The authors would like to thank the editor and the reviewers for their precious time and invaluable comments. The corresponding changes and refinements are highlighted in yellow in the revised paper and are also summarized in our responses below. Authors' responses are in blue. Reviewers' comments are in black. When the manuscript is cited, it is shown in italics.

Referee #2

First of all, I want to thank the authors for this well-structured and clear and concise article. In terms of grammar, I have nothing to add. However, I do have three important issues that need to be addressed.

Major Comments:

1. The first is the issue of proper accreditation. Each of the individual TCCON datasets used, should have their data reference added to the references list (See https://tccon-wiki.caltech.edu/Main/DataLicense for TCCON citation guidelines). These individual per station citations can be found on https://tccondata.org/.

   ➔ Thank you for pointing this out. Following the TCCON data citation guidelines, we added the individual data references for all TCCON station datasets used in this study to the reference list. We also moved the TCCON site table to the main manuscript to clearly present the stations, site information, and corresponding citations. These revisions ensure proper accreditation of the TCCON datasets used in this study.

   Lines 87-88: "*Stations used in this study are shown in Fig. 1 and listed in Table 1.*"

2. The second point pertains to the selection of GOSAT-2 as the standard to which the other satellite products are bias-corrected (step 2 in the overall process). This selection of GOSAT-2 is based on the results listed in Table-3. However, it is not clearly stated if the common data sample on which Table 3 is based, is the training sample or instead comes from the LOSOCV approach. If the first, I consider this to be a weak basis for selection, if the latter it is a stronger one as we are fundamentally interested in the performance of the bias corrected products outside the scope of the training dataset. That said, we also need to take the global distribution of the TCCON network into account, which under-samples large swaths of the globe. In that view it is hard to state, with confidence, based on the analysis performed here, that GOSAT-2 should be taken as the definitive reference. I would very much prefer it if the authors performed 3 different step2 analysis wherein in turn, GOSAT, GOSAT-2 and TROPOMI are taken as a reference. This will allow the authors to perform an intercomparison, assess the impact of this choice on a global scale, identifying regions where things converge and diverge, and make a more thorough determination on whether all 3 of these end products turn out to be valid candidates or that one is superior.

   ➔ We thank the reviewer for this important methodological question. We confirm that Table 4 is based entirely on leave-one-site-out cross-validation (LOSOCV) results, not the

training sample, ensuring that the evaluation is independent of the training data and provides a cross-validated basis for selecting GOSAT-2 as the harmonization reference. We have clarified this in the revised manuscript text and table caption.

In addition, we agree that the spatial distribution of the TCCON network is uneven and does not fully represent the diverse retrieval environments encountered globally. Therefore, we acknowledge that the TCCON-based bias-corrected GOSAT-2 product should not be regarded as an absolute global truth standard. Given this limitation, we used the product with the most stable performance in the LOSOCV evaluation against TCCON as a practical reference scale for multi-sensor harmonization. Because the validation site conditions are excluded from training in LOSOCV, this evaluation provides an independent basis for reference-scale selection within the constraints of the TCCON network. This limitation has been clarified in the revised manuscript and added to the limitation discussion in Section 4.5.

We agree that comparing Step 2 harmonization results using GOSAT, GOSAT-2, and TROPOMI as alternative references would be scientifically valuable. However, TROPOMI showed lower and more variable performance than the GOSAT series across different cross-validation strategies in Table S5, with LOSOCV R^2 = 0.79 for TROPOMI compared with 0.87–0.88 for the GOSAT series. Furthermore, TROPOMI XCH_4 retrievals are known to carry inherent surface-reflectance-dependent biases from the 2.3 μm spectral band (Lorente et al., 2021; Lorente et al., 2023; Somkuti et al., 2025), and previous studies have specifically used GOSAT as a harmonization anchor because of its higher spectral precision and retrieval stability (Balasus et al., 2023; Li et al., 2024; Fan et al., 2024). Figure R6 confirms that TROPOMI shows strong albedo dependence in the standard product. Although ML-based correction reduces this within the TCCON-sampled range, the training data are limited to a narrow albedo range of approximately 0.0–0.30. The GOSAT series, in contrast, shows stable retrieval characteristics even within this limited range, suggesting comparably stable performance beyond it as well. Therefore, we considered TROPOMI a lower-priority candidate as the harmonization reference for the reference-scale selection objective of this study. Instead, we performed an additional analysis comparing GOSAT-anchored and GOSAT-2-anchored harmonization within the GOSAT series, which we considered the more plausible candidate set.

Figure R7 showed that GOSAT-2 provided more inter-satellite co-location samples with TROPOMI across all years, months, and latitudinal zones, which is advantageous for training a more stable and geographically representative harmonization model. Cross-validation (CV) results further showed that the GOSAT-2-anchored harmonization produced lower errors than the GOSAT-anchored harmonization across different CV strategies (Table R2), latitudinal zones, and months (Fig. R8). Both reference frameworks substantially reduced surface albedo and AOT dependencies in the harmonized TROPOMI product (Fig. R9), but the GOSAT-2-anchored framework showed more stable overall performance.

Taken together, these results do not establish GOSAT-2 as an absolute global reference. However, considering the LOSOCV-based TCCON evaluation, the reference-scenario sensitivity analysis within the GOSAT series, and the larger co-location sample with

TROPOMI, the TCCON-based bias-corrected GOSAT-2 product is the most reasonable choice as a practical and well-validated reference scale for the multi-sensor harmonization framework in this study.

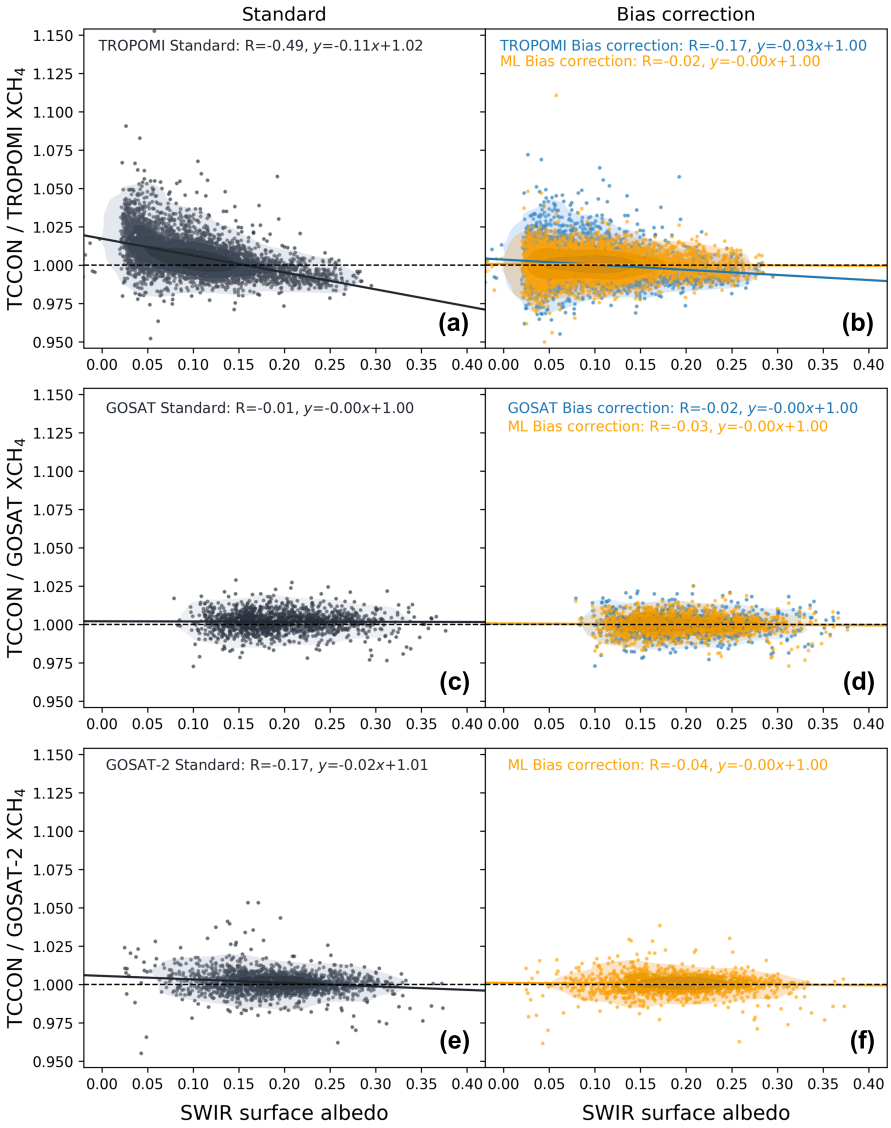

Figure R6. SWIR surface albedo dependence of satellite XCH_4 products relative to TCCON under LOSOCV, across an albedo range of 0.0–0.40. Left column shows standard products and right column shows bias-corrected products for TROPOMI (a, b), GOSAT (c, d), and GOSAT-2 (e, f). Orange points indicate ML-based bias-corrected results and blue points indicate operational bias-corrected results. Solid lines denote least-squares regression fits and the dashed horizontal line indicates the unbiased reference level (ratio = 1). Pearson correlation coefficients (R) and regression equations are reported in each panel.

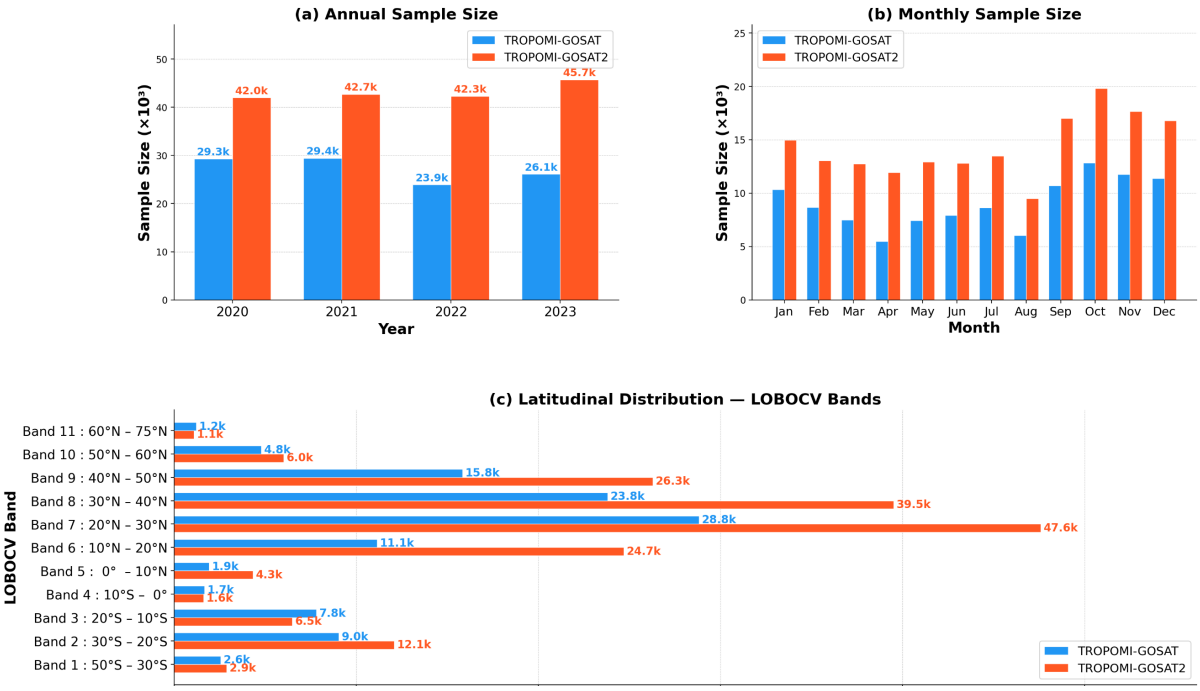

Figure R7. Comparison of inter-satellite colocation sample sizes for TROPOMI-GOSAT (blue) and TROPOMI-GOSAT-2 (orange) pairs used in Step 2 harmonization for 2020–2023. (a) Annual sample size, (b) monthly sample size, and (c) latitudinal distribution across LOBOCV bands.

Table R2. Cross-validation performance comparison of TROPOMI harmonization using GOSAT and GOSAT-2 as alternative harmonization references. Boldface indicates the better-performing value for each metric

		LOBOCV[a]	LOMOCV[b]	LOYOCV[c]
GOSAT as reference	N		112,341	
	R^2	0.86	0.88	0.87
	MAE (ppb)	8.62	8.06	8.39
	RMSE (ppb)	11.32	10.61	11.02
GOSAT-2 as reference	N		183,550	
	R^2	**0.91**	**0.92**	**0.92**
	MAE (ppb)	**6.78**	**6.32**	**6.59**
	RMSE (ppb)	**9.07**	**8.49**	**8.82**

[a] LOBOCV: Leave-One-Band-Out Cross-Validation
[b] LOMOCV: Leave-One-Month-Out Cross-Validation
[c] LOYOCV: Leave-One-Year-Out Cross-Validation

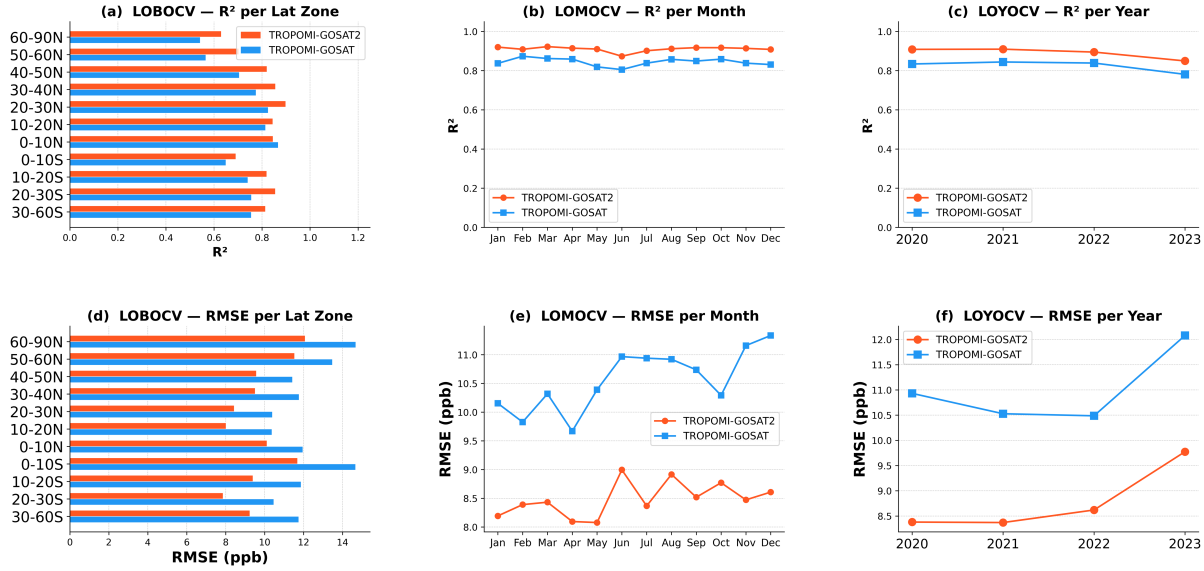

Figure R8. Cross-validation performance comparison of TROPOMI harmonization using GOSAT (blue) and GOSAT-2 (orange) as alternative harmonization references. Panels (a–c) show R^2 results for LOBOCV by latitudinal zone, LOMOCV by month, and LOYOCV by year, respectively. Panels (d–f) show the corresponding RMSE results for LOBOCV by latitudinal zone, LOMOCV by month, and LOYOCV by year.

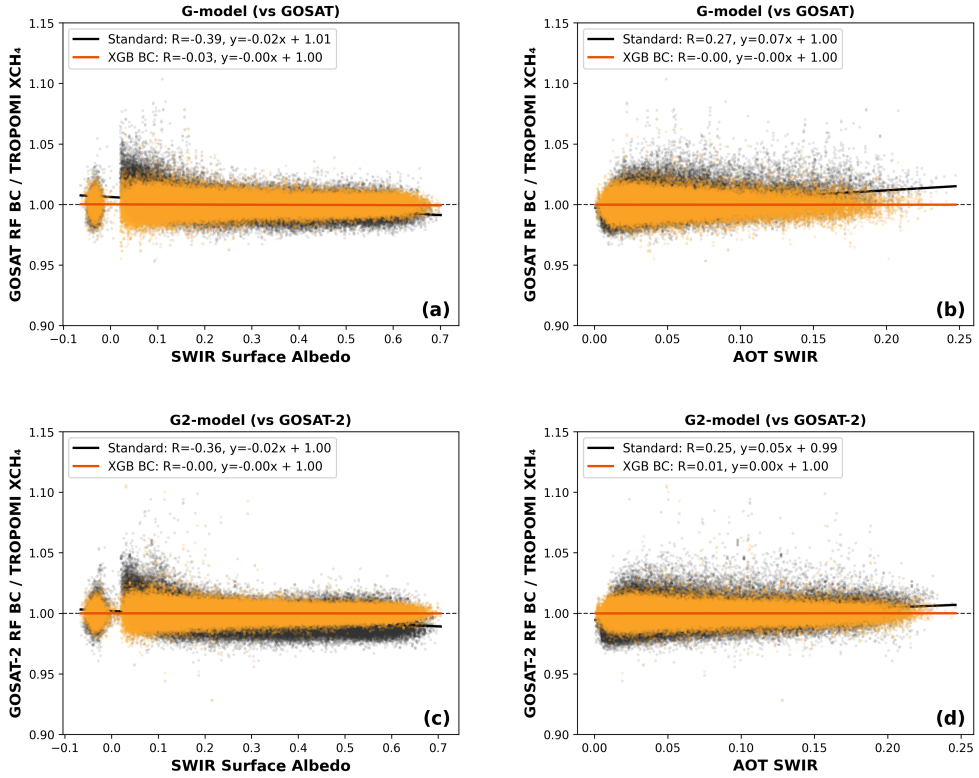

Figure R9. Dependence of the harmonized TROPOMI XCH_4 ratio on SWIR surface albedo (left column) and aerosol optical thickness (AOT SWIR; right column) before (black) and after (orange) XGBoost-based harmonization. Results are shown for TROPOMI harmonized to GOSAT (G-model; panels a, b) and to GOSAT-2 (G2-model; panels c, d). Solid lines denote least-squares regression fits and the dashed horizontal line indicates the unbiased reference level (ratio = 1).

Lines 296-298: *"Table 4 compares the ML-based bias-corrected products from all three sensors evaluated against TCCON using common collocated samples under LOSOCV, ensuring that the reference selection is based on independent out-of-site performance."*

Lines 454-459: "*First, the ML-based bias correction relies on TCCON as the primary reference dataset. Although TCCON provides high-precision ground-based XCH_4 observations, its spatial distribution is limited and does not fully represent the diverse retrieval conditions encountered globally, particularly high-surface-albedo conditions. We partly addressed this limitation through leave-one-station-out cross-validation and additional satellite match-up analyses over broader surface albedo ranges, but independent validation under underrepresented retrieval conditions remains necessary.*"

References:

Lorente, A., Borsdorff, T., Butz, A., Hasekamp, O., aan de Brugh, J., Schneider, A., Wu, L., Hase, F., Kivi, R., Wunch, D., Pollard, D. F., Shiomi, K., Deutscher, N. M., Velazco, V. A., Roehl, C. M., Wennberg, P. O., Warneke, T., and Landgraf, J.: Methane retrieved from TROPOMI: improvement of the data product and validation of the first 2 years of measurements, Atmospheric Measurement Techniques, 14, 665–684, https://doi.org/10.5194/amt-14-665-2021, 2021.

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, https://doi.org/10.5194/amt-16-1597-2023, 2023.

Somkuti, P., McGarragh, G., O'Dell, C., Di Noia, A., Vogel, L., Crowell, S., Ott, L. E., and Bösch, H.: Surface reflectance biases in XCH_4 retrievals from the 2.3 µm band are enhanced in the presence of aerosols, Atmospheric Measurement Techniques, 18, 4647–4663, https://doi.org/10.5194/amt-18-4647-2025, 2025.

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, https://doi.org/10.5194/amt-16-3787-2023, 2023.

Li, K., Bai, K., Jiao, P., Chen, H., He, H., Shao, L., Sun, Y., Zheng, Z., Li, R., and Chang, N.-B.: Developing unbiased estimation of atmospheric methane via machine learning and multiobjective programming based on TROPOMI and GOSAT data, Remote Sensing of Environment, 304, 114039, https://doi.org/10.1016/j.rse.2024.114039, 2024.

Fan, L., Wan, Y., and Dai, Y.: Development of a Multi-Source Satellite Fusion Method for XCH4 Product Generation in Oil and Gas Production Areas, Applied Sciences, 14, 11100, https://doi.org/10.3390/app142311100, 2024.

3. The third point addresses the rank order merging method used in step three, where each 0.1° daily grid is filled with GOSAT-2 if available, then TROPOMI if available, then GOSAT. If the ML harmonization between the satellite products performed in step 2 is successful, I see no reason why this method is superior to simply taking the median of all products in the daily grid cell. If the ML harmonization is unsuccessful, then clearly the rank order creation of the merged dataset isn't the solution either.

   ➔ Thank you for this important comment. We agree that the merging strategy after

harmonization should be clearly justified. To address this, we additionally quantified the overlap frequency of available satellite observations in the valid daily 0.1° grid cells of the final fused XCH_4 product.

Figure R10 shows the overall, daily, and spatial sensor-overlap frequencies for 2020. More than 99% of valid grid cells contained only one satellite observation, whereas cases with two or more simultaneous satellite observations accounted for less than 1% (Figs. R10a, b). Spatially, the frequency of three-sensor overlap was mostly below 0.1% of valid days, and the frequency of two-sensor overlap was generally below 10% (Figs. R10c, d). Under this sparse-overlap structure, the practical advantage of median-based merging as a robust estimator is limited. When only one satellite observation is available, median, mean, and priority-selection methods produce the same value, and the advantage of the median becomes meaningful only when multiple observations are frequently available within the same grid cell.

Therefore, the main purpose of fusion in this study is not to statistically average multiple sensor values within the same grid cell, but to use complementary observations from the GOSAT series in regions or conditions where TROPOMI retrievals are limited. As discussed in Section 4.3, the overall increase in global coverage from fusing the three satellites was modest, and most coverage was contributed by TROPOMI. However, the GOSAT series can provide additional XCH_4 information under some conditions and regions where TROPOMI observations are unavailable; thus, the benefit of fusion lies more in complementary sampling than in a large increase in coverage percentage.

For this reason, we considered the priority-selection strategy to be the most appropriate. ML-based harmonization reduces systematic inter-sensor discrepancies by converting GOSAT and TROPOMI to the bias-corrected GOSAT-2 scale, and Figure 7 shows that the inter-sensor bias distributions became more aligned with reduced spread after harmonization. Nevertheless, residual inter-sensor differences can remain even after harmonization; therefore, a simple mean or median is not necessarily optimal when multiple values are available in the same grid cell. We therefore prioritized the ML-based bias-corrected GOSAT-2 product, which defines the reference scale, followed by TROPOMI and GOSAT based on their agreement with reference and sampling density.

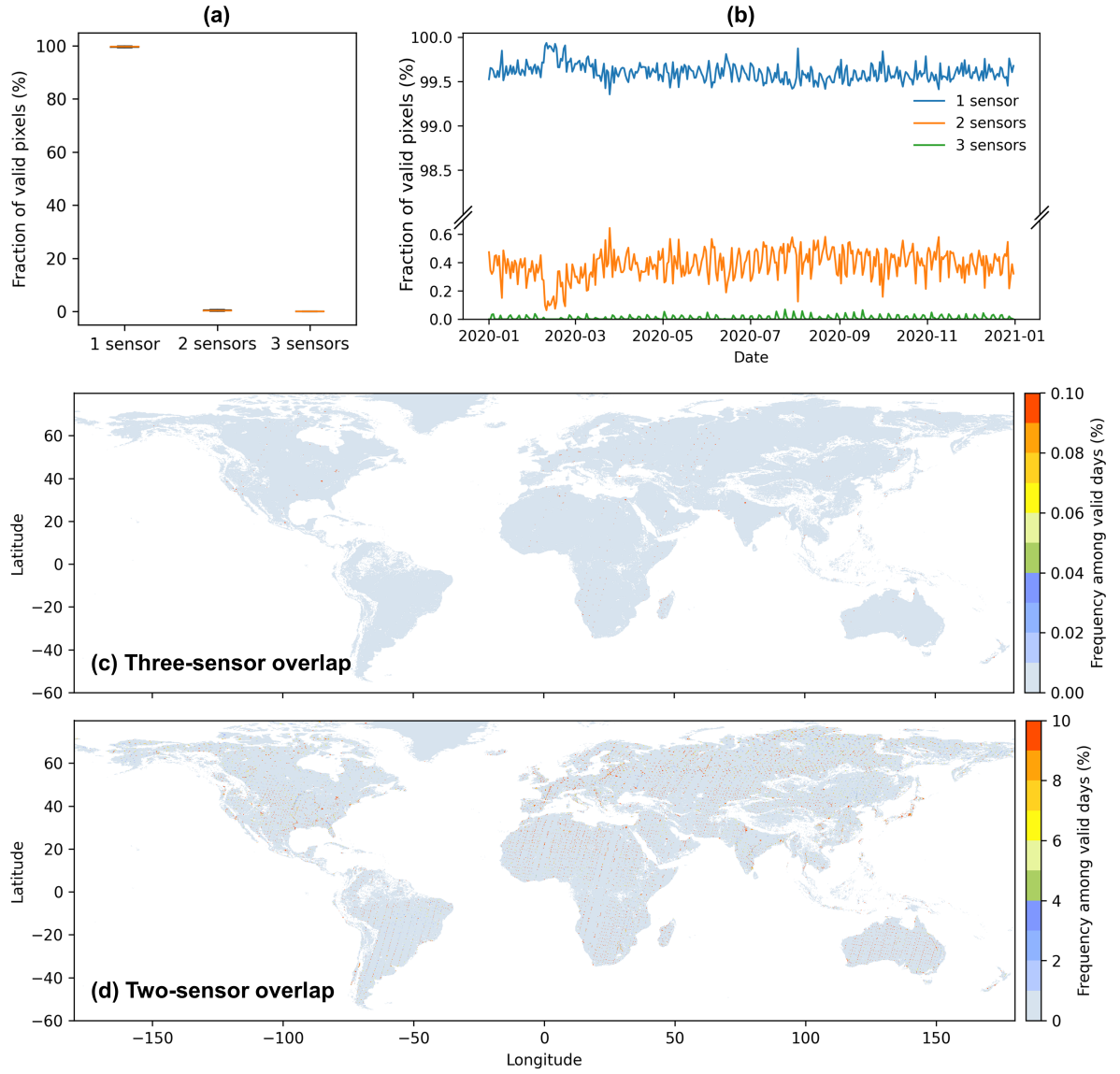

Figure R10. Sensor-overlap frequency in valid daily 0.1° grid cells. Distribution of the number of available satellite observations used for the fused XCH_4 product in 2020. (a) Overall fraction of valid grid cells containing one, two, or three available sensors. (b) Daily variation in the fraction of valid grid cells by the number of available sensors. (c, d) Spatial frequency of three-sensor and two-sensor overlaps, expressed as the percentage of valid days at each grid cell.

Minor Comments:

1. Line 164: Sha et al. 2021 is used as a source for the used validation colocation criteria. Note however that Sha et al. consider the line of sight of the FTIR instrument. To quote the paper:"An effective location of the FTIR measurement on the line of sight (i.e. at a 5 km altitude) is used to do the co-location". This should be acknowledged.

   Sha, M. K., Langerock, B., Blavier, J.-F. L., Blumenstock, T., Borsdorff, T., Buschmann, M., Dehn, A., De Mazière, M., Deutscher, N. M., Feist, D. G., García, O. E., Griffith, D. W. T., Grutter, M., Hannigan, J. W., Hase, F., Heikkinen, P., Hermans, C., Iraci, L. T., Jeseck,

P., Jones, N., Kivi, R., Kumps, N., Landgraf, J., Lorente, A., Mahieu, E., Makarova, M. V., Mellqvist, J., Metzger, J.-M., Morino, I., Nagahama, T., Notholt, J., Ohyama, H., Ortega, I., Palm, M., Petri, C., Pollard, D. F., Rettinger, M., Robinson, J., Roche, S., Roehl, C. M., Röhling, A. N., Rousogenous, C., Schneider, M., Shiomi, K., Smale, D., Stremme, W., Strong, K., Sussmann, R., Té, Y., Uchino, O., Velazco, V. A., Vigouroux, C., Vrekoussis, M., Wang, P., Warneke, T., Wizenberg, T., Wunch, D., Yamanouchi, S., Yang, Y., and Zhou, M.: Validation of methane and carbon monoxide from Sentinel-5 Precursor using TCCON and NDACC-IRWG stations, Atmos. Meas. Tech., 14, 6249–6304, https://doi.org/10.5194/amt-14-6249-2021, 2021.

➔ Thank you for pointing this out. We agree that Sha et al. (2021) used a line-of-sight-based effective FTIR location for co-location, whereas our TROPOMI–TCCON co-location followed the spatial-temporal criteria described by Balasus et al. (2023). We therefore revised the sentence and replaced the citation accordingly.

Lines 171-173: "*We therefore constructed collocated training pairs using satellite-specific spatiotemporal and elevation constraints, adopting criteria from previous studies and official validation strategies (Balasus et al., 2023; Yoshida et al., 2023).*"

Reference:

Balasus, N., Jacob, D. J., Lorente, A., Maasakkers, J. D., Parker, R. J., Boesch, H., ... & Varon, D. J. (2023). A blended TROPOMI+ GOSAT satellite data product for atmospheric methane using machine learning to correct retrieval biases. *Atmospheric Measurement Techniques*, *16*(16), 3787-3807.

2. Line 241: a space is missing between "conditions." and "Given".

➔ Thank you for pointing this out. We have corrected the missing space between "conditions." and "Given."

3. Line 295: please repeat that we are building a 0.1° daily product here.

➔ Thank you for the suggestion. We added the requested clarification that the final output is a daily 0.1° fused XCH_4 product.

Lines 316-317: "*Finally, we generated a daily 0.1° fused XCH_4 product by integrating the harmonized observations from the three satellites.*"

4. Line 308: I would not describe Xianghe as a rural site (it sits within 100 km of Beijing Centre) in a heavily industrialized and urbanized region. There is probably a mix-up with Edwards (described as urban), which 100 km radius touches the outskirts of Los Angeles but in and of itself is situated in the desert.

➔ Thank you for catching this. We have revised the site descriptions to better reflect each station's actual environment. Xianghe is located approximately 50 km east-southeast of

Beijing, within a densely populated and economically active region of China, and is surrounded by urban areas, croplands, and strong anthropogenic emission sources. Edwards is located in the arid high desert of California and is characterized by high surface albedo associated with nearby bright dry lakebeds. Accordingly, Xianghe is now described as a peri-urban site influenced by strong anthropogenic emissions, while Edwards is described as an arid high-desert site with high surface albedo. The descriptions have been corrected in the revised manuscript.

Lines 333-336: *"The three stations represent distinct environments: a peri-urban site influenced by strong anthropogenic emissions (Xianghe, China; mean = 1903.16 ppb), an arid high-desert site with high surface albedo (Edwards, USA; mean = 1880.37 ppb), and a mountainous site (Garmisch, Germany; mean = 1872.98 ppb)."*

5. Paragraph 336 onwards: Here we see an improvement of the fusion product compared to TROPOMI when comparing to GOSAT-2, but all components within the fusion product are ML bias corrected towards GOSAT-2. It is thus basically telling us the same as Figure 7.

➔ We thank the reviewer for this observation and agree that the previous Fig. 10b partially overlapped with the harmonization evaluation already presented in Fig. 7. To avoid redundancy and potential ambiguity in interpretation, we removed the regional discrepancy comparison figure from the revised manuscript. Section 4.3 now focuses on the primary objective of the fusion framework in retrieval-challenged environments: improving observational coverage and spatial complementarity among sensors. We have clarified this focus in the revised manuscript.

6. Paragraph 363 onwards: How exactly was the growth rate calculated. Also, no uncertainty values on the growth rate are given.

➔ We thank the reviewer for this helpful comment. The mean annual growth rates shown in Fig. 11b were estimated using linear regression of the monthly global mean XCH_4 time series over 2020–2023, with uncertainties reported as the standard errors of the regression slopes. The regional interannual growth rates in Fig. 11c were calculated as year-over-year differences (i.e., each month minus the corresponding month of the previous year). Annual markers represent the mean of the monthly growth rates within each year, while error bars represent the standard deviation (σ, $n-1$) of the monthly growth rates within each year. All linear trends were statistically significant ($p < 0.001$). Figure 11 and the corresponding discussion have been revised accordingly.

Lines 387-397: *"Mean annual growth rates were calculated from the monthly time series over 2020–2023, while uncertainties represent the variability of monthly growth rates within each year. The fused product captures a sustained increase from approximately 1850 ppb in early 2020 to nearly 1900 ppb by late 2023, corresponding to a mean global growth rate of* $12.28 \pm 1.00 ppb\ yr^{-1}$ *(Fig. 10b). This agrees well with TCCON* $(13.56 \pm 1.25 ppb\ yr^{-1})$ *and NOAA* $(14.77 \pm 0.64 ppb\ yr^{-1})$*, supporting the ability of the fused dataset to represent large-scale atmospheric* CH_4 *variability and interannual trends.*

Interannual growth rates exhibited strong regional variability (Fig. 10c). The NH (0–80°N), SH (60°S–0°), and the high-emission zone (0–40°N) all peaked in 2021 at 14.3 ± 2.2, 15.4 ± 2.7, *and* 15.5 ± 2.1 *ppb* yr^{-1}, *respectively, consistent with record global* XCH_4 *increases reported for that period (Saunois et al., 2025). Following this peak, growth rates generally declined during 2022–2023, reaching approximately* 6.0 ± 1.8 *ppb* yr^{-1} *in the NH,* 11.4 ± 3.9 *ppb* yr^{-1} *in the SH, and* 6.4 ± 3.0 *ppb* yr^{-1} *in the high-emission zone by 2023, consistent with recent inverse modeling estimates (Pendergrass et al., 2025)."*

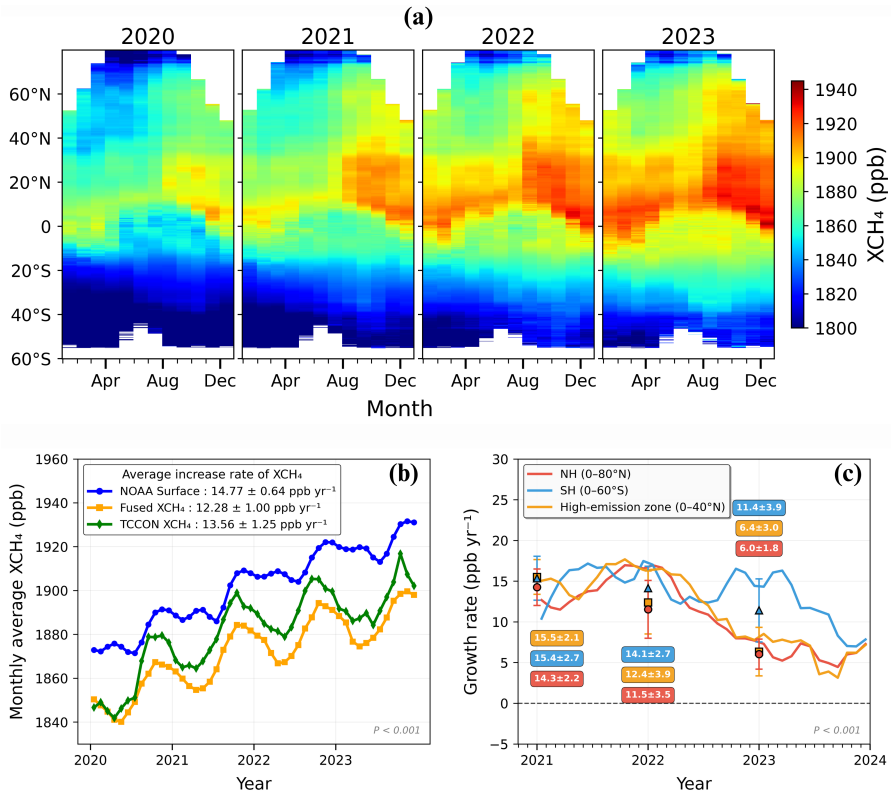

Figure 11. Spatiotemporal characteristics and growth of fused XCH_4 ***during 2020–2023. (a) Monthly latitudinal distribution of fused*** XCH_4 ***concentrations at 0.1° spatial resolution from 2020 to 2023. (b) Comparison of monthly global mean methane time series from the fused*** XCH_4 ***product (orange), NOAA marine surface*** CH_4 ***measurements (blue), and TCCON ground-based*** XCH_4 ***observations (green). Values in the legend indicate mean annual growth rates over 2020–2023 with associated uncertainties. (c) Monthly hemispheric year-over-year differences (solid lines) and annual mean growth rates (markers) for the Northern Hemisphere (NH; 0–80°N, red), Southern Hemisphere (SH; 0–60°S, blue), and high-emission zone (0–40°N, orange). Error bars represent the standard deviation of monthly growth rates within each year. All trends shown in panels (b) and (c) were statistically significant (p < 0.001).***

7. Line: 580: This is a AMTD reference, replace by its non-discussions final paper

➔ Thank you for pointing this out. We have replaced the AMTD reference with the appropriate final published reference and no longer cite the AMTD discussion paper in the revised manuscript.