

Reply to reviewer 1

We thank the reviewer for this comprehensive review. We insert our replies (indicated in blue) to each comment below.

General comments

This study presents an innovative method for estimating total column methane (XCH_4) using a Lagrangian Particle Dispersion Model (LPDM), FLEXPART, and a way to assimilate the data in an atmospheric inverse model, FLEXINVERT. The case study is carried out for Siberia for 2022. The comparison against the “traditional” ground-based inversion showed broad agreement with the inversion using TROPOMI data, and consequently reliability and a good potential in the presented method for estimation of regional CH_4 fluxes. The method sounds applicable for other LPDMs, and could contribute significantly to the atmospheric inverse modelling community with new ways to infer greenhouse gas flux information from satellite data. The fact that LPDMs can be run in much higher resolution than Eulerian transport models will be an advantage for incorporating information from future satellites with much higher spatial resolution. Therefore, this paper is worth of prompt publication after considering a few points below.

- The authors found large differences in modelled XCH_4 depending on background initial mixing ratios. The boundary conditions were optimised to somewhat discriminate the “errors”, but how can it be sure that the signals within the domain is not over constrained by the background? In L280, it is said that the correction of modelled XCH_4 was “largely due to the improvement to the background estimate”, and the posterior fluxes from the TROPOMI inversion did not change much from the prior. I suppose that with less uncertainty in the background, the fluxes would change more. How do you know what is a good balance?

Indeed, with more accurate background XCH_4 estimates smaller uncertainties can be used for the scalars optimizing the initial mixing ratios, and thus the model-observation differences will have a stronger impact on the posterior fluxes.

The optimization of the initial mixing ratios is made by optimizing scalars, which are defined for boxes of the atmosphere (for a given number of latitudinal bands and vertical layers). Thus many retrievals in different parts of the domain will contribute to constraining the same scalars, because the BRR (background receptor relationship) is spread out over different latitudes and vertical layers. On the other hand, the SRR (source receptor relationship) is strongest in the vicinity of the retrieval. Thus there is a different constraint on the initial mixing ratios versus the fluxes. In this way, if accurate background estimates are obtained, even if large uncertainties are assigned to the scalars of initial mixing ratio these should not differ significantly from the prior value of one, and the model-observation differences will more strongly constrain the fluxes. In this way, it is not so important how large the uncertainties on the scalars of initial mixing ratio are, as long as they are sufficiently large to correct a wrong background. Please also see our reply to the comment on L209-213.

- High northern latitudes are challenging regions such that satellite retrievals are associated with various biases, especially those related to seasonal variations may play a significant role. I wonder how much of the differences between the ground-based and TROPOMI inversions were due to these biases. Please discuss.

The posterior fluxes from the inversion using TROPOMI remained very close to the prior fluxes. This is simply because TROPOMI does not provide a strong constraint on surface fluxes at this latitude. Thus, the difference in posterior fluxes between the ground-based and TROPOMI inversions is simply due to the fact that in the TROPOMI inversion the fluxes do not significantly change with respect to the prior whereas in the ground-based inversions they do.

- The inverse model results are associated with various uncertainties. You have discussed and did sensitivity tests on background mixing ratios, but other optimization setups, such as choice of retrieval products, pre-processing methods of the satellite data, prior fluxes and prior uncertainties for observations and fluxes are also important. As this paper does not present variety of sensitivity tests, the conclusion of the paper about CH₄ flux estimates should be presented carefully that the results may change significantly depending on the setups.

We did in fact carry-out some sensitivity tests for the choice of prior uncertainties. We tested varying the prior uncertainties, and found the inversion results to be fairly insensitive to prior uncertainties chosen as 50% or 100% of the prior flux value.

For the observation space uncertainties for XCH₄, these were determined for each super-observation as the quadratic sum of the individual retrieval uncertainties. This was a larger number than the standard deviation of individual retrievals going into each super-observation. The median observation uncertainty was around 12 ppb and did not vary much between months. In addition, we consider a background uncertainty estimate. In total, the observation space uncertainties had a median value of 16 ppb and inter-quartile range of 14 – 18 ppb. We consider this to be a reasonable estimate of the uncertainty of the column average mixing ratios. Increasing the observation uncertainties would not change the result of the TROPOMI inversion, since the posterior fluxes remain very close to the prior fluxes, and increasing the observation uncertainties would bring them even closer to the prior.

We also looked at the Official XCH₄ retrieval product from SRON, however, we found that this contained many spurious XCH₄ values over our domain, which was not the case for the WFMD retrievals. Furthermore, even the bias corrected XCH₄ from the Official retrieval still contains biases due to surface albedo and aerosol size parameter (Balasus et al. 2024). Therefore, we chose to use the WFMD retrievals in our study.

Concerning the conclusions of our paper, we only state that the posterior fluxes remain close to the prior fluxes, and that there was limited uncertainty reduction. This result is

robust to the choice of prior uncertainty (50% or 100%) and would not change if larger observational uncertainties would be used, and we think smaller observational uncertainties would underestimate the uncertainty.

Specific comments

L160: Why do you use area-weighted averages? I understand it is somewhat reasonable for mixing ratios, but for averaging kernels and pressure weighting, I do not fully understand how the area would be affected. Could you also explain how did you take into account differences in number of observations within the aggregated cells?

If one weights the mixing ratios by area, for consistency it is also necessary to apply this to the averaging kernels and pressure weighting.

The analysis column average mixing ratios are defined as:

$$\hat{x} = x_{pri} + A(x - x_{pri}) \quad (1)$$

The mean of column average mixing ratios is:

$$\bar{x} = \frac{1}{S} \sum_{m=1}^M \hat{x}_m S_m \quad (2)$$

where S_m is the area of the ground pixel of the retrieval. Substituting Eq. 1 into Eq. 2 and expanding gives:

$$\bar{x} = \frac{1}{S} \sum_{m=1}^M x_{pri,m} S_m + \frac{1}{S} \sum_{m=1}^M A_m x_m S_m - \frac{1}{S} \sum_{m=1}^M A_m x_{pri,m} S_m \quad (3)$$

The super-observations are calculated as the mean of all retrievals with the centre of the retrieval falling within the defined grid cell. The number of retrievals per super-observation can of course vary. We treat the errors of the individual retrievals as uncorrelated, namely the super-observation uncertainty is calculated as the quadratic sum of the individual retrieval uncertainties.

In the averaging algorithm, we also look at the distribution of retrievals within the grid cell as defined in the current iteration. If there are not retrievals in every quarter of the grid cell, then the grid cell is any how divided in the next iteration to avoid having a super-observation for a grid cell which is not fully represented by the retrievals. We have added this extra information to Section 2.2.

Section 3: Did you include temporal correlation of the state vectors? Please add information somewhere.

The prior error covariance matrix also includes temporal correlation of prior flux errors. The temporal correlation is calculated using exponential decay with time with a correlation scale length of 28 days. We have added this information to Section 3.1.3.

L195: Could you add a figure on spatial resolution? Where were lowest and highest resolutions? Were the resolutions same for the TROPOMI and ground-based inversions, despite the fact that they would have differences in “how strongly the fluxes influence the observations” due to differences in locations and quantity (surface vs total column) of the observations? Please clarify.

We have added figures showing the grid used for the state vector for the TROPOMI and for the ground-based observation inversions to the Supplement and refer to these in Section 3.1.1. The grid was calculated separately for the TROPOMI and for the ground-based observations, since these have different coverage and sensitivity to the fluxes.

L209-213: In later sections, it is said that the background mixing ratios were also optimised. What were the uncertainties in the boundary conditions?

For the TROPOMI inversions, we used a fairly large uncertainty for the scalars of the initial mixing ratios of 5%. This uncertainty was based on a number of tests, which showed that smaller uncertainties did not give sufficient freedom to adjust the background column mixing ratios, resulting in an erroneous seasonal cycle in the posterior fluxes with very large fluxes in spring and smaller fluxes in summer, which was owing to the large positive bias in the background mixing ratios in spring and small positive bias in summer.

For the ground-based observation inversions, the background mixing ratios were also optimized, but for these inversions the background estimate was more accurate and smaller uncertainties were used, of 1%.

We now include this information in Section 3.1.1.

L218: Why did you chose the grid cell sizes of 0.25° and 0.5° ? FLEXPART is run at 0.5° and smallest optimisation spatial resolution is also 0.5° , so why did you chose to have observations at higher spatial resolutions? Can FLEXPART resolve differences well (or what is done) if there are more than one observations within a $0.5^\circ \times 0.5^\circ$ grid cell?

In the Lagrangian model, there is no finite resolution or underlying grid for the modelling of the particle trajectories. Thus, the observations can be represented (via the release of virtual particles) as a point (as is done in the case of ground-based in-situ observations) or as a volume (as in done for the satellite column observations). The $0.5^\circ \times 0.5^\circ$ resolution refers to the grid resolution on which the particles are sampled, and thus the output resolution of the SRRs. By using a finer resolution for the particle releases the atmospheric transport at and around the actual observation can be better resolved.

L223: I suppose number of observation vary a lot within the study period. Please add information about number of observations also perhaps in 14-days temporal resolution, which is your flux optimisation resolution. I would also like to see for both TROPOMI and ground-based observations.

We have included figures showing the number of observations per 14-day time interval for the TROPOMI super-observations and for the ground-based observations (see Supplementary Fig. S2). For the TROPOMI observations, we show the number per time interval as well as latitudinal intervals of 5° to indicate how the number of observations also varies by latitude. The number of observations strongly decreases with latitude, especially in March and October, as expected. There are also fewer observations from

the mid May to mid July in the latitudes north of 50°N owing to cloud cover. The number of ground-based observations also varies strongly with fewer observations from May to June owing to instrumental problems at a number of JRSATION sites.

L225-232:

- The source information for JR-STATIONS are available in Data availability section, but how about other ground-based data?

We have added the information for the other ground-based stations as well.

- Did you process/filter these data at all?

The JRSTATION data are given at a frequency of two observations per hour (except for station BRZ for which there are four observations per hour). These observations were averaged to hourly, which we now state in Section 3.1.2. These data were also corrected for a calibration scale offset of 5 ppb with respect to the WMO CH₄ X2004A scale (the other observations were already on the WMO CH₄ X2004A scale). The in-situ data at station CPA were already provided at hourly resolution. The NOAA flasks at station UUM are sampled in pairs, and these pairs were averaged. For the data obtained from the World Data Centre for Greenhouse Gases (WDCGG), which includes UUM, TER and CPA, quality flags are provided, only data indicated as “invalid” were removed. No other processing was performed.

- For the TROPOMI data, you mentioned that the observation uncertainties were 14-20 ppb. How about for these ground-based observations?

The ground-based observation uncertainties had an interquartile range of 7 to 12 ppb and median value of 8 ppb. We have now added this information to Section 3.1.2.

Section 3.1.3:

- Did all the prior fluxes had estimates for 2020? If not, what did you do? What were the original resolution of the prior fluxes? Did you do any interpolation when original resolution was lower than 0.5°?

The EDGAR-v8 (anthropogenic), LPX-Bern (peatlands, inundated, wet and mineral soils) and the GFED-v4.1s (fire) estimates were all available for the year 2020. The geological and ocean estimates were climatologies. The original resolutions are provided in Table 2. Only the emissions that had lower resolution than 0.5° were the geological ones, which were at 1.0°, these were linearly interpolated to 0.5°.

- Do I understand it correctly that you include ocean emissions, but do not optimise them? How large were the contribution of ocean fluxes to the total fluxes of this domain?

It is correct that we account for the ocean fluxes but do not optimized these. The ocean emissions amounted to 0.5 Tg/y for the domain, and represented 1.6% of the total emissions for the domain (see Table 2).

Section 3.2: I understand that you only optimise total fluxes, but as you find some spatial differences in the flux increments between the TROPOMI and ground-based inversions, can you speculate whether emissions from oil and gas sources have different seasonal patterns in the two inversions?

The differences we see between the posterior fluxes from the TROPOMI versus the ground-based observation inversions, is essentially the difference between the prior and posterior fluxes for the ground-based observation inversions, since the posterior fluxes from the TROPOMI inversions remain very close to the prior. The peak in the prior estimate in summer is due to wetland fluxes, and the increase in the summer in the posterior fluxes from the inversion using ground-based observations is likely attributable also to wetland fluxes, as it is unlikely that oil/gas emissions would increase considerably in summer.

L286, L306: Could you add uncertainty estimates as well?

We have added the uncertainties [here](#).

L295-299: Are the number of TROPOMI observations less than those from the ground-based stations? Do you argue that number of the observations was persistent for all months? I suppose flux uncertainties are larger in summer (as a whole domain) due to contribution of wetlands (although it is not so clear from Figure 6)? How would the retrieval biases possibly play a role that were discussed in e.g. Lindqvist et al. (2024)?

There are, as expected, many more TROPOMI observations than there are ground-based ones, however, in terms of the constraint on the fluxes it is not so much the number of observations but how sensitive these are to the fluxes and what uncertainty they have (see also reply to the comment below). The number of TROPOMI observations did vary somewhat from month to month (see new Figure S2 showing the number of observations per flux time step and by latitude).

The prior flux uncertainty is actually constant for each flux time step in the inversion. The prior flux uncertainty is determined as a fraction of the mean flux over all time steps, with a minimum uncertainty equal to approximately the 10th percentile flux value considering all fluxes (averaged in time) over the domain. This gives equal freedom to adjust the fluxes in all seasons.

Concerning retrieval biases, the full-physics retrieval algorithm, RemoTec, which is used to provide the Official ESA retrieval product, is known to be affected by artefacts due to albedo and aerosol size parameter (e.g. Balasus et al. 2024). The WFMD retrieval algorithm is less susceptible to these artefacts, this is also why we chose to use the WFMD retrieval product. Since our posterior fluxes remain very close to the prior ones, we do not see any influence of possible retrieval artefacts on the fluxes from the inversion.

L318-321: Related to questions above, why do you think that the flux were not as well constrained in the TROPOMI inversions? Satellite data are suppose to have good spatial coverage compared to ground-based data, and with much large number of data, it should, in principal, constrain the fluxes better than the ground-based data. But it is not the case here. Is this a general feature or something specific to high northern latitudes?

For the constraint on the fluxes, it is not only how many observations there are, but how sensitive each observation is to the fluxes and what uncertainty they have.

Ground-based observations are made to the largest extent within the boundary layer (possible exception is high mountain stations, which may sometimes, e.g. at nighttime, be above the boundary layer) and thus are more sensitive to surface fluxes. Satellite observations, on the other hand, have varying sensitivity to different layers of the atmosphere and are not as sensitive to the fluxes at the surface.

Furthermore, the ground-based observations are made at much higher accuracy and precision than the satellite observations, which again means these have a stronger constraint on the fluxes. For instance, the median uncertainty in the observation space for the ground-based observations was 8 ppb, whereas it was 16 ppb for the TROPOMI observations, and the model-observation differences are weighted by the inverse of the square of the observation uncertainties.

Another contributing factor for the poorer constraint from the satellite observations is that the satellite SRRs are always smeared out over large regions, whereas the SRRs for ground-based observations can be much more focussed over a small region. The latter leads to larger increments over the background, at least potentially (i.e., if there are sources in that region). This has two effects:

- i) Larger deviations from the background make the inversion less sensitive to the exact value of the background, and less information is used to optimize the background.
- ii) Since the cost function (in the inversion) includes the quadratic difference between observation and modelled mixing ratio, a few large differences have more impact than a large number of small differences. This means that the ground-based observations (with more focused SRRs) have more impact.

Technical comments

Introduction: Please add information about focus/simulation years

We have now added this information to the last paragraph of the introduction.

L67: Please add references to FLEXPART.

We have now added the reference for FLEXPART v10.4 which is the version used in this study.

Equations: Please use bold fonts for vectors and matrices.

We have corrected all our equations using bold lower case letters for vectors and bold upper case letters from matrices.

L210: ...ERA5 at $0.5^\circ \times 0.5^\circ$ and... ?

Yes, we mean $0.5^\circ \times 0.5^\circ$ and include the second 0.5° in the text.

Table 1. What are the altitudes here? Elevation of the site or height from which FLEXPART trajectories were calculated?

These are the sample air intake heights as stated by the data providers.

Figure 2: Are these of super-observations? Please clarify.

These are the super-observations averaged over the month onto a regular grid of 0.25×0.25 degrees. We now specify this in the caption.

Figure 4:

- You could perhaps consider adding number of observations here?
- Could you also consider adding ranges?

Fig.4 is the figure showing the area-weighted mean XCH₄ over time from the observations and the prior and posterior model. We are not sure where we would add the number of observations here. In any case, we show the number of observations in now in the new Fig. S2.

We are not sure what ranges the reviewer is referring to. If the reviewer means to add the uncertainties for the observations we can do that (i.e., the area-weighted mean uncertainties). As for the modelled XCH₄, we do not calculate uncertainties for these, only we assign an uncertainty in the observation space for the inversion to account for the model representation uncertainty and the measurement uncertainty. This, however, is different from the uncertainty on the modelled XCH₄ values which necessarily depend on the uncertainty in the fluxes as well as in the model representation.

Please consider combining Figures 5 and 8, and Figures 7 and 9. It would be easy to compare between TROPOMI-based and ground-based inversions that way.

We have combined Figures 5 and 8 showing the prior and posterior fluxes and flux increments from all inversions. However, we prefer not to combine Figures 7 and 9, because we want to keep the focus of Fig. 7 on the comparison of the uncertainty reduction for TROPOMI for all months versus for spring and for summer. Also, considering the formatting it will be awkward because it is not possible to fit 4 panels horizontally, and having 3 panels horizontally (i.e. those for TROPOMI) and one panel below would look strange.

References

Lindqvist, H., et al.: Evaluation of Sentinel-5P TROPOMI Methane Observations at Northern High Latitudes, *Remote Sensing*, 16, 2979, <https://doi.org/10.3390/rs16162979>, 2024.