This manuscript by Sicsik-Paré et al. presents a comparative analysis of three XCH_4 products derived from TROPOMI observations—SRON, BLENDED, and WFMD. The authors examine various parameters influencing discrepancies among the XCH_4 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_4 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.

2. The WFMD product reports the highest absolute XCH_4 values among the three datasets and shows the largest bias relative to TCCON (7.5 ppb; see Table C2). While the WFMD XCH_4 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?

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.

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.

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

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.

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.

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

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.

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.

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

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.

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.