Authors' Response to Reviews of

Operational XCH_4 Retrievals from MethaneSAT: Demonstrating Sensor Performance for Constraining Regional Methane Emissions

,

RC: ***Reviewers' Comment***, AR: Authors' Response, ☐ Manuscript Text

1. General Comments

RC: ***A strong and important paper documenting the first operational MethaneSAT XCH4 retrievals. The stripe-correction and validation work is careful overall. Comments below are mostly requests for clarification or additional detail; points 2 and 7 affect the headline accuracy claims and I'd like to see them addressed.***

AR: We thank the reviewer for their positive assessment of the paper and for their careful and constructive comments. We address each point individually below.

2. Specific comments

RC: ***1. L4, spatial resolution. Native pixels are ∼110×400 m^2 but precision is reported at 2×2/5×5 km at L14-16. A sentence on why this resolution was chosen, presumably to support cloud/aerosol screening and stripe/ISRF characterization rather than the reported precision itself, would help readers.***

AR: We perform the retrieval at native resolution to improve cloud/aerosol clearing and more accurately account for the ISRF (and potentially reduce errors from issues such as inhomogeneous illumination which can bias results near hard dark/bright boundaries like coastlines.

The requirement 5×5 km^2 spatial scale comes from emissions inversion OSSEs performed during mission planning (mentioned at line 429 of preprint). The 2×2 km^2 value was in reference to this, stating we could meet the mission precision requirement at that scale. I think I made the wording confusing as the other reviewer also raised this, so I have updated the line to be clearer:

> This corresponds to a precision of better than 3 ppb when aggregated to 2×2 km^2, meeting the pre-launch precision requirement at a much finer spatial resolution than the specified 5×5 km^2.

RC: ***2. L19-20, Section 4.2, CO2 prior choice. The mean bias against TROPOMI is similar under either prior (∼0.1 ppb), but the regression slope and hemispheric behavior differ substantially: under the operational GGG2020 prior the slope is 1.069, driven by an 8 ppb low bias in the Southern Hemisphere (vs. +2 ppb in the North); under CAMS the SH bias drops to 1.4 ppb and slope to 0.991. The abstract's slope figure of 0.99 is explicitly tied to the CAMS substitution. Please state more clearly in the abstract/conclusions which prior underlies which reported number.***

AR: Now we quote both numbers to make this clearer. In the abstract:

> Comparisons with XCH_4 from TROPOMI show excellent agreement, with a mean bias of 0.1 ppb. The regression slope improves from 1.069 to 0.99 when XCO_2 from the CAMS greenhouse gas forecast, rather than the GGG2020 prior used in the retrieval, is used to convert retrieved CH_4 and CO_2 columns to XCH_4.

In the conclusions:

> MethaneSAT agrees within 0.1 ppb averaged across all colocations. The regression slope improves from 1.069 to 0.99 when XCO_2 from the CAMS greenhouse gas forecast, rather than the GGG2020 prior used in the retrieval, is used to convert the columns of CH_4 and CO_2 retrieved to XCH_4. This improvement is consistent with GGG2020 that represents latitudinal background CO_2 rather than local XCO_2 influenced by nearby emissions and transport.

RC: ***3. L106-108, dataset selection. Please clarify how the 340 scenes were selected from the 1160 processed to L2, and ideally show how they compare to the full dataset in albedo/SZA/region, so readers can judge how representative the validation statistics are.***

AR: We have clarified in the text how the 340 scenes were prioritized:

> These scenes were prioritized for reprocessing based on low cloud cover. Priority was also given for scene selected as calibration targets or as strong candidates for emissions quantification.

Many of the excluded scenes are a result of severe cloud cover, due to the the low weight of cloud forecast given in the target scheduler in the early phase of the mission. An earlier version of the dataset was originally used to derive the homogeneous validation scene subset in 3.3. We used this to prioritize those granules in the reprocessing. The remaining scenes were selected by the L4 team, based on emissions inversions from the previous processing that were of the most scientific interest.

Below I am showing the PDFs comparing the distribution of SZA, Albedo and latitude for the full dataset and the reprocessed subset. You can see that the distributions between them are quite similar. The reprocessed scene tends to have a few more scenes with higher SZA, likely a result of excluding fully cloudy scenes in the tropics.

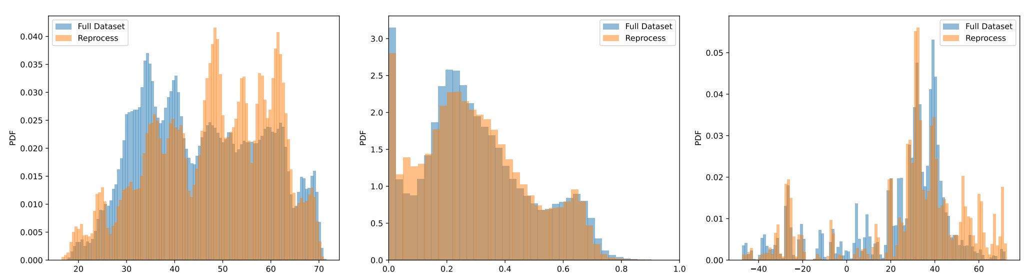

Figure 1: Probability densities of solar zenith angle, albedo, and latitude for the full (blue) and reprocessed (orange) MethaneSAT datasets.

RC: ***4. Section 3.2, prior covariance tuning.*** γ_{CO2} ***= 10 is selected to give "near unity PBL averaging kernel responses" in Fig. 2, a reasonable but qualitative criterion. A brief comment on sensitivity of the reported precision/bias numbers to this choice would strengthen the section.***

AR: The $\gamma_{CH_4} = 1$ prior already produced near-unity PBL averaging kernel response, so we tuned γ_{CO_2} to match it. This ensures that a light-path enhancement affects both the retrieved columns by a comparable magnitude, and thus largely cancels out in the proxy ratio.

The sensitivity of the precision and bias to this choice is quantified in the discussion immediately following the figure: Scaling $\mathbf{S_{CO_2}}$ degrades single-pixel XCH_4 precision by only 4.9 ppb for a typical scene, while reducing the albedo-dependent bias from ~ 5 ppb to below 1 ppb under a representative light path perturbation. The value was selected based on a coarse parameter search as sufficient to meet mission requirements. We agree that we could further refine γ_{CO_2} through a more exhaustive optimization, but given the modest sensitivity of precision to this parameter, we do not expect the main results in section 4.1 and 4.2 to significantly change by further tuning.

RC: ***5. Section 3 ordering. The DoFS-based quality filter in 3.1 depends on the retrieval's averaging kernels, which in turn depend on the prior covariance tuning introduced afterward in 3.2. Consider reordering for clarity.***

AR: We agree, and have reordered the manuscript so that the prior covariance tuning section is before the quality filtering section.

RC: ***6. Eq. 13-14, Figure 4. The precision estimator assumes "natural XCH4 variability occurs over spatial scales larger than a few pixels" (L270), which may not hold near plume edges. Since this feeds into homogeneous-scene selection and the stripe correction training set, a brief comment on sensitivity here would be useful.***

AR: Each cross-track position has about 600 along-track pixels, so pixels intersecting a plume edge make up a small fraction of the sample used in Eq. 13, limiting their impact on the standard deviation. We also manually reviewed the candidate scenes and did not find it to be a problem in practice. Figure 4 shows an example scene containing a large plume to the left of the image, and the precision estimator does not show any obvious sensitivity to it. Finally the filter is based on the relative difference between $\sigma^{\text{pix}}_{X_{\text{CH}_4}}(x)$ and $\sigma^{\Delta_t}_{X_{\text{CH}_4}}(x)$. Since a plume would increase both terms, with the pixel-level standard deviation affected at least as much, this relative comparison should remain robust even for scenes with more plume contamination than shown here.

RC: ***7. Section 3.4, destriping. PLS alone reduces*** σ_b ***to 6.1 ppb; adding the wavelet step pushes it below the 3.5 ppb theoretical noise floor, which the text attributes to the band-pass filter suppressing some real Fourier modes along with the stripe noise. A wavelet-only run (without the PLS step) would help clarify how much each stage of the correction actually contributes to the final result.***

AR: The plot below shows Figure 8 of the paper with the PDF of scenes destriped with only the wavelet filter. The distribution of stripe-noise the wavelet-only method closely matches the case with the data that have been destriped with PLS followed by the wavelet filter. This is expected, since σ_b is derived from cross-track finite differences, making it most sensitive to the pixel-to-pixel stripe structure, which is exactly what the wavelet filter is designed to remove whether or not the PLS has already been applied.

However we don't think this result implies PLS is unnecessary. The wavelet filter, by construction, only modifies the horizontal detail coefficients The coarse approximation coefficient is left untouched at every decomposition level, to avoid supressing genuine large-scale XCH_4 gradients within the scene. Consequently, a smooth, low cross-track frequency bias trend, e.g. due to the CO_2 window ISRF narrowing toward the FPA

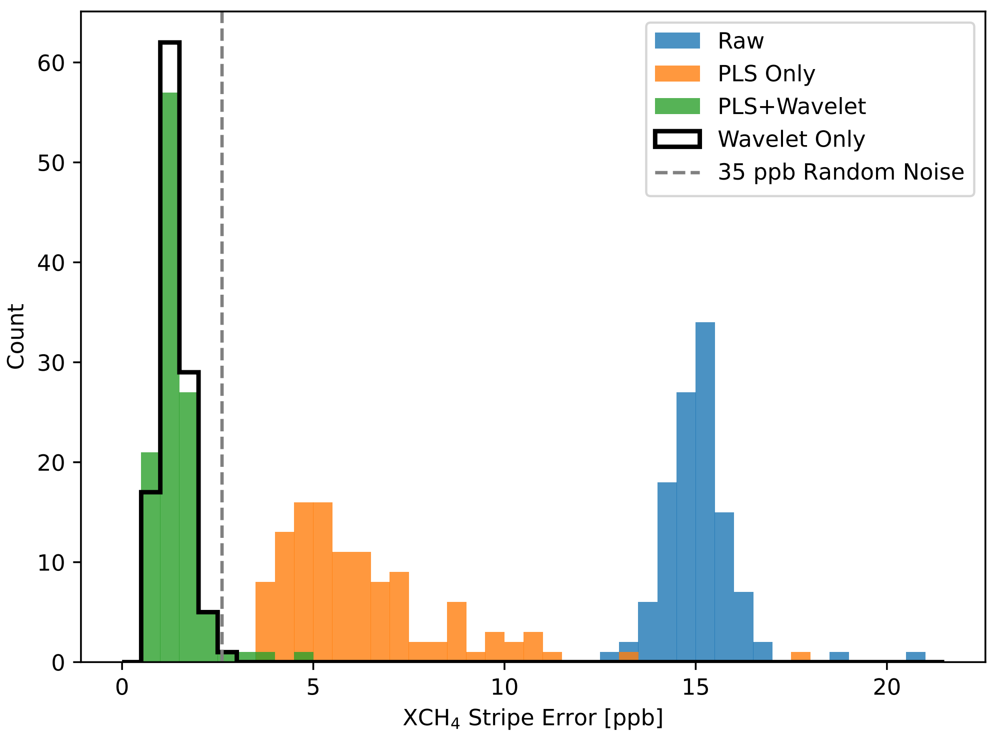

Figure 2: Reproduction of Figure 8 from the main text, with the wavelet-only case added (black line)

edges documented in Section 3.4, would be encoded in this approximation coefficient and pass through the wavelet step unchanged, independent on the methods hyperparameters L and σ_{f_n}. Since σ_b is insensitive to this broad trend, the above wavelet-only test is not useful for ruling out this failure mode, and we currently do not have a metric that can isolate this directly. PLS uses an independent physical predictor (ISRF squeeze) rather than spatial frequency separation and is therefore not subject to this limitation. We have added a note on this point in the supplement (Section S2)

> We note that this tuning exercise, like the wavelet filter itself, targets fine-scale, pixel-to-pixel stripe variation. A smooth, low cross-track frequency bias trend would instead be captured by the coarse approximation coefficient $a_{L,m,n}$, which is never filtered at any L., and would therefore pass through the wavelet step unchanged regardless of tuning. The PLS correction, which uses an independent physical predictor rather than spatial frequency separation, is not subject to this limitation, and we consider it complementary to, rather than redundant with the wavelet step.

RC: ***8. Section 4, independent validation.The TROPOMI and L1/L2 self-consistency checks are useful but not fully independent of the retrieval framework. Was a comparison against controlled methane release data considered?***

AR: We had planned to participate in the Stanford 2025 multi-satellite controlled release campaign (Reuland et al., 2025). However the satellite was in safe mode during the experiment, so we have no data available for validation.

We also note that the main aim of MethaneSAT is to constrain CH_4 fluxes at the km scale. To our knowledge, there is not yet an appropriate controlled release experiment for validating an observing system's ability to constrain diffuse fluxes of this kind. In light of this, we are currently limited to comparisons against other satellites capable of measuring absolute XCH_4, where we can compare the small enhancements mapped by both.

RC: ***9. TCCON comparison. Since GINPUT/GGG2020 (used as the CO2 prior) is itself developed for TCCON (L130-131), a direct MethaneSAT–TCCON comparison could help clarify the prior-sensitivity question in point 2, where coverage allows.***

AR: We appreciate the suggestion, but we suspect it may offer limited additional information for two reasons. We only have 40 scenes with coincident measurements in the full dataset, and only 6 in the reprocessed one. TCCON also carries a significant processing lag, and data were largely unavailable at the time we were analyzing the MethaneSAT observations for the manuscript.

GGG2020 is designed to represent latitudinal background XCO_2 rather than the exact concentration at the target location. Whilst this difference does not induce significant error in the retrieved columns, we showed using it to rescale the CH_4/CO_2 column ratio to XCH_4 does cause a discrepancy.

RC: ***10. Section 4.3, aerosol dependence. The Libya PICS analysis covers one aerosol regime (bright surface, mineral dust). It would help to know whether targets with other aerosol types are available to test the robustness of the CO2-proxy approach more broadly.***

AR: We focus on mineral dust because it is the dominant source of aerosol-induced bias in the retrieval globally. This dominance arises from its comparatively large particle size, yielding a large size parameter in the SWIR and correspondingly a higher extinction efficiency. Most other common aerosol types are smaller, with extinction efficiency dropping off rapidly towards the SWIR. Dust is even more relevant to MethaneSAT, because its target deck is heavily weighted towards oil and gas basins in North Africa and the Middle East, and thus towards bright scenes influenced by dust. The Libya site further allowed us to attribute XCH_4

gradients to aerosol effects with some confidence, since it is far from known CH_4 sources and thus XCH_4 can be assumed approximately uniform a priori.

Hygroscopic aerosols (e.g. sea salt, sulfate) are likely the other major source of significant scattering in the SWIR, as growth under high RH can increase their effective size to values high enough to impact the retrieval. Testing this regime is more difficult however, as their targets are typically located over scenes with other CH_4 sources, and thus we cannot cleanly attribute retrieved XCH_4 gradients to aerosol effects vs. real variability. We therefore consider the dust case examined to be the most critical to establish the retrieval robustness first.

RC: ***11. Data availability. Given the paper's role as a mission-performance demonstration, consider releasing the 340-scene validation subset (with homogeneous-scene flags) as a standalone dataset alongside the gated access form.***

AR: We appreciate the suggestion. We would like the clarify that access to the dataset is already effectively unrestricted. All requests submitted through the access form to date have been approved. The data are hosted on google cloud, and downloading them only incurs the cloud providers data egress cost. Given the size of the 340 scene subset (), we would prefer not to duplicate it via a separate hosted release.

We note, however, that the bias correction presented in this paper was an intermediate refinement, developed using this validation subset, that was not itself propagated to the standard access system. The data currently retrievable for these collection IDs instead reflect either an earlier bias correction (predating this paper) or a more recent one from a subsequent full reprocessing, which incorporates a further update (a plume mask applied prior to wavelet filtering) on top of the correction presented here. As a result, neither version currently obtainable through the standard access process will exactly reproduce the bias-corrected XCH_4 values shown in this paper, although we expect the differences to be modest. We will provide the list of collection identifiers and homogeneous-scene flags used in this analysis as supplementary information; readers wishing to obtain the exact bias-corrected XCH_4 values used in this paper may contact the authors directly.

3. Minor

RC: ***Eq. 6: briefly define h in words on first use.***

I think I did but the wording was not very clear. I have updated the sentence introducing equation 6 from

> To propagate this measurement error into a proxy precision for XCH_4, we compute the Jacobian of XCH_4 with respect to the state vector ($\mathbf{h}$):

to:

> To propagate this measurement error into a proxy precision for XCH_4, we compute $\mathbf{h}$, the Jacobian of XCH_4 with respect to the state vector $\mathbf{x}$:

RC: ***L195-199: clarify whether the averaging kernels in Fig. 2 include the full state vector or just CH4/CO2 layers.***

AR: Right now it is just computed from the CO2 and CH4 layers (similar to the formulation used by OCO2. At some point I had a code branch with a version of the calculation using full state vector but it didn't make much difference for the retrieval setup presented here. I have updated the text to clarify:

Here we consider only changes to $\mathbf{n}_\mathrm{g}$ caused by the CO_2 and CH_4 state sub-vectors. This differs from the treatment of measurement noise (Equation 8) which propagates uncertainty through the full state vector $\mathbf{x}$. We find that interference from other state vector elements is small, supporting this approximation.

RC: ***Fig. 12: correlation coefficients per scene range 0.46-0.86; worth stating whether cloud-contaminated scenes are excluded from any summary statistics or only flagged qualitatively.***

AR: We exclude all scenes that are filtered by each algorithms respective quality flagging scheme. We have updated the caption to reflect this.

Pixels flagged for poor quality by either algorithm, including for cloud contamination, are excluded in both the maps and correlation coefficient calculation.

References

Reuland, F., Adams, T., Galvin, K., Kort, E. A., and Brandt, A. R.: Large-Scale Controlled Methane Releases for Satellite-Based Detection and Emission Quantification of Point-Sources, , 2025.