We thank the reviewers for their comments, in black. Responses are in blue with references to the changed lines in the updated manuscript referenced in **bold** where applicable.

Reviewer 1 Comments

Major comments:

1. The study adopts a step-wise approach, using the TROPOMI posterior as a prior for the MethaneAIR inversion. The authors mention that concatenating data was avoided due to temporal mismatches. However, have the authors considered or tested a joint inversion where both observation vectors are included in a single cost function with appropriate error covariance matrices as used in the paper "Global methane budget and trend, 2010–2017: Complementarity of inverse analyses using in situ (GLOBALVIEWplus CH4 ObsPack) and satellite (GOSAT) observations"(https://acp.copernicus.org/articles/21/4637/2021/)? A discussion on why the step-wise approach is superior or more practical in this context would be valuable.

We did not test a joint inversion with both observation datasets joined in a single cost function. The motivation for the stepwise approach is twofold.

1. Because of the temporal mismatch between the inversions (monthly mean emissions for TROPOMI vs daily mean emissions for MethaneSAT) we find the stepwise approach more appropriate due to the known sub-monthly temporal variability of emissions
2. Because there exists observational bias between these datasets we prefer to optimized the boundary conditions for each separately to better reflect the individual instrument characteristics

We have added additional detail to the existing discussion of this topic on **lines 241-244:** "Concatenating the TROPOMI and MethaneAIR data into a single observation vector would not be as effective because of the different time scales over which the observations operate, **so that the optimized emissions are different (daily for MethaneAIR, monthly for TROPOMI), and also because the forward model uses different boundary conditions for each instrument.**"

2. In the Permian case (RF06, Figure 1), TROPOMI XCH4 is significantly lower (11 ppb on average) than MethaneAIR. It is surprising that the independent inversions yield nearly identical total emissions (84.8 vs. 85.9 t/h) despite this substantial systematic offset. Could the authors clarify how the inversion achieves this consistency? Could this result primarily be driven by the adjustment of boundary conditions to compensate for the observational bias?

This consistency is indeed owed to the optimization of boundary conditions for each instrument's inversion compensating for instrument bias. This reenforces our stepwise approach justifications outlined in our response to question (1). We have added additional text in Section 3.1 to clarify this point on **lines 337-341**

"The MethaneAIR and TROPOMI inversions, driven by independent observations, show **general** consistency in their posterior emissions, totalling 85 t h^{-1} in the MethaneAIR inversion and 86 t h^{-1} in the TROPOMI inversion. **This is despite an average 11 ppb lower mean observed concentration in the region for TROPOMI compared to MethaneAIR and very**

different spatial distributions (Figure 1). Systematic bias between the two instruments would be corrected by the boundary conditions, while differences in spatial distributions can be attributed to variability in transport or emissions over the different sampling periods."

3. The coefficient 4.7 in equation (3) is unclear. Is this a fitted slope from the resampling experiment? Moreover, given that RF06 and RF08 likely differ in spatial heterogeneity, the authors should justify whether this single empirical constant is applicable to both flights, or if flight-specific coefficients would be more appropriate.

This is a linear regression fit from the resampling experiment. Our goal was to create a generalized statement on representation error for the instrument with our available data. Representation error makes up a relatively small amount of our total error budget. Text has been added on **lines 207-215** to clarify these points:

"To estimate this representation error, we select all 12×12 km2 grid cells in our **RF06 and RF08** datasets where there is full MethaneAIR observational coverage, successively remove data from 10% of each grid cell area and calculate the root-mean-square error (RMSE) of the resulting super-observations compared to the fully observed grid cells. **Linear regression of the RMSE versus the observed area fraction C yields** the representation error standard deviation σ_R (ppb):

$$\sigma_R = 4.7\,(1 - C) \tag{3}$$

The representation error affects only one third of the MethaneAIR super-observations for a given scene (Figure 1, middle panels) and even there it is generally small compared to the model transport error."

Minor comments:

1. The Introduction (Line 39) states that TROPOMI provides continuous data at a resolution of 5.5×7 km2. While this is true for this study period (2021), TROPOMI XCH4 products were initially provided with a 7×7 km2 resolution and switched to 5.5×7 km2 on August 6, 2019. This context should be corrected.

We have added text on **lines 40-41** to include this more complete information on the TROPOMI observation characteristics:

"**Prior to August 6, 2019, nadir pixel resolution for the instrument was at 7×7 km^2 resolution. TROPOMI**"

2. Please provide the specific temporal information and native spatial resolution for the GHGI and Omara et al. (2024) emission inventories.

Both inventories are at 0.1°×0.1° resolution with annual mean estimates. Clarifying text to this effect has been added to **lines 246-250**

"We use for the inversion the default prior emission inventories at 0.1°×0.1° resolution from IMI 2.0 (Estrada et al., 2025) including the gridded US EPA Greenhouse Gas Inventory (GHGI) (Maasakkers et al., 2023). We supersede the emissions from the oil/gas sector with those from Omara et al. (2024), which include more measurement-based and facility-specific information than the GHGI **and are also produced at 0.1°×0.1° resolution. Both GHGI and the Omara et al. (2024) inventory report emissions as annual mean estimates.**"

3. MethaneAIR observations occur in the morning (10:00–12:00 LT), whereas the TROPOMI overpass typically occurs in the early afternoon (~13:30 LT). Apart from the transport errors mentioned, how are diurnal variations in emissions and boundary layer height accounted for when comparing or integrating these two datasets?

We address these concerns through additional text on **Lines 126-131**

"This may be expected from TROPOMI measuring a monthly average **at 13:30** local time versus MethaneAIR measuring a single day at **10:00-12:00 local time** and could reflect temporal variability in transport as well as in emissions. The effect of variability in transport is accounted for in the inversion **through the hourly meteorological data including mixing depths driving the GEOS-Chem CTM, but temporal variability of emissions is not accounted for due to lack of reliable information. Varon et al. (2023) previously reported a 25% week-to-week variability in emissions from the Permian.**"

4. When aggregating very-high-resolution MethaneAIR pixels to the 12-km grid, is strict mass conservation or column-weighting maintained?

Column weighting is maintained through the construction of a characteristic averaging kernel for each super observation as addressed in **lines 147-149.**

Mass conservation is maintained. We add text to clarify this on **lines 93-95.**

"**Here we co-add pixels to 1×1 km^2 resolution with a mass-conserving regridding algorithm to display the structure of the MethaneAIR data, and we further co-add to 12×12 km^2 resolution for application of the IMI.**"

Reviewer 2 Comments

Special comments:

- Figure 4. Why were posterior methane concentrations from TROPOMI not compared in this study, while methane emissions were compared with TROPOMI in Figure 3? Also, the same question for the following figures.

  The ability of the IMI to improve the fit to observations with TROPOMI is well documented and not of interest in this work. We add text to address this in **lines 325-328**

“Figure 4 shows X_{CH4} differences between GEOS-Chem simulations driven by these emissions and the MethaneAIR observations. **The reduction in X_{CH4} differences when using posterior versus prior emissions indicates the ability of the observations to optimize emissions in the framework of the IMI. This was demonstrated before for monthly TROPOMI observations in the Permian (Varon et al., 2022) and we focus here on MethaneAIR.**”

- Section 3.1. The author posted two figures first in this section, but only a few words about the concentrations shown in Figure 4. I suggest adding more discussion of concentrations, as the spatial distributions driven by posterior and prior emissions were still different.

Additional text has been added to address the respatialization of emissions in the inversion and its effects on the concentration biases shown in Figure 4 **lines 331-332**

“The three inversions correct that bias by increasing emissions to 83-86 t h^{-1}. RMSEs are also improved from 9.8 ppb to 5.2-5.7 ppb. **The respatialization of the emissions by the inversion corrects the systematic low bias on the southwestern edge of the domain.**”

- Section 3.2, Any differences in the prior emissions between GHGI and Omara et al., 2024 in the region of Uinta?

Because of the lower total magnitude of emissions in the Uinta case, signal is generally much weaker which was why we chose to focus on the Permian scene for discussion of the varied prior emissions experiment. We add text to address this in **lines 408-411**

“**Our analysis of the Uinta observations including much less information content than the Permian thus supports the Permian results that MethaneAIR and TROPOMI are consistent in their optimization of emissions, and that TROPOMI provides useful prior information for MethaneAIR inversion.**”

- Line 328, The data shown here was not consistent with the values in Figure 8.

We have corrected the value for accuracy and consistency, now on **line 406**

“It further improves on the bias relative to the MethaneAIR observations, down to **1.2** ppb, with an RMSE of 3.6 ppb.”

- Is the framework developed in this study also suitable for other point-source methane observation satellites, such as PRISMA or GHGSAT, rather than MethaneSAT?

Unfortunately, the observing window of other point source observing satellites is too small for this method the be applied. PRISMA (30 km window) and GHGSAT (12 km window) would only be able to produce a few total 12 km super observations for a given scene. Because 12 km is the finest resolution available in GEOS-Chem, the integration of these instruments is not feasible at this time.

- Some small errors. Lines 101, change “10 and 12 local time” to 10:00 and 12:00 local time”, similar to that in line 104. Line 208, change “Total” to “total”.

We have the suggested edits to the time notation. The capitalization of total was an error of a missed period rather than a capitalization issue. The period before Total has been added.

Reviewer 3 Comments

The manuscript presents an interesting extension of the IMI framework to integrate high-resolution airborne methane observations. It is indeed timely, given that high-resolution empirical datasets both from remote sensing and aerial surveys are growing. Therefore, the results discussed are encouraging and indicate a useful application of an existing framework. The paper is also well written and easy to follow. However, I have a few concerns as listed below, which should be addressed to strengthen the paper.

1. The chosen argument for validating the integration

Authors chose to show the robustness of their integration by comparing a one-day airborne observation with monthly mean TROPOMI data. I believe, this is not sufficient to demonstrate that the framework performs equivalently for high-resolution data. Methane emissions from oil and gas systems are known to exhibit substantial temporal variability as also noted by the authors, including intermittency and episodic releases. As a result, the “remarkable” agreement between a snapshot airborne inversion and a monthly mean satellite inversion could be partly coincidental unless the authors can demonstrate that emissions in the study region are relatively stable over time. So mere comparison of posterior totals does not necessarily imply similar observational constraint.

Nevertheless, to strengthen this aspect of the study, I suggest that the authors:

- provide evidence for limited temporal variability in emissions over the study period, OR
- perform additional tests of consistency between the airborne and TROPOMI inversions. For example, running inversions with multiple prior estimates (authors can use 0.75x,1x, 1.25x of the same prior or use multiple different priors altogether) and examining whether the airborne and TROPOMI inversions “respond” similarly to changes in prior magnitude and prior uncertainty. So, instead of just the “posterior totals”, comparing the “behavior across different settings” would provide a more robust demonstration that the airborne observations are being assimilated in a manner consistent with the TROPOMI-based framework.

We have adjusted language throughout including the removal of the phrase “remarkable” to more accurately reflect the consistency of the results to be encouraging rather than conclusive. We instead point out that the emissions are “generally” consistent between approaches.

Regarding the variability attributable to temporal variability, we now include text acknowledging the impact of temporal variability on these results in **lines 223-225**

"Comparison of (1) and (2) evaluates the consistency between the emissions inferred by MethaneAIR and TROPOMI, **acknowledging the potential for unresolved temporal variability of emissions between the daily time scale optimized by MethaneAIR and the monthly timescale optimized by TROPOMI.**"

2. Selection of prior uncertainty (GSD) based on concentration RMSE

The prior uncertainty (GSD) is selected by minimizing the RMSE between GEOS-Chem simulations (using posterior emissions) and the same observations used in the inversion. While this is an interesting approach to balance the regularization and data fit, it can introduce a degree of circularity, as the same dataset is used both to constrain and evaluate the inversion. I would suggest authors to compare their optimal prior GSD with independent estimates (e.g., based on inventory uncertainty or facility-level variability)

Because uncertainty within the inventories is defined in normal space, and for yearly mean emissions, it is not appropriate to be directly used in our inversion. Estimates of facility level variability themselves are highly uncertain and selection of these parameters directly would amount to guesswork/selective inclusion. Instead, optimizing the RMSE effectively constrains these parameters while guarding against overfit through the analysis done in section 2.4 and Figure 2. We have added text to clarify this and reference to similar analysis done in Lu et al. 2021 and Jacob et al 2002 **lines 283-296**

"The geometric standard deviation of (GSD) of the lognormal error PDF for the prior emission estimate is a particularly important parameter **but we do not have independent information for it.** We select it as shown in Figure 2 on the basis of the resulting RMSE of the GEOS-Chem simulation with posterior emissions compared to MethaneAIR observations. **This effectively seeks a minimum for the cost function in fitting the inversion results to the observations, as is commonly done to fit inversion parameters (Jacob et al., 2002; Lu et al., 2021).** A low prior error GSD does not allow the inversion to fit the observations, but a high prior error GSD **also limits the ability of the inversion to fit the observations by not taking advantage of prior information,** particularly when using a high-quality prior estimate as with Omara et al. (2024)."