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: ***This study presents the first operational CH4 retrievals from MethaneSAT using the CO2-proxy method. It assesses the performance of the retrieval algorithm against mission requirements and provides an explicit analysis of several potential sources of systematic error. The paper represents an important contribution to the field and is well suited to the scope of Atmospheric Measurement Techniques. It can be published once the following points have been adequately addressed and clarified.***

AR: We thank the reviewer for their assessment of the paper and for their detailed and thought-provoking comments, which we address individually below.

2. Specific comments

2.1. Section 3.1

RC: ***Quality filtering is based on the O2 band at 1.2μm. Are dayglow emissions (adding radiance that is not reflected sunlight) somehow considered/modelled to achieve the required accuracy needed for quality filtering? Isn't dayglow a problem for dark surfaces and/or high solar zenith angles? Please elaborate a bit more on the topic.***

AR: Not accounting for oxygen airglow in the 1.2 micron band would indeed by a significant source of error. We currently account for it in the surface pressure retrieval by fitting a column scaling factor. The airglow volume emission rates are modeled following Sun et al. (2018), using line parameters from the GGG2020 spectral database (to be consistent with the linelists used to model absorption). The airglow priors use a parameterization based on SCIAMACHY airglow retrievals (Sun et al., 2022). In general we do find strong correlations with retrieved surface pressure differences over land. Sun et al. (2018) also evaluated how separable the O2 airglow and absorption features were, and found that at the resolution of MSAT they can be retrieved simultaneously.

We have found the dark surface over the ocean useful for validating the airglow retrieval and prior, as it is a significant fraction of the signal. Since the airglow variation is low over the scale of a MSAT scene, we have considered using ocean pixels as a means to better constrain the pixels with higher surface albedos.

More details will be available in a manuscript on the O2 band retrieval to be submitted within the next few months. Unfortunately, it was not ready in time for this publication so I was unable to cite it.

RC: ***Is $\triangle$XCO2 really needed as an additional quality criterion in a proxy retrieval? The assumption is that CH4 and CO2 are spectrally close and therefore react in virtually the same way to changes in path length, so that the effect becomes negligible in the ratio.***

AR: The issue is that there are some pixels in the 1.6 micron band that do not have an O2 band observation: These occur at the along-track edges due to a timing issue, and at a cross-track index at the center of the scene where there is a missing cross-track in the O2 band. We use fractional changes in the CO2 column (units molecules/area) relative to the prior as a backstop to essentially fill in these gaps. Since the CO2 column is known a priori within a 1%, the absolute columns can be used for this purpose. We reduce the a priori CO2 biases below this by comparing the effective air columns inferred from CO2 and O2 where their pixels overlap.

RC: ***What is the $\triangle$XCO2 threshold used for quality filtering?***

AR: We exclude fractional changes in CO2 column relative to the prior above 2%. This represents 20 hPa for a 1000 hPa surface pressure, making the threshold approximately consistent with the O2 Band filter.

2.2. Section 3.2

RC: ***The argument not to use profile scaling is not fully convincing. I think a retrieval response exceeding unity is no problem, because such behaviour is fully characterised by the averaging kernels, which are specifically provided to describe the retrieval sensitivity and allow proper interpretation of the results. It is true that a scaling factor retrieval is structurally incapable of representing changes in vertical shape. But as written in the text (L157), the retrieved profiles are highly sensitive to the assumed a priori covariance. This suggests that much of the apparent flexibility is supplied by the assumed covariance model rather than by information contained in the measurements. As a result, it is not evident that profile retrievals are inherently superior to profile scaling; they simply replace one explicit assumption (fixed profile shape) with another (assumed covariance). Please elaborate a bit more on this trade-off (between minimising assumptions about the profile shape and minimising dependence on an assumed covariance model) in the manuscript. How much independent vertical information (DoFS) do the measurements actually contain on average (quality cut-off is at 0.6 according to L138)?***

AR: We agree that a non-unity CH_4 averaging kernel response is not by itself a problem: given a candidate true CH_4 field (e.g. in a flux inversion), the averaging kernel can be applied directly to produce the equivalent retrieved field for proper comparison. This is standard practice and accounted for in the AK formalism, as the reviewer notes.

However the proxy method also relies on the CO_2 averaging kernel to cancel the effect of light path modifications (e.g. from aerosol scattering) in the CH_4/CO_2 ratio. Unlike a CH_4 enhancement, a light-path perturbation has no corresponding "true field" that can be modeled and corrected through the averaging kernel after the fact; it is an unknown nuisance signal, not a quantity of interest. Cancellation of this this nuisance signal in the ratio is only possible if the CO_2 and CH_4 averaging kernels respond similarly to it. If they do not, the light-path effect leaves a residual bias in XCH_4 that cannot be corrected using the averaging kernel alone, unlike the CH_4 response to a genuine enhancement - It provides a tunable degree of freedom, via the layer correlation structure, that allows us to bring the CO_2 and CH_4 averaging kernels into closer alignment. From earlier testing, the averaging kernels of the tuned GOSAT covariance matrices used here were closer compared the those of column scalings.

We would also like to clarify our use of DoFS and the prior covariance in this context, as we agree the framing in the text could invite confusion. We do not rely anywhere in this work on a Bayesian interpretation of the posterior (e.g. using the posteriori covariance as a genuine uncertainty estimate). The two quantities we use, the averaging kernel and measurement error propagated precision (Eq 13-14) are deterministic functions of the retrieval gain and Jacobians, and are valid description of the linearized retrievals actual behaviour regardless of whether $\mathbf{S_a}$ represents true atmospheric variability. DoFS by contrast, is more naturally read as an information content statement, and here the choice of $\mathbf{S_a}$ matters. The median DoFS in our dataset (subject post quality filtering, including a DoFS cutoff of 0.6 for both CO_2 and CH_4) is 1.04 for CH_4 and 0.98 for CO_2 (90% range 0.88-1.14 and 0.77-1.08 respectively), but these values are a consequence of our covariance tuning rather than a fixed, measurement only property. Given the single pixel SNR, $\mathbf{S_a}$ must be relaxed well beyond physically realistic atmospheric variability simple to allow individual pixel retrievals to be informed by the measurements (a physically realistic prior would yield near-zero DoFS at the single pixel level). An alternative would be to aggregate pixels prior to retrieval to improve SNR, but along with sacrificing the native resolution of the instrument could also create its own systematic errors (e.g. in ISRF modeling of aggregate pixels, impacts of inhomogeneous illumination and increased susceptibility to cloud contamination). We therefore consider our relaxed, tuned covariance a necessary design choice.

Finally, we note that the relative correlation structure of the prior covariance (as opposed to the overall scaling γ) is inherited from the University of Leicester GOSAT CO_2-Proxy retrieval, and has not been further optimized in this work. We don't believe this undermines our conclusions - The γ tuning presented in Section 3.2 demonstrates empirically that, under this inherited structure, the light-path perturbations consistent with our quality filtering thresholds do not induce XCH_4 errors exceeding our accuracy requirements.

RC: ***This concern seems to be particularly relevant when the prior covariance is parameterised by adjusting only scalar scaling factors in the tuning process, while the covariance structure is fixed (Equation 3). Doesn't changing the scaling factors simply alter the balance between the measurement and the prior? How sensitive are the retrieval results to the choice of γ_{ch4} and γ_{co2}?***

AR: Yes that is correct: adjusting γ changes the balance between measurement and prior, and we do not intend this as anything other than a deliberate regularization choice, as discussed in our response to the previous comment. Our reason for adjusting γ_{CO_2} specifically (rather than γ_{CH_4} is that the unscaled ($\gamma_{CH_4} = 1$) prior already produces near-unity CH_4 averaging kernel response across the range of conditions sampled by MethaneSAT, so no adjustment was needed to bring it in line with the CO_2 response.

The sensitivity of precision and bias to both γ_{CH_4} and γ_{CO_2} is already shown in Figure 2 and discussed in Section 3.2. We believe this directly answers the question, and have therefore left this text unchanged.

2.3. Section 3.4

RC: ***The stripe error of the PLS+Wavelet combined destriping approach falls below the level expected from purely random, normally distributed noise. This is attributed to the suppression of Fourier modes due to the Wavelet destriping (L424). Although this may have a small effect, I do not think that significant real features are suppressed, provided that σ_f has been chosen appropriately. I would be more inclined to suspect that the 'over-correction' is more related to the previous PLS regression. Wouldn't a wavelet correction alone be enough? Have you tried to examine what the result would look like? A wavelet-only correction would also have the advantage that the PLS-calibration on homogeneous scenes would not be required, which is an additional source of uncertainty.***

AR: We suspect part of the concern here may stem from an ambiguity of how we described the PLS step, which could give the impression that it produces a per-pixel correction using per-pixel ISRF information, similar to the wavelet step. To clarify: The PLS model predicts a single, along-track invariant cross-track bias pattern for each scene, taking along-track median ISRF squeeze factors as inputs to the PLS model. The resulting pattern is subtracted uniformly from each row. It is not a per-pixel correction, and does not use per-pixel ISRF information. We have added a sentence clarifying this:

> As with $b(x)$, the predicted correction is a single value per cross-track index. In operational use this is calculated using the along-track median ISRF squeeze factors and subtracted uniformly from every row.

Regarding whether the wavelet-only correction would be sufficient: We did test this, and found the resulting σ_b distribution to be similar to the PLS+wavelet (see reviewer 1 response). However, σ_b is derived from cross-track finite differences, making it primarily sensitive to fine-scale, pixel-to-pixel fine structure. It is not well suited to to detect smooth, low cross-track frequency bias trends, which in the wavelet decomposition are encoded in the aproximation coefficients rather than the filtered horizontal detail coefficients, and therefore pass through the wavelet filter unchanged regardless of tuning. We believe such trends are physically plausible in our data (e.g. the CO_2-window ISRF narrowing towards the FPA edges, Section 3.4). Since we do not have a metric to directly isolate them, we cannot rule out their presence by the wavelet-only method reducing σ_b sufficiently.

More fundamentally, even a well tuned wavelet filter cannot distinguish a smooth cross-track bias trend from a real smooth XCH_4 gradient, since both present identically in the spatial frequency domain. This is why σ_{f_n} and L must be tuned as conservatively as possible to avoid supressing real structure (Section S2), which limits how aggressively a purely frequency-based method can address this kind of error. PLS avoids this ambiguity bceause it is based on an independent physical predictor, uncorrelated with true atmospheric XCH_4, rather than spatial separation within the XCH_4 field itself.

We agree that the PLS calibration introduces an additional source of uncertainty tied to the representativeness of the homogeneous validation scenes, and have discussed this in Section 2.1. We consider the trade-off as acceptable, given the complementary failure modes between the two methods.

2.4. Section 4.1

RC: ***Why is the precision on L447 reported on a 2°×2° grid and not on 5°×5° as the other estimate and the mission requirement?***

AR: I meant to convey that we meet the prelaunch precision requirement at a much higher spatial resolution (2x2 km^2). I have changed the line to state this more clearly:

> 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.

2.5. Section 4.2

RC: ***L452: As there are at least three different TROPOMI products, please state explicitly "... operational TROPOMI XCH_4 retrievals of version ...".***

AR: We now specify its the operational product and quote the exact processor versions:

> We compare MethaneSAT to the operational TROPOMI XCH_4 retrievals produced by SRON/ESA,

using processor versions 2.7.1 and 2.8.0 (from 8 September 2024). We restrict the intercomparison to this version as earlier versions did not include a stripe correction.

RC: ***L455: Is there a reference that TROPOMI exhibits larger systematic biases than GOSAT? I think, concerning station-to-station bias (stddev of the individual biases at TCCON sites), both instruments are very similar and close to the TCCON accuracy.***

AR: We do not have a single reference that states this claim in these exact terms, but it is discussed in related literature. In their recent review of satellite methane observations, Jacob et al. (2022) note that the reliance on TCCON to diagnose variable bias is limited by network sparsity (concentrated in northern midlatitudes), and instead reference GOSAT as a well-calibrated alternative, citing a variable bias of 2.9 ppb relative to TCCON for the University of Leicester CO_2-proxy retrieval (Parker et al., 2020). They further show TROPOMI biases relative to GOSAT can exceed 20 ppb in some regions, which they attribute to TROPOMIs coarser spectral resolution, and unavailability of the proxy method in the 2.3 micron band. They ultimately recommend comparing to TROPOMI to GOSAT for the region of interest before interpreting TROPOMI data there (Chen et al., 2022). This is consistent with Balasus et al. (2023), who use GOSAT to bias-correct TROPOMI XCH_4 and find the largest corrections occur over deserts. In such regions TCCON observations are unavailable, so even though there are large differences regionally, they note that the change in TCCON bias in the corrected product is small.

RC: ***L459: The station-to-station bias (5.1 ppb in Lorente et al., 2023) is the much more important measure than the mean bias (-5.3 ppb in Lorente et al., 2023). I do not find a -5.6 ppb in the cited paper.***

AR: We agree that the mean bias was slightly misquoted. In light of the reviewers suggestion, we now report the station-to-station bias (5.1 ppb), as we agree it is a more relevant measure of spatial bias.

> The station-to-station variability of the TROPOMI bias relative to TCCON, a measure of spatial bias, is most recently reported as 5.1 ppb (Lorente et al., 2023).

RC: ***Figure 11: The MethaneSAT retrievals differ quite noticeably in detail for the two priors used. Just how different are the two prior data sets from one another?***

AR: Given the sensitivity of XCH_4 to the XCO_2 used for scaling (~ 4.5 ppb error per ppm XCO_2 error) the ~ 6.6 ppb reduction in the Southern Hemisphere bias between GGG2020 and CAMS based conversions in Figure 11 corresponds to a mean difference of only $\sim$1.5 ppm XCO_2 fields in this region. This illustrates that a relatively modest, spatially coherent difference between the two datasets is sufficient to produce the shift shown in Figure 11. We expect differences of at least this order due GGG2020s design to only simulate background XCO_2 levels.

RC: ***L491: You write that CAMS likely cannot properly capture the true subgrid XCO2 variations at the scale of a MethaneSAT target. But can GGG2020 actually resolve those variations? Or actually not (better not at all than wrong)? Isn't it a general problem if the prior sensitivity is so high that the result depends on this capability? Or are you not talking about the prior of the CO2 retrieval, but rather the modelled XCO2 scaling as used in Equation 1? Please clarify in the text what is meant by 'prior' in each instance, as modelled CO2 is incorporated into the XCH4 retrieval in two different ways.***

AR: Yes the reviewers framing is correct: GGG2020 cannot resolve subgrid XCO_2 variations at all, and this is by design, as it is intended to represent only smooth latitudinal background CO_2. As the reviewer suggests, this is a case where representing no spatial subgrid structure is better than representing it incorrectly, as this will only add another source of subgrid bias to the XCH_4 field.

To directly answer the second part of the comment: this section refers to the XCO_2 value used in the proxy conversion from the retrieved columns to XCH_4, not the CO_2 prior used within the retrieval itself. We agree this distinction, and the terminology used to describe it, needs to be made explicit throughout. We address this in the last comment.

RC: ***Figure 12: The colour scales for the second and third column seem to be saturated; please increase the spanned value range appropriately. Why isn't the same colour scale used for all three columns?***

AR: The three columns are intentionally shown with different color scales for a specific reason. The left column shows native-resolution MethaneSAT data, where single-pixel noise is large; a color scale wide enough to avoid saturation here would be too broad to show meaningful structure in the lower resolution columns. The center and right columns compare MethaneSAT and TROPOMI at a common spatial resolution, and share an identical, saturated color scale to enable a direct visual comparison between the instruments. The scale is set to the 5th-95th percentile range of the concatenated MethaneSAT and TROPOMI data shown in the two panels, rather than chosen manually, so the degree of saturation is applied consistently and is not tuned to enhance the apparent agreement between the two instruments. The color bar triangular caps indicate the scale is saturated (this is the default way matplotlib does it). We can clarify this rationale explicitly if the reviewer feels this is not sufficiently clear.

2.6. Section 4.3

RC: ***Figure 13: The proxy approach seems to work well for low-level clouds. But which scenes would the quality filter described in Section 3.1 have filtered out in the map anyway? Would there be any problematic scenes left at all? What about increases in methane that may be partially masked by clouds? (XCH4 [CO2 proxy] cannot see below the clouds)***

AR: Figure 13 intentionally shows all retrieved data in the scene, without applying the standard quality filters in Section 3, specifically to illustrate the light-path effects of both dust and aerosols before filtering. Applying the cloud filter ($|O_2\Delta p| > 20$ hPa) would remove the pixels directly over cloud, which show large negative $O_2\Delta p$ from direct reflection (note colorbar saturation at 5-95th percentiles), but retain most of the surrounding clear sky pixels. This includes those adjacent to the clouds showing evidence of light-path lengthening, since these remain within the filter threshold.

Regarding CH_4 enhancements obscured by clouds: these are robustly caught by the $O_2\Delta p$ and would thus appear as data gaps. We do not claim to be able to retrieve these.

RC: ***L536: Is it really dust, it looks more like small patchy clouds? A similar figure for a strong desert dust aerosol event would also be interesting. The scene in Figure 15 also looks quite clean regarding aerosols.***

AR: We agree that a stronger dust event will be of interest, but note that our quality filter, which rejects pixels with $|O_2\Delta p| > 20$ hPa is designed to reject scenes where light path effects from clouds and dust are too large to be reliably corrected. Since this filter is intended to bound the conditions under which we claim the proxy method meets mission requirements, evaluating cases within this range, rather than strong events that fall outside it, is a more relevant test of robustness for our purposes. Figure 14 extends this assesment across all 5 Libyan calibration targets currently processed, one of which (11-September-2024) shows larger XCH_4 variability with $O_2\Delta p$.

Regarding dust, the scene does show evidence of a mild background dust loading, visible in the lower left corner where a decrease in albedo coincides with a decrease in retrieved $O_2\Delta p$. This is consistent with light-path shortening from backscatter over a darker surface.

MethaneSAT does observe stronger dust events over oil and gas production regions, though these are similarly subject to the quality filter, and the strongest events are excluded for the same reason. The remaining more moderate cases can be complicated by separate genuine XCH_4 variation from real enhancements, which makes it difficult to cleanly separate real XCH_4 structure from that induced from dust. Libya4 was chosen because it lacks local CH_4 sources to avoid this problem.

2.7. Section 4.4

RC: ***Also here, please make clearer what is conceptually meant by prior. This section is about the modelled CO2 scaling 1:1 the retrieved XCH4 according to Equation 1 (and not about the prior for the CO2 retrieval itself), right? It would be helpful to use two different terms throughout the entire manuscript to distinguish between the two applications currently referred to as prior, even if the same data set is used.***

AR: We agree, and will revise the manuscript to consistently distinguish between the two quantities both referenced as "prior". That is, (1) The retrieval prior used when determining the CO_2 and CH_4 vertical columns, and (2) The XCO_2 proxy value used to convert the columns to XCH_4. In the current operational product, both are derived using GGG2020. We only use CAMS to convert the columns to XCH_4. We will revise the text throughout to be clear each time what GGG2020 and CAMS are being used for.

3. Technical corrections

RC: ***L90: the the***

AR: Fixed in updated manuscript

References

Balasus, N., Jacob, D. J., Lorente, A., Maasakkers, J. D., Parker, R. J., Boesch, H., Chen, Z., Kelp, M. M., Nesser, H., and Varon, D. J.: A blended TROPOMI+GOSAT satellite data product for atmospheric methane using machine learning to correct retrieval biases, Atmospheric Measurement Techniques, 16, 3787–3807, , 2023.

Chen, Z., Jacob, D. J., Nesser, H., Sulprizio, M. P., Lorente, A., Varon, D. J., Lu, X., Shen, L., Qu, Z., Penn, E., and Yu, X.: Methane emissions from China: a high-resolution inversion of TROPOMI satellite observations, Atmospheric Chemistry and Physics, 22, 10 809–10 826, , 2022.

Jacob, D. J., Varon, D. J., Cusworth, D. H., Dennison, P. E., Frankenberg, C., Gautam, R., Guanter, L., Kelley, J., McKeever, J., Ott, L. E., Poulter, B., Qu, Z., Thorpe, A. K., Worden, J. R., and Duren, R. M.: Quantifying methane emissions from the global scale down to point sources using satellite observations of atmospheric methane, Atmospheric Chemistry and Physics, 22, 9617–9646, , 2022.

Lorente, A., Borsdorff, T., Martinez-Velarte, M. C., and Landgraf, J.: Accounting for surface reflectance spectral features in TROPOMI methane retrievals, Atmospheric Measurement Techniques, 16, 1597–1608, , 2023.

Parker, R. J., Webb, A., Boesch, H., Somkuti, P., Barrio Guillo, R., Di Noia, A., Kalaitzi, N., Anand, J. S., Bergamaschi, P., Chevallier, F., Palmer, P. I., Feng, L., Deutscher, N. M., Feist, D. G., Griffith, D. W. T., Hase, F., Kivi, R., Morino, I., Notholt, J., Oh, Y.-S., Ohyama, H., Petri, C., Pollard, D. F., Roehl, C., Sha, M. K., Shiomi, K., Strong, K., Sussmann, R., Té, Y., Velazco, V. A., Warneke, T., Wennberg, P. O., and Wunch, D.: A decade of GOSAT Proxy satellite CH_4 observations, Earth System Science Data, 12, 3383–3412, , 2020.

Sun, K., Gordon, I. E., Sioris, C. E., Liu, X., Chance, K., and Wofsy, S. C.: Reevaluating the Use of O2 a1g Band in Spaceborne Remote Sensing of Greenhouse Gases, Geophysical Research Letters, 45, 5779–5787, , 2018.

Sun, K., Yousefi, M., Chan Miller, C., Chance, K., González Abad, G., Gordon, I. E., Liu, X., O'Sullivan, E., Sioris, C. E., and Wofsy, S. C.: An optimal estimation-based retrieval of upper atmospheric oxygen airglow and temperature from SCIAMACHY limb observations, Atmospheric Measurement Techniques, 15, 3721–3745, , 2022.