We thank the reviewers for their suggestions and comments on the manuscript. Please see below our responses to the comments. We have listed out the reviewer comments in black and the replies in blue.

Reviewer #1

This manuscripts uses TROPOMI and GOSAT observations to constrain methane emissions in Colombia at high resolution for 2023. The authors develop a high-resolution gridded bottom-up inventory of methane emissions. Using GEOS-Chem as a forward model, they solve an analytical Bayesian inversion to update the bottom-up inventory. The main value is the development of a high-resolution gridded emissions inventory constrained by satellite observations that yield specific suggestions for improving bottom-up estimation methods. I find the work to be novel, well-written, a valuable addition to the field, and suitable for ACP. However, I have major questions regarding validation of posterior emissions and robustness of the sector attribution.

Thank you for taking the time to review and help improve our manuscript.

Major comments:

1. L341-342: This argument makes sense. However, I do not believe the posterior error variance is shown in the main text or SI. Could the authors provide further justification of this claim by comparing posterior error variance at each grid cell to the variance of the inversion ensemble, perhaps in the supplement? Additionally, it would be useful to see a comparison of the prior and posterior error variance at each grid cell.

Thank you for this suggestion. In response, we now include a new analysis in Figure S2 comparing the posterior standard deviation derived from $\hat{\mathbf{S}}$ with several ensemble-based uncertainty metrics across spatial aggregation scales, including the ensemble standard deviation, the ensemble standard deviation including the posterior variance of each ensemble member, and the ensemble range. This analysis shows that the ensemble range is the most conservative uncertainty measure at all aggregation scales, while also clarifying how the other uncertainty metrics depend on spatial averaging. We have added this discussion to the ensemble subsection (lines 362-369).

2. The posterior emissions would be more compelling if they were independently validated. To my understanding, only in-sample validation has been shown (Figure 7). Could the authors show that the posterior emissions improve simulation of TROPOMI and GOSAT observations that are held out of the inversion? Either in 2023 or in a different year?

We agree that independent validation would strengthen this manuscript. For this work, we included all available TROPOMI and GOSAT observations for 2023 to maximize the observational constraint over Colombia. We also include a comparison with coal point-source observations from GHGSat, CarbonMapper, and EMIT in Figure 8, but we agree that this does not constitute a complete independent evaluation. To our knowledge, there are no additional in situ methane observations available over Colombia for 2023 that could be used for independent validation. Withholding a fraction of observations as validation dataset would not help because (unlike ML) the Bayesian inference solution is regularized. Using observations from another year to validate the 2023 posterior emissions would also be difficult because wetland emissions, which make up the largest portion of total emissions, vary substantially from year to year. We are currently conducting a follow-up analysis looking at seasonality and trends from 2019 to 2024 over Colombia. In this future work we plan to follow your suggestion by using TROPOMI observations for the inversion and reserving GOSAT observations for independent evaluation. We now clarify the lack of available validation data in lines 582-583.

3. Figure 6: it looks like the largest changes to the emissions occurred in regions coincident with observations (Figure 7). Could the authors show a map of the diagonal elements of the averaging kernel

(i.e., Aii) or DOFS averaged over regions to show where posterior emissions are constrained by observations? If there are regions that are not constrained well by the observations, can the authors discuss them? This is partially addressed with the discussion of Carbon-I (L538) but it would be clearer if this was directly shown.

Thank you for this suggestion. We have now included a supplementary figure showing the averaging kernel sensitivities (Figure S3) and added corresponding discussion in the text (lines 345-347, 380, 418-419, 444-456, 623-624).

4. In general, I am skeptical of the posterior sector attribution for non-point sources (wetlands, livestock, rice) in regions where these emissions are co-located (e.g., wetland emissions in the La Mojana region and Magdalena River with livestock emission in Figs. 4 and 3). The authors assume that posterior sector emissions in each grid cell are proportional to prior sector emissions. While total sector emissions errors are generally uncorrelated (Figure S2), this does not mean that finer-scale attribution is correct. High uncertainty in the spatial distribution of prior non-point source emissions will propagate to the posterior sector emissions. The following comments relate to this point:

We agree that inversion results depend heavily on the prior distribution of emissions. This is a key point of this paper that motivates our construction of Colombia-specific inventories using national bottom-up data rather than using global emission inventories. We now highlight this and the uncertainty in the text (lines 26-28, 121-123, 381-385, 663-665).

a. The authors briefly discuss this uncertainty in L548-550. However, because this is a key assumption affecting interpretation, the sector attribution method should be made clearer in the abstract and conclusion.

We have added further discussion of this throughout the text as described above, but we specifically modified lines 26-28 in the abstract and lines 663-665 in the conclusion.

b. Figure S2 shows low posterior error correlations between sectors, but this does not necessarily demonstrate that sectors are independently constrained by the observations. Could the authors provide the sector-resolved averaging kernel (i.e., WAWT) as another diagnostic to assess whether co-located sectors can be meaningfully distinguished (e.g., whether the diagonal elements are large and off-diagonal elements are small), or whether sector attribution is primarily driven by prior assumptions?

Thank you for this suggestion. We have added a supplementary figure showing reduced averaging kernel sensitivities by sector (Figure S7) and discuss this in the text (lines 444-456).

c. Can the authors provide more discussion of how the uncertainty in the distribution of the prior wetland emissions could affect posterior sector emissions? While sensitivity tests using global inventories (Figure S2) are helpful, the problem remains underdetermined, and these tests may not fully explore the uncertainty space.

Yes. We have added further discussion of this to the manuscript (lines 381-385, 646-651). We are also currently conducting a follow-up study optimizing emissions at monthly temporal resolution to better constrain the wetland source.

d. Figure 8: Could the high posterior emissions factor in Magdalena Medio have been a misattribution of wetland emissions to livestock emissions? The GLWD + LPJ prior shows weak but notable emissions in that region. If the prior wetland emissions were biased low or too spatially concentrated, could the posterior livestock emissions be biased high?

Thank you for this suggestion. We now state this possibility more clearly in lines 640-642.

e. It would be useful to see a map of the change in prior and posterior sector emissions to better evaluate sector attribution. This is done for the wetland emissions in the La Mojana region (Figure S4), but difference plots would be useful for the sector emissions in general.

We have added a figure showing emission differences for major sectors (Figure S5).

Minor comments:

5. It would be nice to see an explicit comparison of the spatial distribution of wetland emissions from GLWD + LPJ compared to LPJ alone. e.g., a difference plot.

Thank you for the suggestion. We have now added a difference plot as a supplementary figure (Figure S1).

6. It is not clear to me how the seasonal cycle of wetland emissions is considered in the inversion. I assume that the K matrix contains this information via the time-resolved forward model? If that is the case, could the authors make that clearer?

Yes, that is correct. We optimize annual emissions such that the inversion assumes the temporal distribution of emissions follows that of the prior estimate. The time-resolved forward model and the resulting K matrix carry the effect of that seasonality into the inversion. We now clarify this in the manuscript (lines 341-343).

Reviewer #2

This manuscript presents a 12 km x 12 km methane inversion over Colombia using Integrated Methane Inversion framework. The study combines satellite measurements from TROPOMI and GOSAT with their national bottom-up inventory from the Biennial Update Report and GLWDv2 inventory for wetland emissions. The inferred posterior emission fluxes indicate substantial upward revisions in the anthropogenic emissions, particularly in oil & gas, and coal sectors. The study is timely and relevant, and can be accepted after addressing the following comments:

Thank you for taking the time to review and help improve our manuscript.

Line 85: The authors claim that the high resolution allows them to directly attribute their posterior emission to specific sectors. However, 12 km is still too coarse to resolve all of the sectors. As a result, authors merge several collocated sectors in this study. I recommend that the authors rephrase this claim to avoid overstatement and more explicitly acknowledge that sectoral attribution is only partially resolved at this scale, and in some cases relies on prior information or aggregation assumptions.

Thank you for pointing this out. We do not intend to claim that all sectors are fully resolved at 12-km resolution. Rather, our point is that 12-km represents a substantial improvement over previous work and allows stronger sectoral interpretation. We now soften the language accordingly (lines 100-103) and further highlight this point in the methods section (lines 381-385).

Line 101: The prior inventory is constructed to match the national total in BUR, which introduces an implicit constraint on the overall emission magnitude. This raises concern regarding the interpretation of posterior uncertainties. For example, if the national total in BUR is underestimated, the prior uncertainties will be underestimated as well. Typically, the posterior uncertainties are smaller than the prior uncertainties and that will mean that posterior uncertainties will be underestimated in that case. A sensitivity analysis using an alternative prior would help assess the robustness of the inferred emissions and their uncertainties. If that is computationally expensive to run, then I would recommend conducting sensitivity analysis with scaling the prior and evaluating on independent observations similar to Figure 7.

We appreciate this comment. The purpose of the prior inventory in the inversion is to provide a bottom-up constraint that is as close as possible to the true emissions, based on the best available information for Colombia. The inversion then determines the optimal emissions through the balance between fit to this prior estimate and fit to the satellite observations. For this reason, we place strong emphasis on constructing the prior from local data rather than using an alternative prior that would be less representative of national conditions. We now emphasize this more clearly in the manuscript (lines 121-123). To assess the robustness of the inferred emissions, we use an inversion ensemble in which we vary the prior error standard deviation from 50% to 200%. We report posterior uncertainty as the range across this ensemble, which allows for a broader departure from the prior and provides a practical measure of robustness while retaining the same best-estimate prior.

Line 104: Were the emissions from the point sources assumed constant or did they have temporal variability based on persistence or repeated observations? Additionally, if the BUR total emissions were preserved, does that mean increasing oil and gas emissions require decreasing other sector emissions?

Thank you for this question. For the addition of point source observations to the gridded, bottom-up inventory, we do not estimate grid cell emission rates based on the point source observations; we just ensure that the location of each point source detection has emissions present for that sector in the prior estimate. We preserve BUR sectoral totals unless more up-to-date information from Ardila et al. (2025a) is available. The total anthropogenic BUR emissions are not preserved. We now clarify this in lines 126-130.

Line 203: Minor grammatical error: "national total estimate of of 0.049 Tg a-1".

Thank you, we have corrected this error.

Line 242: The analysis focuses on the year 2023, however, the rationale for selecting this specific year is not clearly justified. Given that TROPOMI measurements are available from 2018 onward, it would be valuable to understand why a longer temporal analysis was not conducted. Extending this study period could help assess the robustness of the results, capture interannual variability, and better constrain sector emissions. The authors should clarify their choice of 2023 and discuss whether their framework could be applied to a multi-year analysis.

We selected 2023 for this analysis because it is the year for which the supporting datasets used to construct the prior estimate are most complete, including updated inventory information from Ardila et al. (2025a) and point-source observations used to improve the prior spatial allocation. We now clarify this in lines 277-279. We agree that extending the analysis across multiple years would be valuable, and we now mention this in the text (lines 646-651). We are presently working on a multi-year analysis from 2019 to 2024 as a follow-up to this work.

Line 297: Why both prior and observational error covariance matrices are assumed to be diagonal? Not having off-diagonal terms in the error covariance matrices can limit the inversion's ability to correct structural biases, and the posterior fluxes may retain spatial patterns from the prior.

We agree that including off-diagonal terms in the prior and observational error covariance matrices can better represent correlated errors. In this study, however, we assume diagonal covariance matrices because the correlation structures of prior and observing system errors are not well constrained over Colombia. Introducing off-diagonal terms without a defensible basis would require additional assumptions that could themselves strongly influence the posterior. Instead, following standard practice when these error statistics are uncertain, we test the sensitivity of the inversion to the magnitude of the assumed errors through the inversion ensemble. We have revised the text to clarify this (lines 336-337). We also now explain that improving the treatment of errors would improve this work (lines 643-645).

Figure 6: The spatial structures in the posterior emissions look very much similar to the prior, given the limited observations, off-diagonal terms in the error covariance matrices may help in constraining the posterior fluxes better.

We appreciate this suggestion. We do not use off-diagonal terms for the reasons described above, but this could be a solution in other regions with sparse observational coverage and better knowledge of error correlations.

Line 340: The authors used the posterior uncertainty as the evaluation metric to select the best posterior estimate of emissions. Does the comparison between model simulated and observed concentrations also follow the same order of performance?

We do not use the posterior uncertainty to select the best estimate of emissions. The results shown in the main text are from the base inversion, which uses the median choice of each inversion hyperparameter, while the ensemble is used to characterize uncertainty. We now clarify this distinction in the manuscript (lines 391-392). All ensemble members successfully reduce the bias between modeled and observed concentrations.

Line 350: The authors compare the GEOS-Chem model bias relative to the TROPOMI and GOSAT observations and highlight significant improvements over the bottom-up inventory. However, it is not clear if the observations used in this analysis and Figure 7 are independent observations from this inversion or not. I strongly recommend using independent observations as typically the observations used in the flux inversion will have a better fit.

The purpose of Figure 7 is to show that the inversion effectively reduces the bias between the satellite observations used in the inversion and the modeled concentrations based on emissions. These observations are therefore not independent from the inversion, and we now make that more explicit in the text (line 417). We agree that comparison with independent observations would be preferable, but to our knowledge there are no in-situ methane observations in Colombia for 2023 that could be used for validation. We do include a comparison with coal point-source observations from GHGSat, CarbonMapper, and EMIT in Figure 8, but we agree that this does not constitute a complete independent evaluation. In future work, we plan to exclude GOSAT observations from the inversion and instead use them for independent evaluation. For this work, we now clarify the lack of available validation data in lines 582-583.

Line 512: Can there be any physical reasoning behind the decrease in the reservoir emissions? For example, reservoir emissions are generally the highest in the first two decades and decrease over time after that.

The downward adjustment in reservoir emissions is not consistent across the ensemble, partly due to spurious information propagated from livestock in the posterior estimate (we now indicate this in line 596). We therefore avoid over-interpreting this decrease and instead focus on the more robust signal of increased emissions at the five reservoirs where we find a significant emission adjustment.