Response to Referee #3

We thank Julia Marshall for constructive comments. Our responses are shown below in blue.

One of the most scientifically interesting findings is the temporal dependence in the differences between the GOSAT and the TROPOMI data (compared through a model), shown in Figure 6. The fact that the largest seasonal features correlate with changes of surface albedo suggests directions for the future improvement of (TROPOMI) retrievals algorithms. I think it would be worth highlighting this finding in the abstract as well.

Thanks for the suggestion. We add a sentence in the abstract highlighting this finding, which is mainly related to Northwest China and Kazakhstan.

Like the last reviewer, I am also puzzled by how small the error bars on the posterior flux estimates are. The authors mention that these uncertainties are "optimistic", as they neglect systematic uncertainties due to the inversion setup. This is not ideal, and should be improved in the future, but does not invalidate the focus of the GOSAT-TROPOMI comparison presented here.

Thanks for the comment.

Minor comments that need to be addressed

L326: The WLG site show(s) a relatively low posterior what? (Also, it's not the site itself that is unable to capture variability, as this sentence implies.) How can the simulations simultaneously show "relatively good agreement" and "relatively low posterior (correlation)" at WLG? Relative to what? Is it really "day-to-day" variability that is poorly captured in the case of flask data? I guess the sampling is roughly weekly, although the bizarre x-axis in S9 makes this impossible to say for sure (see comment about the supplement below).

We now change to "*Both posterior simulations … achieve **reasonable** agreement at PDI, UUM, and WLG…*" and "*…inability to capture **sub-seasonal** variability*". We have also changed the x-axis in S9 (It's now renamed to S10).

Figure 5 is lacking a colour bar!

The figure has been updated.

L506-507: Based on the error bars, only CSC and NEC show consistent (and not just qualitatively similar) posterior emissions. This is important. Either the uncertainties in the posterior emissions estimates are grossly underestimated (likely true), or the datasets are fundamentally inconsistent with one another (also likely true).

We already discuss in the manuscript the inconsistency between the two inversions over NWD and SXJC. For BAN and CNC, the evidence that the two inversions are inconsistent is not as strong. Note the error bar shown in Figure 3 is one standard deviation (1σ).

Figure S9: Something very strange is going on with the x-axes. It looks like the observations are spaced equally, and the ticks for the months are adjusted to fit. This happens in such a way that there are always measurements on the first of each month, which is suspicious. This is not a normal way of presenting timeseries data, and should be fixed. Are there really only 23 FTIR measurements from HF for the whole year?

We have updated Figure S10 following the suggestion about x-axes. The figure shows daily-average observations. For HF, there are in total 29 days of measurements. After filtering out days with large solar zenith angles, only 23 daily averages are used for the analysis.

Technical/stylistic comments/typos:

We have incorporated the following comments in the revised manuscript.

- L30: Suggested rephrasing: The methane emissions inferred from GOSAT observations are … higher than those from TROPOMI observations…
- L31: These -> The
- L38: I think the “Ganges Plain” is usually referred to as the “Indo-Gangetic Plain”. I would recommend changing it throughout.
- L67: a number is missing regarding the TROPOMI footprint.
- L94: Is it already clear before this study that the two sets of satellite data differ in their “regional accuracy”? If so, a citation is needed! Otherwise, remove this.
  We have changed it into “regional bias”, and a citation has been added.
- L114: are -> is
- L126: annual average XCH_4 on the 0.625°× 0.5° grid for GOSAT and TROPOMI -> XCH_4 measured by GOSAT and TROPOMI, annually averaged on the 0.625°× 0.5° grid.
- L127: over THE Mongolian…
- L129: of multiple measurements fall in -> when multiple measurements fall within
- L163: located distant -> far
- L182: more discussed -> discussed further
- L183: early -> earlier (or previous)
- L204: either “a biased boundary condition” or “biased boundary conditions”
- L264: overestimate -> overestimates
- L296: estimate -> estimates
- L303: applying A traditional regularization
- L431: are in India -> either “that are in India” or “in India”
- L431: subscript in XCO2
- I would really like to see a figure of the 600 spatial clusters included in the supplement. See Figure S5.
- L540: benefit -> improve
- Caption of Figure S10, last line: the bias correction factor greater than -> the bias correction

factor is greater than

Brief responses to the following comments:

- L79: A bit confused by the use of "variable" here. Is it required?

  The term "variable bias" is used as in Jacob et al. (2022).

- L97, but also in abstract and throughout the study: I guess North(ern) India should be capitalized throughout? As is every region of China… This is certainly how Wikipedia does it: https://en.wikipedia.org/wiki/North_India

  We checked papers published in ACP. It appears that either northern India or North India is ok.

- L149: Were the *in situ* data temporally filtered? Generally nocturnal measurements are not well represented by models.

  Daytime measurements are used to compare with simulations. This information is now added to the text.

- L190-191: WetCHARTs is an ensemble product, which specific ensemble member was used? Is there any concern about double-counting rice emissions as wetlands?

  We use the average from the WetCHARTs ensemble. This information is now added to the text. Rice and wetland emissions used in the simulation show distinct spatial distribution. Therefore, the double-counting issue should not be a concern.

Reference

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, Atmos. Chem. Phys., 22, 9617-9646, https://doi.org/10.5194/acp-22-9617-2022, 2022.