Author response - RC2

We thank the reviewer for their thoughtful comments. We respond to them below: the comments are copied hereafter and shown in black, our author responses in blue; suggested new manuscript text is indicated in green. New line numbers in the revised manuscript are provided.

This manuscript by Sicsik-Paré et al. presents a comparative analysis of three XCH_8 products derived from TROPOMI observations—SRON, BLENDED, and WFMD. The authors examine various parameters influencing discrepancies among the XCH_8 datasets, highlighting aerosol scattering and SWIR albedo as primary contributors. Additionally, the study utilizes these products in inverse modelling with the CHIMERE model and a data assimilation system to estimate surface methane fluxes.

Overall, the manuscript provides valuable insights for the atmospheric methane community, especially in the context of inverse modelling and satellite product evaluation. It demonstrates the potential and limitations of different XCH_8 retrievals and their influence on flux estimates. I recommend publication in ACP after addressing the following general and specific comments.

General comments:

1. The manuscript is generally well-structured but somewhat lengthy in certain sections. For instance, the detailed description of the CHIMERE model in Section 2 could be moved to the appendix to improve the flow and readability of the main text.

The feedbacks from the different reviewers all suggested to improve concision, especially in Section 2. This section was reorganized to remove unnecessary statements that did not support the main message of the article. In particular, we have made the description of CHIMERE more concise. However, we prefer to keep this subsection in the main text, as the description of the model set-up is usually part of the main text for inversion articles.

2. The WFMD product reports the highest absolute XCH_8 values among the three datasets and shows the largest bias relative to TCCON (7.5 ppb; see Table C2). While the WFMD XCH_8 dataset shows moderate differences from both prior and posterior simulations compared to SRON and BLENDED, the fluxes derived from WFMD are consistently the lowest (Figure 10b). This discrepancy raises discussion here.

Could the lower flux estimates be a result of overestimated background subtraction? Furthermore, WFMD includes approximately 30% more retrievals than the other products. Could this may be due to less stringently filtered, allowing residual uncertainties to remain, which could contribute to the observed differences in flux estimates?

The lower emission estimates derived from the WFMD inversion are indeed a crucial element of analysis in this study. This element is discussed at the end of Section 3.3. We did not manage to identify a unique origin of these strong negative increments: the distribution of obs-sim is overall positive (Fig. 9) but with values closer to 0 than SRON, so we expect lighter increments than for SRON inversion – this is not the case. We investigated the background optimization (strong negative increments could have compensated strong positive increments on the background, which is not the case) but it was not conclusive either (Fig. D4, D5).

This is likely due to a from subtle balance within the inversion process: the differences in spatial distributions of background increments and emission increments with the other products could come from differences in the separation of background and emissions when using WFMD observations. The high number of retrievals could also result in overfitting in the optimization process, and/or residual XCH_4 uncertainties due to less stringent filtering. This topic has to be further explored to allow better understanding of the process.

The results of Rona Thompson presented at EGU 2025 also showed similar relative differences between products (Prior: 27.4 Tg/yr, SRON: 26.4 Tg/yr, BLENDED: 23.7 Tg/yr, WFMD: 19.0 Tg/yr, ICOS: 22.9 Tg/yr). These simulations were made with another model (FLEXPART). It cannot be considered a strict validation, but it suggests that WFMD observations do indicate lower CH4 emissions than the other products in Europe in 2019.

Reference: Thompson, R., Schneider, P., and Stebel, K.: Using different TROPOMI XCH4 retrieval products in atmospheric inversions of CH4: a comparison and reconciliation over Europe, EGU General Assembly 2025, Vienna, Austria, 27 Apr–2 May 2025, EGU25-9567, https://doi.org/10.5194/egusphere-egu25-9567, 2025.

3. While the study uses surface observations for the inversion, have the authors considered assimilating ground-based total column data, such as from TCCON or COCCON? These measurements are widely used to validate satellite products, as both observe total atmospheric columns and are therefore directly comparable. Incorporating such datasets could provide a more representative constraint on regional methane emissions and may help improve flux estimates, particularly in regions where surface observations are sparse or not fully representative.

The use of TCCON data in this study is limited to the comparison of XCH_4 datasets and is presented in Annex C. We do not use the data for the comparison of the emissions estimated from the inversions, thus it is not used as a validation dataset. We use in-situ ground-based measurements for the validation (Sections 3.3, 3.5).

We have considered the assimilation of ground-based total columns from TCCON or COCCON. We decided to assimilate in situ measurements as an independent dataset because the 19 stations used in this study offer higher spatial coverage than the 5 (COCCON) + 7 (TCCON) sites available in 2019 (with 2 redundant sites in Karlsruhe and Sodankylä).

Specific and technical comments:

Line 126-127: the authors reference previous work to justify scaling the errors by a factor of 2, but a brief explanation here would enhance clarity for readers unfamiliar with that context.

Considerations on the definition of errors have been grouped in Section 2.1.3 (L.195-214). Specifically, we have described more thoroughly the error in the SRON product, and explain that the product authors « suggest to multiply them with a factor 2 to reflect the scatter of errors in the TCCON validation ». This is based on the ReadMe of the SRON product (see reference to Landgraf et al., 2024, page 10 of the document), but no further explanation is available.

Line 207: Figure 3 is referenced before Figure 2 in the text. The order of figures should be revised for logical flow.

The order of references has been reversed in the text (L. 175,176).

Line 207-208: The phrase "BLENDED lower concentrations" is unclear. Does this mean that only the lower end of BLENDED values agrees well with GOSAT? Consider clarifying this point. Additionally, including GOSAT's spatial distribution or mean value in Figure 3a would improve visualization and support the comparison.

The explanations have been clarified: the statement that BLENDED XCH_4 are lower than SRON and WFMD for all months is quantified (L. 184,185), and the statement that BLENDED is the closest to GOSAT is mentioned (« Only BLENDED aligns rather well with GOSAT, due to its correction of the TROPOMI-GOSAT bias »).

We chose not to compare averages of TROPOMI XCH4 with GOSAT XCH4 for the readability of the Figure. Also, the sampling is very different for both instruments: the comparison of averages can therefore deceive the reader in thinking than the differences are due to biases between products, whereas they can be caused by differences in coverage.

Figure 3: It seems the errors for SRON and BLENDED largely overlap in panels (b) and (c). If this is the case, please clarify it in the figure caption.

BLENDED errors are retrieved from SRON, so errors from both datasets are exactly equal. It is explained in the main text at L.129 and 196. The sentence « SRON and BLENDED errors are exactly similar, as BLENDED errors are directly retrieved from the SRON product. » has been added to the caption of Fig. 3 to improve readability.

Line 240: typical of SWIR retrievals >>> type?

This has been changed to « characteristic of SWIR retrievals ».

Line 408: The analysis indicates that differences are closely linked to across-track pixel index, aerosols, and SWIR albedo. While the latter two are physically intuitive, the role of across-track pixel index is less clear, as it represents only a positional identifier. Could these index be indirectly capturing information related to albedo, aerosol, or stripping artifacts in the retrievals? Further clarification would be beneficial.

The link between the positional across-track ID and striping patterns is described in L.342. A reminder « across-track pixel index (thus striping patterns) » has been added at the mentioned location (now L.365) to remind the reader of the link between the ID and striping effects.

Line 519: Background removal is known to be challenging. The manuscript would benefit from a brief explanation of how the background is defined and removed in this study.

The background mentioned here is composed of the lateral and top boundary conditions, the initial conditions and the stratosphere concentration field that are taken from the CAMS global product. It is defined in Section 2.4.2 and we have added here a reference to this section.

Line 531-532: result of >>> result from?

The correction has been done.

Line 560: The authors note poor improvement over Scandinavia. Could extending the study period improve coverage and reduce uncertainty through increased data availability? This point deserves further discussion.

Poor improvement over Scandinavia is due to the lower coverage in this region (Figure 1), in particular in winter (many pixels with no observations for JFM and OND in Figure D3).

The extension of the study has been considered to generalize the analyses and capture seasonal signals (only one year is not sufficient, some temporal variations could be specific to this year). It has not been implemented because of the already high number of simulations performed in this study, and their high computation cost. However, it is not clear that extending the study period would reduce uncertainty: the increased coverage would provide more data, but in the meantime the inversion would have more emission and background components to constrain, resulting in the same « constraint capacity » of the inversion. The ratio between the number of observations in Scandinavia and the one in other parts of the domain would remain approximately similar, thus the inversion would behave similarly and provide low constraints in Scandinavia.

The study of high latitude regions has to be the object of dedicated studies, which has already been the case using TROPOMI data: the studies of Tsuruta et al. (2023) and Lindqvist et al. (2024) are cited in the manuscript, see also the more recent study of Kivimäki et al. (2025).

Reference : Kivimäki, E., Aalto, T., Buchwitz, M., Luojus, K., Pulliainen, J., Rautiainen, K., Schneising, O., Sundström, A.-M., Tamminen, J., Tsuruta, A., and Lindqvist, H.: Environmental drivers constraining the seasonal variability of satellite-observed methane at Northern high latitudes, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2025-249, 2025.

Appendix C: COCCON is increasingly used as a complementary dataset to TCCON, particularly in regions lacking TCCON coverage. The authors could also consider incorporating COCCON measurements to provide additional reference data for satellite validation, thereby enhancing the spatial representativeness of the evaluation.

See the answer to General comment #3: we agree that COCCON would provide additional data enhancing the comparison of TROPOMI XCH_4 products. The use we make of TCCON data in this study is limited to the comparison of XCH_4 datasets: as it is not directly used for the comparison of the emissions estimated from the inversions, it is not an evaluation of our results but more a complementary comparative description of XCH_4 products. For this reason, the section was moved to Appendix and we did not consider using more data from COCCON. For further studies, it will be very beneficial to use both observational networks to enhance the spatial representativeness of the analyses.