Reply to Review 1.

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

The paper is mostly well written, except for the section comparing posterior OH to ATOM results. The main concern I have is that the method and results in this paper are not fundamentally different from those in the Qu et al. ACP 2021 paper which also uses TROPOMI data for quantifying emissions for essentially the same time frame. The main difference between this paper and the Qu et al. paper is the version of the data, which is ostensibly more accurate than the data used in Qu et al. but then there is no discussion on how this improved data set changes, or potentially improves the results over and above the Qu et al. results.

For acceptance, the paper needs to better describe the difference from those in Qu et al, and how these results are an improvement ; I think there are sufficient results in here for this purpose (e.g. monthly estimates allow for attributing some components of the methane budget). In addition, you could compare with the Qu et al. 2022 paper (methane surge) which uses GOSAT data for 2019; in principal the improved TROPOMI data sets should result in better comparisons with the GOSAT based results for this time period.

We have revised the manuscript as follows to better highlight the differences between our study and that of Qu et al.

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)."

As the reviewer suggests, we have also revised the draft to provide more details on the updated data version employed here:

Section 2: "Relative to the albedo-corrected product, the prior TROPOMI version exhibits high biases over North Africa, the Middle East, and the western US, and low biases over Amazonia, the eastern US, central Africa, and eastern China (Lorente et al., 2021)."

Another issue is a lack of discussion on uncertainties; I see them reported in final estimates but its not obvious how they are computed, are these buried in the text somewhere ?(Im pretty sure I read through the entire text, 2.5 times + browsing). A more extensive discussion on uncertainties should be in Section 2.

This is now explained more prominently in the text:

Section 2: "In the following, we interpret the multi-inversion mean as our base-case solution and the range as the corresponding uncertainty estimate."

We have also added a new analysis to further characterize the downscaled solution accuracy using the OSSE described in-text:

Section 2: "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)."

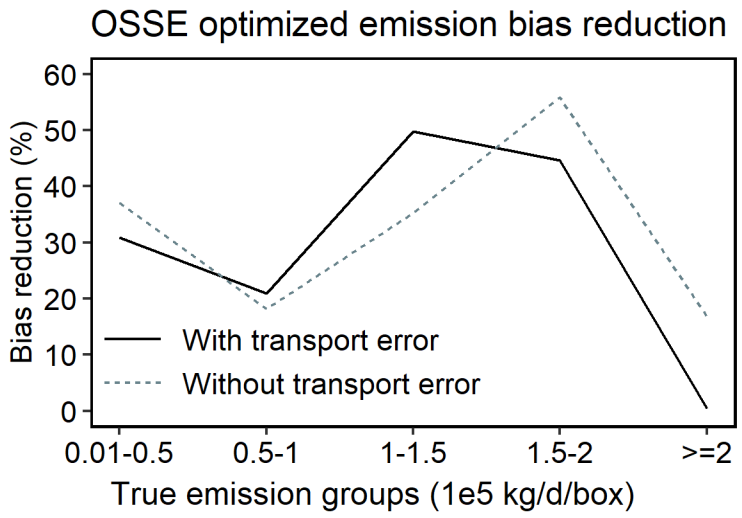

Figure S8. Downscaling bias reduction as a function of emission magnitude, based on 1-month Observing System Simulation Experiments (OSSE) over North America (see main text for details).

Note that Im not convinced that the downscaling approach described here is sufficient by itself to merit publication as it is not (obviously) an improvement over the optimal estimation based approach described in the un-cited Cusworth et al. 2021 paper (see subsequent comments).

We strongly disagree with this perspective. The downscaling method developed and applied here is a novel contribution and uses an entirely different strategy from that of Cusworth et al. (2021). Cusworth et al. (2021) combine the posterior and prior uncertainties to obtain sector-based emission estimates. Our method takes a different approach, combining the prior error estimates directly with the (oversampled) TROPOMI observations themselves to project emissions from coarse to fine resolution. A unique advantage of our approach lies in this use of the sub-model-grid satellite information (which is lost in a standard lower-resolution inversion) to inform the downscaling. Further details are provided in our replies to the more specific comments below.

We now compare our method with that of Cusworth et al. (2021) in text as follows.

Section 2: "Compared to existing emission downscaling methods that rely on prior and posterior error covariance estimates (Cusworth et al., 2021), or are based solely on satellite data (Liu et al., 2021), our approach is unique in combining the prior emission information (and its uncertainty) with the oversampled TROPOMI observations themselves. Variable weighting between these terms permits greater influence from the observations when the prior emissions are more uncertain. The method

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

Specific Comments:

Abstract, "… indicate rapid increases in Middle East"; the way this sentence is currently written implies you base this statement on satellite observations.

We have clarified the Abstract as follows:

"Emissions from fossil fuel activities are strongly underestimated over the Middle East (+5 Tg/y), where bottom-up inventories suggest rapid increases over the past decade, and over Venezuela."

Abstract: state that you are estimating monthly values

We have revised the Abstract as suggested:

"We apply a new downscaling method to map the derived monthly emissions to 0.1°×0.1° resolution, using the results to uncover key gaps in the prior methane budget."

Abstract: You stated you used observations of CO and OH, but I don't see any description of these data in Section 2. Also see comments on comparisons to CO and OH to ATOM below

We have revised the abstract to clarify that the CO and OH data are used for evaluation:

"Employing remote carbon monoxide (CO) and hydroxyl radical (OH) observations with independent methane measurements for evaluation, we infer from TROPOMI a global methane …."

We have also added a description of the evaluation datasets to Section 2, as suggested:

"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)."

Section 2.1 Page 3: As stated in the general comments, the analogous paper here is from Qu et al. ACP 2021 which uses V1.03 TROPOMI data whereas you use the Lorente et al. based corrections; make that difference clear here. Note that as far as I can tell there is fundamentally no difference between yours and the Qu et al. results, notwithstanding the improved XCH4 data sets you are using.

Can you add discussion on this difference in Section 2.1 and then add more comparisons to the Qu et al. 2021 results in Section 4?

We have revised the text in section 2 to explain the differences between our study and that of Qu et al., as described earlier.

We have also modified the discussion in Section 4 to include comparisons with Qu et al. (2021):

"Eight of the ten nations in Table 2 (China, India, US, Russia, Brazil, European Union, Pakistan, Indonesia) are likewise identified by Worden et al. (2022) as among the top ten anthropogenic emitters globally. Our inferred anthropogenic fluxes for the US and China agree well (within ~10%) with the GOSAT-based results from Worden et al. (2022) and with the GOSAT+TROPOMI results from Qu et al. (2021). Anthropogenic emissions derived here are likewise within 10% of the Worden et al. (2022) results for India and the European Union, with both studies lower (20–50%) than Qu et al. (2021). Our results for Russia and Iran are 21–28% higher than the GOSAT-based estimates, mainly reflecting oil, gas, and coal emissions, and ~40% lower for Brazil, mainly due to livestock. Emissions for Pakistan and Indonesia agree to within 1% for the TROPOMI- and GOSAT-based results (Worden et al., 2022). However, we find here that anthropogenic emissions from Bangladesh (7 Tg/y versus a prior of 4 Tg/y) are 3× higher than the GOSAT estimate (2 Tg/y), while adjacent emissions from Myanmar (4 Tg/y) are half the GOSAT estimate. Worden et al. (2022) conclude that the GOSAT-derived emissions for Myanmar are anonymously high due to impacts from their prior assumptions; we attribute much of that flux to Bangladesh and show later that it mainly arises during the South Asian monsoon."

We have also added comparisons with GOSAT-based inversion results to section 5 and 6, as suggested by the reviewer in a later comment:

"Our derived global wetland fluxes are ~20% higher than previous GOSAT-based estimates (145-148 Tg/y: Ma et al., 2021; Zhang et al., 2021), with similar latitudinal distribution to that found by Ma et al. (2021)."

and

"The 2019 global anthropogenic methane emissions obtained here are modestly (12%) higher than GOSAT-based results for 2010–2018 (336 Tg/y; Zhang et al., 2021), with both results pointing to higher-than-predicted biotic emissions (consistent with isotopic constraints; Nisbet et al., 2016)."

Section 2.6, page 7, Provide rationale for why you are downscaling to 0.1 degree resolution, especially since it depends on priors which can vary considerably (uncorrelated at 0.1 degree resolution) depending on choice of prior as you note in the text. As far as I can tell, the downscaled results are not used thereafter in the paper, is that correct? (Note that in the Worden et al. 2022 paper, we downscaled so that we can then upscale more accurately to each country; the other reason for the OE based downscaling (Cusworth et al. 2021) we developed is to step us towards using top-down emissions estimates for updating gridded inventories at this scale).

The existing text already provided some rationale for this. We have now elaborated on that, adding the additional motivation mentioned by the reviewer regarding top-down satellite-informed emission inventories (as indicated below). The downscaled results are in fact used later in the paper to map the advance and retreat of East Asian emissions.

Section 2: "We present here a new method to spatially downscale the satellite-derived emissions for potential use in models. This downscaling is necessitated by the fact that the current GEOS-Chem adjoint model does not have global simulation capability at finer than 2° × 2.5° resolution. Furthermore, each of the 2-year inversions performed here required >12,000 CPU hours (>80 days on multiple processors) to converge, making higher-resolution optimizations computationally impractical. However, the inventories employed as prior, as well as the TROPOMI observations themselves, contain information at much finer scales (e.g., 0.1° × 0.1° and 7 × 7 km2)—and thus contain additional high-resolution constraints that are neglected by the 2° × 2.5° inversions. We therefore leverage this information to spatially downscale the optimized emissions to 0.1° × 0.1° …"

As described earlier, we have added a caveat regarding the reliance on robust prior error estimates:

Section 2: "The method thus assumes robust prior error estimates, a caveat that also applies to Cusworth (2021) and similar methods."

How does this downscaling approach compare to the optimal estimation based approach to downscaling discussed in Cusworth et al Earth Environ 2, 242 (2021). Can you perform a test(s) similar to what is shown in Cusworth et al. to ensure you are preserving information from original grid and downscaled grid? Your co-author A. Bloom designed these tests for Cusworth et al. so you could ask him for details. Note that I would be ecstatic for an additional vetting of this OE/Cusworth approach by the Dylan / Daven crew… we are pretty sure we got the math right as we used two different approaches to arrive at the same result (the Cusworth / Bloom and the Bowman approach, with Worden moderating), but given that its a 30+ equation derivation some additional vetting is desired.

Also cite Liu, M. et al. A New Divergence Method to Quantify Methane Emissions Using Observations of Sentinelâ5P TROPOMI. Geophys Res Lett 48, (2021), as a potential way to use satellite data to identify and quantify emissions at these same fine spatial scales.

This downscaling approach differs significantly from that of Cusworth et al. (2021) and we have added a discussion of this point to the paper as described earlier.

The test mentioned by the reviewer and presented in Cusworth et al. (2021) specifically evaluates the sectoral partitioning of that method. Our approach is purely spatial so such a test is not applicable. Instead we performed a dedicated OSSE experiment that demonstrates the performance and robustness of the method. This evaluation is described in the paper and we have now included the additional evaluation shown in (new) Figure S8 and described above.

Section 3.2. As a reader I did not understand either the rationale for the comparison to ATOM, or how I should interpret the comparison…. This section basically needs a re-write. Note that our group at JPL also attempted to use the ATOM OH estimates but decided against it (although this was a few years ago) because we did not have a good sense of the accuracy, especially since OH is tricky to measure; some discussion is needed on the ATOM OH accuracy to better interpret the comparison between your inversion results and these in situ results. Also, what did you intend to conclude from the comparison to CO?

We have added new information to Section 2 regarding both the motivation for using the ATom data and the measurement 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)."

We have also modified Section 3.2 to include additional statistical tests and to 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."

Regarding the rationale for the ATom comparisons and the conclusions drawn from them, we believe these are made clear in the updated version from the following statements:

Section 2: "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."

Section 3: "Remote observations of OH and CO (the primary OH sink) from the ATom airborne campaign (Wofsy et al., 2018) also point to an OH underestimate in the optOH solution."

Section 3: "When co-optimizing methane emissions and loss we thus find that the solutions can achieve a good fit to the TROPOMI data themselves but degrade model agreement with other observations of methane, OH and CO. We conclude that solving for global methane sources and sinks based solely on satellite observations of methane itself remains an under-constrained problem—even with the dense TROPOMI data coverage."

Section 5.0, Compare against the Ma et al. 2021 and Zhang et al. 2021 wetland results which suggest ~149 Tg CH4/yr total…this again might be a TROPOMI versus GOSAT issue as TROPOMI data results in lower livestock emissions than those from GOSAT in Brazil, which in turn would likely balance to the wetlands, relative to the GOSAT based results. A discussion here on these differences is needed.

Section 6, again compare these totals to the GOSAT based estimates (there are several now available). Discussion on potential TROPOMI / GOSAT differences are needed as well.

We have now added comparisons to GOSAT-based results as follows:

Section 5: "Our derived global wetland fluxes are ~20% higher than previous GOSAT-based estimates (145-148 Tg/y: Ma et al., 2021; Zhang et al., 2021), with similar latitudinal distribution to that found by Ma et al. (2021)."

Section 6: "The 2019 global anthropogenic methane emissions obtained here are modestly (12%) higher than GOSAT-based results for 2010–2018 (336 Tg/y; Zhang et al., 2021), with both results pointing to higher-than-predicted biotic emissions (consistent with isotopic constraints; Nisbet et al., 2016)."

In addition, the paper already compared the country-level emissions with GOSAT-based estimates, as follows:

Section 4: "Eight of the ten nations in Table 2 (China, India, US, Russia, Brazil, European Union, Pakistan, Indonesia) are likewise identified by Worden et al. (2022) as among the top ten anthropogenic emitters globally. Our inferred anthropogenic fluxes for the US and China agree well (within ~10%) with the GOSAT-based results from Worden et al. (2022) and with the GOSAT+TROPOMI results from Qu et al. (2021). Anthropogenic emissions derived here are likewise within 10% of the Worden et al. (2022) results for India and the European Union, with both studies lower (20–50%) than Qu et al. (2021). Our results for Russia and Iran are 21–28% higher than the GOSAT-based estimates, mainly reflecting oil, gas, and coal emissions, and ~40% lower for Brazil, mainly due to livestock. Emissions for Pakistan and Indonesia agree to within 1% for the TROPOMI- and GOSAT-based results (Worden et al., 2022). However, we find here that anthropogenic emissions from Bangladesh (7 Tg/y versus a prior of 4 Tg/y) are 3× higher than the GOSAT estimate (2 Tg/y), while adjacent emissions from Myanmar (4 Tg/y) are half the GOSAT estimate. Worden et al. (2022) conclude that the GOSAT-derived emissions for Myanmar are anonymously high due to impacts from their prior assumptions; we attribute much of that flux to Bangladesh and show later that it mainly arises during the South Asian monsoon."

Section 6.2, Note that reports to UNFCC from Russia have varied considerably over the years, this should be discussed here (e.g. Scarpelli et al. 2021 versus Scarpelli et al. 2022).

We thank the reviewer for pointing this out and now discuss it in-text:

Section 6: "Indeed, subsequent revisions (year-2019; Scarpelli et al., 2022) to the UNFCCC-2016 inventory used here have strongly reduced fossil fuel emission estimates for Russia (e.g., 21 Tg/y from oil in year-2016 vs. 2 Tg/y in year-2019) due to updated emission factor assumptions."

7.0 Conclusions (and to some extent abstract). The paper implies that missing sources can be identified through the downscaling approach, but this is not possible if you are using prior emissions for the downscaling. Also, how can the Venezuelan source simultaneously be lower than the prior and inline with trend estimates? These are different quantities. I think you mean something else here.

The approach can in fact identify missing sources because it directly incorporates the downscaled TROPOMI observations themselves. As described in the paper, it is only the sectoral partitioning that relies solely on the prior, not the flux magnitude or location.

We have simplified the text about Venezuela for improved clarity:

Section 7: "We estimate long-standing fossil fuel sources in Venezuela at 4.8 (3.8–5.3) Tg/y, 28 (2–41)% higher than the prior estimate (which is for year-2016)."

7.0: Line 540 Conclusions about waste and agriculture priors being too small… yes we are finding this to be the case with all the other published TROPOMI and GOSAT based inversions, please reference these other papers.

We added comparisons to other relevant studies in the results and discussion sections, as described in our earlier replies.

7.0 Conclusions / Line 555: This conclusion is potentially very interesting but needs additional vetting. For one, how much of yearly Indian and Southeast Asian underestimate is due to the underestimate in the Monsoon seasons? In addition, how much of this is affected by smoothing error, which is not directly calculated using your method, but you could calculate by using different priors; basically we are finding significant impact of smoothing error, or alternatively cross-correlation of a change in one emission onto another, for emissions and their trends in this region.

Rather than using alternative priors, we have used a suite of inversions that employ substantially different optimization frameworks. These will be affected by smoothing to different degrees and we interpret and discuss our results in the context of the resulting uncertainty as diagnosed from the range across these solutions.

The reviewer's question about how much of the Indian/SE Asian underestimate falls during the monsoon is addressed by the following:

Abstract: "More than 45% of the global upward anthropogenic source adjustment occurs over India and southeast Asia during the summer monsoon (+8.5 Tg in Jul–Oct)."

Section 6: "Approximately 80% of India's annual rain falls during the summer monsoon (IPCC, 2021), and across this Jul-Oct season we find that methane emissions from the India and Southeast Asia boxes in Figure 2b are underestimated by 37 (15–45)%. The resulting seasonal flux increase then accounts for over 68% of the total annual emission correction."

References: You can peruse the Worden et al. ACP paper for missing references on GOSAT inversions that you can then compare to in the text; this same comment was made by reviewers of our Worden et al. paper.

We have revised the draft and cited these references accordingly, as indicated in our earlier relies.