to identify the missing sources better. Their primary take-home messages are i) The use of XCH4 observations is not adequate to provide reliable constraints on OH; this is why authors gave up on the optOH result; ii) the middle east underreports the energy-sector CH4 emissions; iii) In South and Southasia, there is a strong degree of correlation between CH4 agriculture and waste emissions and some hydrological variables such as more precipitation (or to be more precise, the runoff) during Monsoon seasons; iv) the reported emissions related to the oil and gas industry over the US in the latest bottom-emission inventory (EDGAR v6) is not too different than the top-down estimates made from this study. In general, the paper has important implications for the CH4 budget and regulations. However, the results are too optimistic because careful error quantification is lacking. In addition, some key aspects of inversions need to be clarified. A major revision is required to bring this draft to a publishable level.

We have addressed the reviewer's comments regarding uncertainty quantification and clarification; please see detailed responses to the reviewer's more specific comments below.

Major comments:

Inversion: While I understand the importance of using an adjoint model for implicitly resolving the source-receptor relationship without having to rerun the forward model multiple times, the inversion framework comes with a significant weakness which is its inability to gauge the confidence level in the final estimates (i.e., the posterior error). The paper should inform the readers about this major weakness (introduction, conclusion, and Table 2) and highlights studies such as Qu et al. 2021, which reported the AKs of the top-down estimates because they used an analytical inversion.

Our prior OSSE analyses (Yu et al., 2021a) showed that posterior error reductions calculated as the reviewer describes for methane do not in fact correlate with more accurate flux estimates when transport errors and spatial emission errors are present (as is generally the case). We have now added this information to Section 2:

"Adjoint 4D-Var inversions do not directly provide posterior error estimates. Methods are available to indirectly derive such estimates (Bousserez et al., 2015; Yu et al., 2021a). However, our previous Observing System Simulation Experiments (OSSEs; Yu et al., 2021a) showed that for methane the computed posterior error reductions do not correlate with more accurate flux estimates in the presence of model transport errors and spatial emission biases."

Because of this, we instead use an alternative strategy that combines multiple inversion frameworks and a wide suite of independent observations (for CH_4, CO, and OH) to evaluate our solutions and their uncertainty. This is now clarified in Section 2.4:

"Here, we instead combine multiple inversion frameworks (Section 2.5) with an ensemble of independent observations (Sections 2.1, 3) to test the robustness of our results and characterize the associated uncertainties."

And in Section 2.1:

"We use a large suite of independent measurements to evaluate the inversions. These include methane columns from TCCON (2014), a global network of Fourier transform spectrometers, and methane mole fractions from ObsPack (near-real time version v2.0; (2021)), a global compilation of ground-based and airborne measurements. We further use CO and OH measurements from the Atmospheric Tomography (ATom) airborne campaign (Wofsy et al., 2018) to test inversion success at separately optimizing methane sources and sinks. ATom featured pole-to-pole profiling (0.2 to 12 km) during four seasons over four years. The flight design is thus well-suited to determine whether the optimized OH fields improve or degrade global model simulations of OH itself and of CO (whose dominant sink is reaction with OH). Measurements of CO during ATom were performed using the NOAA Picarro instrument with an estimated uncertainty of 3.6 ppb (Chen et al., 2013). OH measurements during ATom employed the Airborne Tropospheric Hydrogen Oxides Sensor (ATHOS), with an estimated uncertainty of 0.018 ppt (1-minute average; Brune et al., 2020)."

To further address this comment we have now added a new OSSE-based evaluation of the flux accuracy achieved with our 4D-Var inversion + spatial downscaling approach, as described in our subsequent replies below.

Finally, we now highlight the differences between this study and that of Qu et al. (2021) as suggested:

Section 2: "Recent inverse analyses by Qu et al. (2021) likewise examined the global methane budget using TROPOMI (and GOSAT) observations. Our study advances on that work in several ways. First, we optimize monthly rather than annual fluxes to identify seasonal patterns of variability. Second, in place of a traditional analytical optimization we combine 4D-Var with new inverse formalisms for better identification of missing sources. Third, we develop a new downscaling approach to constrain emissions at high resolution, and use this framework to elucidate flux mechanisms and missing sources. Finally, our analysis leverages an updated TROPOMI methane product (Lorente et al., 2021) that corrects an albedo-dependent bias present in the version used by Qu et al. (2021)."

For example, can we trust the optimized CH4 emissions using the TROPOMI XCH4 over water or high SZA? are the reported top-down estimates statistically significant?

For the first point, we do not use retrievals over water or with high SZA (> 70°) and this is stated in Section 2.1 ("We omit high-latitude (>60°) observations and require quality filter QA > 0.5 (Sentinel-5 Precursor/TROPOMI Level 2 Product User Manual: Methane, 2022) to avoid errors associated with high solar or viewing zenith angles, low surface albedo, excessive aerosol loading, clouds, terrain roughness, and measurement noise (Lorente et al., 2021)") and Section 2.2 ("… over oceans (which lack TROPOMI XCH_4 data).."). For the second point, as stated in-text we use the spread across diverse inversion frameworks to diagnose the level of confidence in the derived emissions. The associated uncertainty ranges are shown in Tables 2 and S2. Areas where derived emissions are not distinguishable from zero are also indicated visually as the shaded & hatched regions in the relevant figures (6, S12, S16).

In addition, several aspects of the inversion need to be further clarified: i) It needs to be explained how the 4D-var framework is applied when the inversion window is as wide as a 2-years average.

As requested we have now clarified the inversion time window and temporal resolution in Section 2:

"Our inversions run continuously from 01/2018 to 02/2020, optimizing monthly grid-total methane emissions and 26-month mean hemispheric OH concentrations. To minimize any effects from initial conditions and to allow sufficient observational constraints throughout the analysis period we focus interpretation on the 18-month period from 05/2018 to 10/2019. Annual values discussed later are for 11/2018–04/2019 plus the mean of 05–10/2018 and 05–10/2019."

ii) the TROPOMI full-physics algorithm relies on the prior profiles meaning the retrieved XCH4 is a piece of information on top of an ignorant model; I do not see any mention of if TROPOMI XCH4 was recalculated with GEOS-Chem prior profiles to ensure that only the true information from the satellite radiance is used for the inversion (Page 39 in http://www.tropomi.eu/sites/default/files/files/publicSentinel-5P-TROPOMI-ATBD-Methane-retrieval.pdf). This task should be done iteratively because the GEOS-Chem profiles change after each inversion iteration.

Yes, this is stated in Section 2.2 and we now clarify that this is done at each inversion iteration:

"For all model-satellite comparisons (and at each inversion iteration) the GEOS-Chem output is sampled according to the TROPOMI observation operator at the overpass time and location."

iii) it is unclear if the TROPOMI data have been scaled up to the resolution of GEOS-Chem in the inversion; if not, the difference in their spatial representativity will result in a perceived bias which can be problematic.

We have now clarified this in the text. "For inversions on the 2° × 2.5° model grid, we first average the TROPOMI observations to this resolution."

iv) how do the errors associated with the vertical diffusion in GEOS-Chem impact your result? The model error parameter is lacking in the analysis.

As described in Section 2.4, our observing system error estimates are computed from the residual standard deviation between observations and prior simulations. This approach implicitly accounts for non-systematic measurement errors and model transport errors , and averages 11 ppb in our case. In separate ongoing work, we performed model simulations with an alternative vertical mixing scheme, and this leads to discrepancies that average 4.26 ppb. Explicitly accounting for such an effect would not appreciably change the error estimates employed here (e.g., (11^2 + 4.26^2)^0.5 = 11.8).

v) is the state vector the total CH4 emissions, or is it sector-based?

We optimize total emissions rather than on a sectoral level. We have now clarified this in multiple locations.

vi) why did not the authors use the glint mode to account for off-shore emissions?

Coverage locations for the glint mode off-shore observations vary from day to day and developing a framework to employ the sunglint data was beyond our scope.

Downscaling: Two central problems exist: i) can the two-year average TROPOMI XCH4 truly capture the spatial variance in XCH4 at 0.1x0.1 degrees? By oversampling TROPOMI pixels, we may lose spatial variance (a smoothing effect) more than we reduce the random noises. The authors need to prove that a two-year averaged TROPOMI data can resolve the length scales of plumes at the resolution of 0.1x0.1 degree; if not, that resulting spatial representativity error induced by oversampling can potentially hinder reaching a 0.1x0.1 degree information.

First, the TROPOMI nadir footprint is ~7 km, substantially smaller than the 0.1 degree grid size (~11 km) that we average to. Second, previous work has already demonstrated the ability of TROPOMI to resolve sources at far finer than 0.1 degrees. For example, Maasakkers et al. (2022, doi.org/10.1126/sciadv.abn9683) used other high-resolution satellites (GHGSat) and WRF simulations (at 3 km) to identify methane point sources, and showed that the TROPOMI methane measurements oversampled to 0.01 degrees (i.e., 10 times finer than employed in our work) can resolve these sources. The figure below, reproduced from Maasakkers et al. (2022) clearly demonstrates the TROPOMI ability to resolve features that are significantly smaller than 0.1 degrees. In our work we applied similar gridding methods but employ a coarser resolution of 0.1 degrees. We have added a citation of Maasakkers et al. (2022) to our paper.

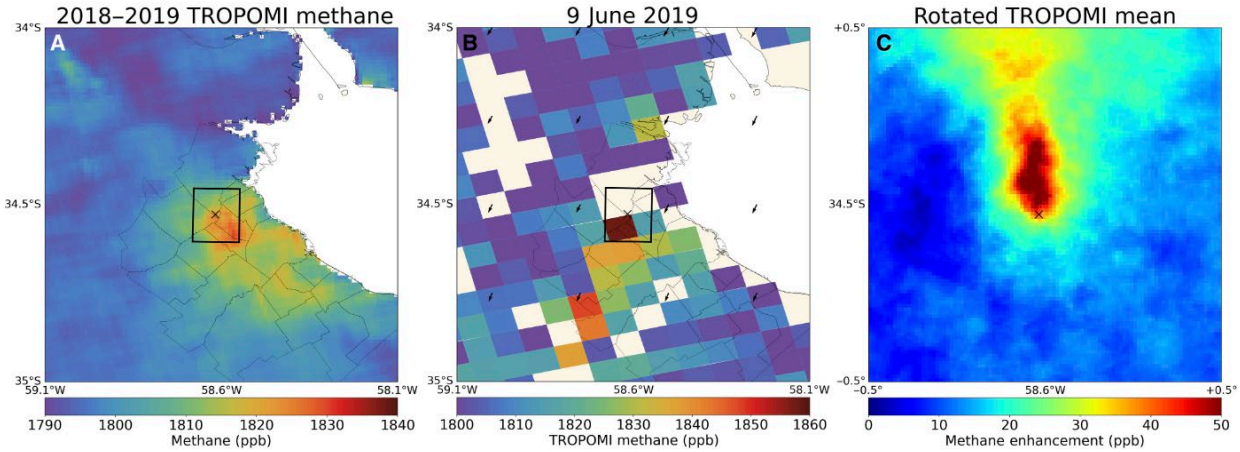

Figure reproduced from Maasakkers et al. (2022, doi.org/10.1126/sciadv.abn9683). TROPOMI observations over Buenos Aires (Argentina). (A) Mean 2018–2019 TROPOMI methane concentrations oversampled (i.e., accounting for the full footprint of the observation) on a 0.01° grid. The Norte III landfill is indicated by the black cross; also shown are a GHGSat window centered on the TROPOMI-derived target (thick lines) and the Greater Buenos Aires municipalities [thin lines]. (B) A single TROPOMI overpass on 9 June 2019 exhibiting a methane plume downwind of Buenos Aires with wind arrows representing ERA5 10-m winds. (C) The 2018–2019 wind-rotated average giving a clear (north-oriented) plume signal indicating a concentrated source.

ii) The proposed downscaling method (Eq2) heavily relies on assumptions about prior errors/information. Even the observational term depends on the prior fraction of emissions (fk). As noted by the authors, the prior CH4 emissions do not agree with other bottom-up emission inventories

(R2=0.01?), so how can one fully trust a downscaling output when it heavily relies on questionable prior information?

The key point here is that for our approach incorrect prior emissions are acceptable provided that the errors are appropriate: when uncertainties are high, the observational term get weighted more heavily. Our spatial downscaling uses an error-based weighting term to balance between the prior information and the long-term TROPOMI data:

"As shown in Figure S7, the resulting downscaling relies most frequently on the prior information, particularly for low-emission areas, but hotspots and locations with higher prior uncertainties are preferentially informed by the observations."

We have now added a caveat regarding the reliance on robust prior errors, as suggested:

"The method thus assumes robust prior error estimates, a caveat that also applies to Cusworth (2021) and similar methods."

Third, the fk term does not impose any influence from the prior spatial distribution as the full $\sum_{k\in j}\left(f_k\left(y_{2y,k}-\ y_{bg,i}\right)\right)$ term in the denominator is simply a scalar that ensures conservation of the derived 2°x2.5° emissions.

Would it be more sensible to use the posterior error/distribution for this part (under the condition in which the inversion framework was analytical, permitting the calculation of the posterior error)?

As described in our earlier replies, our prior OSSE work (Yu et al., 2021a) has shown that larger error reduction does not in fact correlate with more accurate flux estimates.

As a result of these two combined complications, I challenge the authors to provide an error estimation for this downscaling method and propagate them to the emissions maps and statistics (especially Table 2). Your study did not inform the posterior errors due to the use of adjoint; now, the lack of an uncertainty estimation for the downscaling part (which is the most crucial selling point of the paper) appears as an oversight.

We have already addressed the comment about posterior errors and the issues with their use. The paper also already includes a dedicated OSSE analysis demonstrating the applicability and robustness of the downscaling approach presented here. To further address this point we now also include a downscaling error assessment based on the flux bias reduction achieved in the OSSE. This has been added to the text as follows:

"Figure S8 shows that the downscaled OSSE solution reduces the prior bias by 17%–56% for sources exceeding 1000 kg CH_4/box/day (accounting for 99% of the domain-wide emissions) when not subject to transport error. In the presence of transport error, the downscaling method has limited success for the very largest sources (>2×10^5 kg/box/day), but nevertheless exhibits strong bias reduction (21%–50%) for sources between 1×10^3–2×10^5 kg/box/day (96% of domain-wide emissions)."

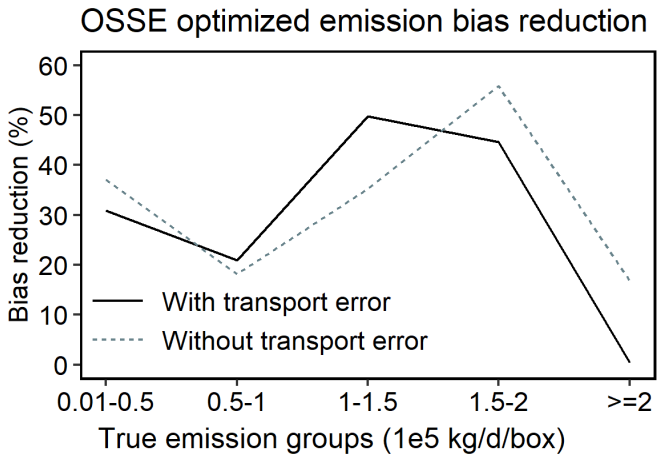

Figure S8. Downscaling bias reduction as a function of emission magnitude, based on 1-month Observing System Simulation Experiments (OSSE) over North America (see main text for details).

Comparison to ATOM: I need clarification on this comparison. Based on the author's discussion on optOH, they suggested that a strong El Nino year (2018-2019) led to lower-than-average OH mixing ratio (8.27e5 molec/cm3). But then they compared their constrained model in different years (<2018) with ATHOS OH measurements and concluded that their optOH is vastly underestimated, reinforced by the overestimation of CO. If 2018-2019 was a unique timeframe, how could one generalize the comparison results from other years to the 2018-2019 period?

First, the model is sampled at the time of measurements in all cases, so that each comparison employs the same timeframe between model and observations. Second, the ATom comparisons span 2 years and show consistent patterns. Third, we also evaluated the inverse solutions using methane measurements for the inversion timeframe from TCCON and ObsPack. We arrive at consistent conclusions in all cases, lending support to our interpretation.

Furthermore, I see a few issues here i) ATHOS OH can easily contain up to 30% error; have the authors considered the measurement errors in their comparison? Given the observational errors, I encourage applying a statistical test to know if the differences are real.

First, we have added new information to Section 2 describing the ATom measurements and their uncertainties:

"We further use CO and OH measurements from the Atmospheric Tomography (ATom) airborne campaign (Wofsy et al., 2018) to test inversion success at separately optimizing methane sources and sinks. ATom featured pole-to-pole profiling (0.2 to 12 km) during four seasons over four years. The flight design is thus well-suited to determine whether the optimized OH fields improve or degrade global model simulations of OH itself and of CO (whose dominant sink is reaction with OH). Measurements of CO during ATom were performed using the NOAA Picarro instrument with an estimated uncertainty of 3.6 ppb (Chen et al., 2013). OH measurements during ATom employed the Airborne Tropospheric Hydrogen Oxides Sensor (ATHOS), with an estimated uncertainty of 0.018 ppt (1-minute average; Brune et al., 2020)."

Second, we now include additional statistical tests and interpret the model-measurement differences in the context of measurement uncertainty:

"With the exception of ATom 3, the mean model OH biases with respect to ATom observations are ~80% lower for fixOH than for optOH (mean differences are all significant based on a paired t-test at 95% confidence). These optOH results exhibit a consistent OH underestimate (averaging 0.020–0.044 ppt) that exceeds the 0.018 ppt measurement uncertainty. Biases in the simulated background CO levels are likewise lower (by 7–87%) in the fixOH simulations, with a clear CO overestimate for optOH (Figure S9). Again, the mean fixOH/optOH differences are all statistically significant at the 95% confidence level, with model-measurement discrepancies for optOH (7–12 ppb) exceeding the 3.6 ppb measurement uncertainty."

ii) have the authors looked into the measured OHR to see if there are missing sources (such as VOC) in their model?

This is out of scope as our model runs employ tagged methane and tagged CO simulations, which are offline and do not simulate other VOCs.

What is the implication of underestimating OH in the optOH scenario? Should we perform a multispecies inversion with TROPOMI CO and HCHO to provide an additional constraint on OH?

Our subsequent work is indeed pursuing analyses along these lines. We made the reviewer's point about the potential value of multi-species inversion at the end of the conclusions:

"We address the issue in this work by employing validated OH fields from a chemical transport model, but future methane inversions can benefit from incorporating additional datasets (e.g., CO, methyl chloroform, formaldehyde) as constraints on the methane sink (McNorton et al., 2016; Rigby et al., 2017; Turner et al., 2017; Wolfe et al., 2019)."

In theory, letting OH see the CH4 feedback (optOH) is suitable, but why should the ATOM analysis discourage this meaningful practice? What if your default CO simulations are too high, and the optOH highlights that tendency? I'm left with many questions because the authors needed to dig into the problem more deeply.

We arrived at this conclusion not just based on the ATom analysis but rather based on multiple consistent lines of evidence. First, independent evaluation of the posterior simulations against ObsPack observations reveals an optOH overestimate of methane concentrations, which would be consistent with an OH underestimate. Second, the ATom OH observations also point to an OptOH underestimate of OH. Third, the ATom CO observations reveal an optOH CO overestimate, which is also consistent with too-low OH. We agree with the reviewer that other factors can affect model-measurement agreement for CO but considered together these multiple consistent lines of evidence provide robust support for our interpretation as presented in the paper.

Specific comments:

P1. L16. I do not think you ever used CO as an observational constraint for the model. This sentence is misleading.

We have revised this text as suggested:

"Employing remote carbon monoxide (CO) and hydroxyl radical (OH) observations with independent methane measurements for evaluation, we infer from TROPOMI a global methane source of 587 Tg/y and sink of 571 Tg/y for our analysis period."

P2. L38. Does really the recent enhancement in CH4 need to be better understood? I suggest adding more recent studies discussing the role of reduced NOx due to the lockdown on OH and CH4. There must be a recent study from Jacob's group regarding the increases in wetlands and permafrost CH4. This part needs more references in general.

As suggested we have now added additional citations of recent studies by Stevenson et al. (2021) and Peng et al. (2022) discussing NOx emissions, wetland emissions, and their contributions to the recent methane increase.

P2. L40. After reading the abstract saying that sinks and sources cannot be resolved with a high-resolution satellite, I found this sentence regarding transformative advancement somewhat contradictory.

We have now revised this text:

"Employing remote carbon monoxide (CO) and hydroxyl radical (OH) observations with independent methane measurements for evaluation, we infer from TROPOMI a global methane source of 587 Tg/y and sink of 571 Tg/y for our analysis period."

P2. L43. Shouldn't we also have an overrepresented source? If a source is underrepresented, another source should compensate for it.

We have now revised this text: "missing and unexpected sources".

P2. I found the second paragraph of the introduction imbalanced. The paper utilized remote sensing data, so I highly suggest comparing the pros and cons of using different remote sensing observations.

We thank the reviewers for the suggestion. We prefer to keep this paragraph as-is, and compare our results with previous GOSAT-based inversions in the discussion section.

P2. L60. Who came up with this R2 value? The agreement is unsettlingly low. Please provide a reference.

We calculated this R^2 value directly from the cited inventories.

P2. The third paragraph needs to include the temporal representation error between different emission inventories and the fact that CH4 emitters can vary from time to time at a relatively short temporal scale.

We have revised this text as suggested:

"Meanwhile, current inventories also lack the ability to predict emission sporadicity (e.g., Irakulis-Loitxate et al., 2022; Pandey et al., 2019), while temporal representation errors can also arise between inventories due to time lags associated with their development."

P3. L71. In the abstract, you said you had constrained CO, but here you imply that they will be used for evaluation.

We have now revised the abstract to clarify this point.

P5. L147. Why is the regularization factor applied to So instead of Se? We are less confident in Se compared to So. Another way (which should be the same as finding the maximum curvature in the L-curve) to find the optimum regularization factor is to scale Se several times and find the knee point in averaging kernels vs. the factor, although this might not be possible with the adjoint.

We apply the regularization factor to So following many previous studies (e.g., Jacob, et al., 2022, https://doi.org/10.5194/acp-22-9617-2022). We thank the reviewer for pointing out the knee point method.

P6. L167. How sure are you that the muted response of the model to a higher error in OH is not due to the lack of the degree of freedom? An analytical inversion would be able to answer it.

This is indeed one of our main conclusions: "evaluation of the inverse solutions indicates that methane sources and sinks cannot be simultaneously resolved by methane observations alone—even with the dense TROPOMI sampling coverage". In other words, the lack of degrees of freedom prevents robust simultaneous optimization of both sources and sinks.

Please see our earlier replies regarding analytical versus adjoint methodologies.

P6. L177. What is the implication of this low gamma value? The prior error is too uncertain or the observations are less noisy compared to the So?

The gamma value of 0.1 reflects the large number of observations compared to the size of the state vector.

P7. L213. Why should we accept this ad-hoc definition as XCH4 background? Any concerete evidence?

This XCH_4 background definition was determined via line search of the OSSE output. Specifically, we systematically varied the XCH_4 background parameters over a wide range and selected those that resulted in the best downscaling performance (i.e., giving the strongest consistency with the true local scale emissions). We have now revised the text to clarify this:

"This background definition was determined via OSSE analysis (described below). Specifically, the corresponding parameters were varied systematically over a wide range to identify values yielding the best consistency with the true underlying fine-scale emissions."

P9. Section 3.1. How can the errors in the soil uptake by methanotroph influence these results?

Soil uptake is only estimated to account for ~6% of the total methane sink, whereas oxidation by OH accounts for ~90% (Saunois et al., 2020). While uncertainties in other sinks can also exist, we therefore focus on OH here as it is the dominant sink and exerts the largest influence on methane source inversions. We have revised the text to clarify this point, as follows:

"Other minor sinks, such as soil uptake, are also uncertain but not addressed here."

P10. In the first paragraph, I encourage using absolute numbers from Figure 3 to describe the reduction in bias, such as (from -13.8 to 8.8 ppbv).

We have revised this text as suggested:

"Ground-based methane column (XCH_4) observations from the TCCON network (GGG2014 (2014)) show comparable improvements over the prior for both the fixOH and optOH solutions (and for their individual member inversions), with 71% (from -12.9 ppb to 3.8 ppb) and 66% (to 4.3 ppb) mean bias reductions, respectively (Figure 3, Table S1). However, global in-situ measurements from ObsPack (near-real time version v2.0; (2021)) reveal a 93% (from -13.8 ppb to -0.9 ppb) absolute mean bias improvement for the fixOH framework compared to just 39% (to 8.4 ppb) for optOH (Figure 3, Table S1). Figure 3 further shows that the optOH solutions overcorrect the prior negative bias with respect to ObsPack, providing a first piece of evidence for a methane sink underestimate in this inversion."

P10. L299-301. I'm afraid I have to disagree with saying that TROPOMI is dense, but we cannot fully resolve the source/sink of CH4. It would help if you had more than CH4 to get OH right (such as HCHO and CO constraints), which has nothing to do with the densityTROPOMI XCH4 observations.

Indeed, we agree with the reviewer and have addressed this point in our conclusions section:

"We address the issue in this work by employing validated OH fields from a chemical transport model, but future methane inversions can benefit from incorporating additional datasets (e.g., CO, methyl chloroform, formaldehyde) as constraints on the methane sink (McNorton et al., 2016; Rigby et al., 2017; Turner et al., 2017; Wolfe et al., 2019)."

P10. In the second paragraph, you should specifically mention what emissions are used for the retrospective simulations.

This information has now been added to Section 2.3:

"Simulations to evaluate posterior model performance for CO and OH employ anthropogenic emissions (for CO, NOx, and VOCs) from the Community Emissions Data System (Hoesly et al., 2018), the 2016 EPA NEI v1 (NEIC, 2019), and the Air Pollutant Emission Inventory (APEI, 2020). Corresponding biogenic and biomass burning emissions are obtained from the Model of Emissions of Gases and Aerosols from Nature (MEGANv2.1; Hu et al., 2015), and QFED (Koster et al., 2015)."

P11. Is rice cultivation part of the wetland?

No, we include rice cultivation in the anthropogenic category rather than in the wetland category. We have revised the text to clarify this.

"The TROPOMI-derived wetland fluxes (excluding rice) total 173 (155–182) Tg/y globally, representing 29 (26–31)% of the total methane source and 88 (84–91)% of the natural source."

P12. Why talk about livestock in the wetland sections?

Because here we are discussing uncertainties in the model partitioning between livestock and wetlands.

P12. 380. Where is the South Sudd in the figure?

It is shown as box 13 in Figure 2b and we have now clarified this in the text:

"Figure 2a shows that the South Sudd wetlands (box 13 in Figure 2b) are a major methane hotspot that is underestimated in the prior model by a column average of 41 (21–65) ppb."

P13. Can you provide more physical explanations of why these wetland emission models disagree so much? Is it due to their parameterization or the need for more information about water nitrogen content, heat content, depth of wetland, sulfate content, etc.?

These wetland emission models are from the global carbon project and Saunois et al. (2020) describes their schemes as well as the associated discrepancies. Rather than repeat their discussion here we have added a citation as follows:

"Across the wetland regions examined above, Figure S14 shows that our optimized emissions fall towards the middle of the land-surface model estimates from the Global Carbon Project (GCP); details on these bottom-up models and their differences are provided by Saunois et al. (2020)."

P14. The second paragraph: are we so clueless about the wetland anaerobic activity to use a simple correlation analysis? How about the soil nitrogen, water temperature, depth, oxygen content, etc.?

While other factors can indeed control wetland emissions, our aim here is to set the stage and motivate future mechanistic studies similar to what the reviewer suggests. We now clarify this in the text.

P14. L421. Please add a fraction of the total for each sector.

We prefer to leave this sentence as-is, since adding all of the fractions impedes readability. The reader can readily calculate the fractions for themselves from the numbers provided in the sentence.

P15. L444. Are they missing from other top-down emissions too? The bar is usually low for bottom-up emission inventories, especially in developing countries.

As stated in the prior sentence we are referring here specifically to the inventories used as prior in our analysis.

P17. L 515. Does correlation explain causation?

Not necessarily and that is why we used the word "supportive" rather than something more definitive.

P18. L 560. It's not about the density of TROPOMI data but a piece of factual information from XCH4. We need more compounds not denser data.

We agree, please see our earlier replies regarding the potential utility of multi-species constraints.

P16. is this number of available pixels really a lot? Please provide the percentage for a hypothetical situation when clouds were not present.

We believe the absolute numbers stand on their own in this context.

Editorial comments:

P1. L16. What do you mean by separately resolved?

We have changed this wording to "simultaneously".

P1. L20-21. It is vague; does the hydrological adjustment come after or before?

We believe this wording is sufficiently clear as-is.

P1. L22. The sentence (Fossil fuel emission...) is awkward.

This sentence has been revised as requested.

P1. L23. Many -> several

We prefer to keep this wording as-is.

P2. L39. What do you mean by strong heterogeneity? Spatial or temporal?

We have clarified this sentence as requested.

P2. L42. inverse -> inversion

We have modified this sentence to "Here, we apply these data in a 4D-Var inversion + spatial downscaling framework to quantify the 2018–2019 global methane budget and determine the importance of missing and unexpected sources."

P2. Please use +- for a normal range. 18+-1 is shorter and neater. Please apply this to the entire manuscript.

The uncertainty range is not always symmetric about the mean and so we list the mean and range separately.

P8. L242. What do you mean by "spatial source uncertainty"?

We have revised the wording to "spatially biased emissions".

P7. Eq.2. The i->j is weird; what do you mean?

This describes the downscaling from coarse grid i to fine grid j.

Figure 4. The panels are too small.

We have modified the figure to improve visibility.

Figure 5 needs to be enlarged. This is the most critical figure, which is hard to see.

The figure has been enlarged as suggested.

Table2. The numbers in the parenthesis are just the deviation in a defined box, not an actual error. Please inform the readers about it.

We have added a footnote to the table to explain this, as suggested.