Authors’ responses (egusphere-2026-1034)

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 #1

This study builds a 0.1° × 0.1° daily XCH_4 product covering 2020–2023. The authors do this by bias-correcting GOSAT-2 with TCCON as reference, then bias-correcting GOSAT and TROPOMI with the TCCON-informed GOSAT-2 product as reference, and then finally filling each daily 0.1° × 0.1° grid cell with the bias-corrected data from the three sensors, giving priority to GOSAT-2, then TROPOMI, then GOSAT. This fused product performs well against withheld TCCON data and provides additional coverage past that of TROPOMI in challenging retrieval environments. The presentation quality is excellent, but I encourage the authors to consider the following comments at their discretion to improve the scientific quality and significance.

Major Comments:

1. While TCCON is an excellent reference dataset, there are downsides to using it as the basis of your bias correction (e.g., limited independent validation data remains). Do the GOSAT-2 and TCCON co-locations represent diverse enough conditions (e.g., in surface albedo) to justify its use as reference? Figure 6 shows a small range of surface albedos relative to the global distribution (cf. Figure 6 in Balasus et al., 2023). In the case of TROPOMI, a small-area approximation can be used to generate data with more diverse retrieval conditions for reference purposes, either training or validation (Lorente et al., 2021).

   ➔ Thank you for this insightful comment. We agree with the reviewer’s opinion. Although TCCON provides a high-precision reference, its spatial distribution is limited and therefore cannot fully represent the diverse retrieval conditions encountered globally. In particular, most TCCON sites used in this study are located in areas with SWIR surface albedo below approximately 0.4 (Fig. 1). We therefore clarified this limitation of TCCON-based bias correction in the revised manuscript and performed additional analyses to assess the dependency on surface albedo.

   Using the TCCON co-location dataset (Fig. R1), the standard TROPOMI product showed a strong negative dependency on surface albedo, with R = -0.49 (Fig. R1(a)). The operational bias-corrected product, which applies the a posteriori albedo-bias correction described by Lorente et al. (2021), reduced this dependency, but a residual dependency remained, with R = -0.17 (Fig. R1(b)). In contrast, machine learning (ML)-based bias correction reduced this dependency to R = -0.02. For GOSAT and GOSAT-2, the standard products already showed weaker surface albedo dependencies than TROPOMI, and the dependencies remained low after ML-based correction, with R = -0.03 and R = -0.04, respectively (Fig. R1(c)–(f)). These results indicate that, within the albedo range represented by the TCCON sites, the ML-based bias correction effectively reduced surface-albedo-related bias.

To further evaluate the stability of the final target (i.e. TCCON-based bias-corrected GOSAT-2), under surface albedo conditions broader than those sampled by TCCON sites, we performed an additional satellite match-up analysis between GOSAT bias-corrected product and TCCON-based bias-corrected GOSAT-2 (Fig. R2). Because the GOSAT retrievals are relatively less sensitive to surface albedo and has been used as a reference or validation dataset in previous studies (Lorente et al., 2021; Balasus et al., 2023; Li et al., 2024; Fan et al., 2024), we used it as a complementary reference to evaluate the stability of the final target under broader albedo conditions. This analysis covered a surface albedo range of approximately 0–0.8. The TCCON-based bias-corrected GOSAT-2 showed a weaker dependency than the standard GOSAT-2 product, with R decreasing from -0.20 to -0.12.

We also evaluated the surface albedo dependency of TROPOMI harmonized to this target scale (Fig. R3). The dependencies observed in the standard and operational bias-corrected TROPOMI products were almost removed by applying ML-based harmonization. This result indicates that harmonization to the TCCON-based bias-corrected GOSAT-2 target scale effectively mitigated the surface-albedo-related bias in TROPOMI.

Therefore, although these additional analyses do not eliminate the intrinsic limitation of the TCCON-based approach, the site cross-validation within the TCCON albedo range and the satellite match-up analysis over a broader surface albedo range together provide complementary evidence that the TCCON-based bias-corrected GOSAT-2 using ML can serve as a reasonable common reference scale for multi-sensor harmonization.

Lines 309-312: "*This inter-sensor harmonization further assessed the major retrieval-parameter dependencies identified in the TCCON-based analysis over broader parameter ranges (Fig. 8). After ML-based harmonization, the SWIR surface albedo dependency of TROPOMI and the ΔPs dependency of GOSAT were nearly removed, indicating improved consistency with the ML-based bias-corrected GOSAT-2 scale.*"

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 LOSOCV strategy and additional satellite match-up analyses over broader surface albedo ranges, but independent validation under underrepresented retrieval conditions remains necessary."*

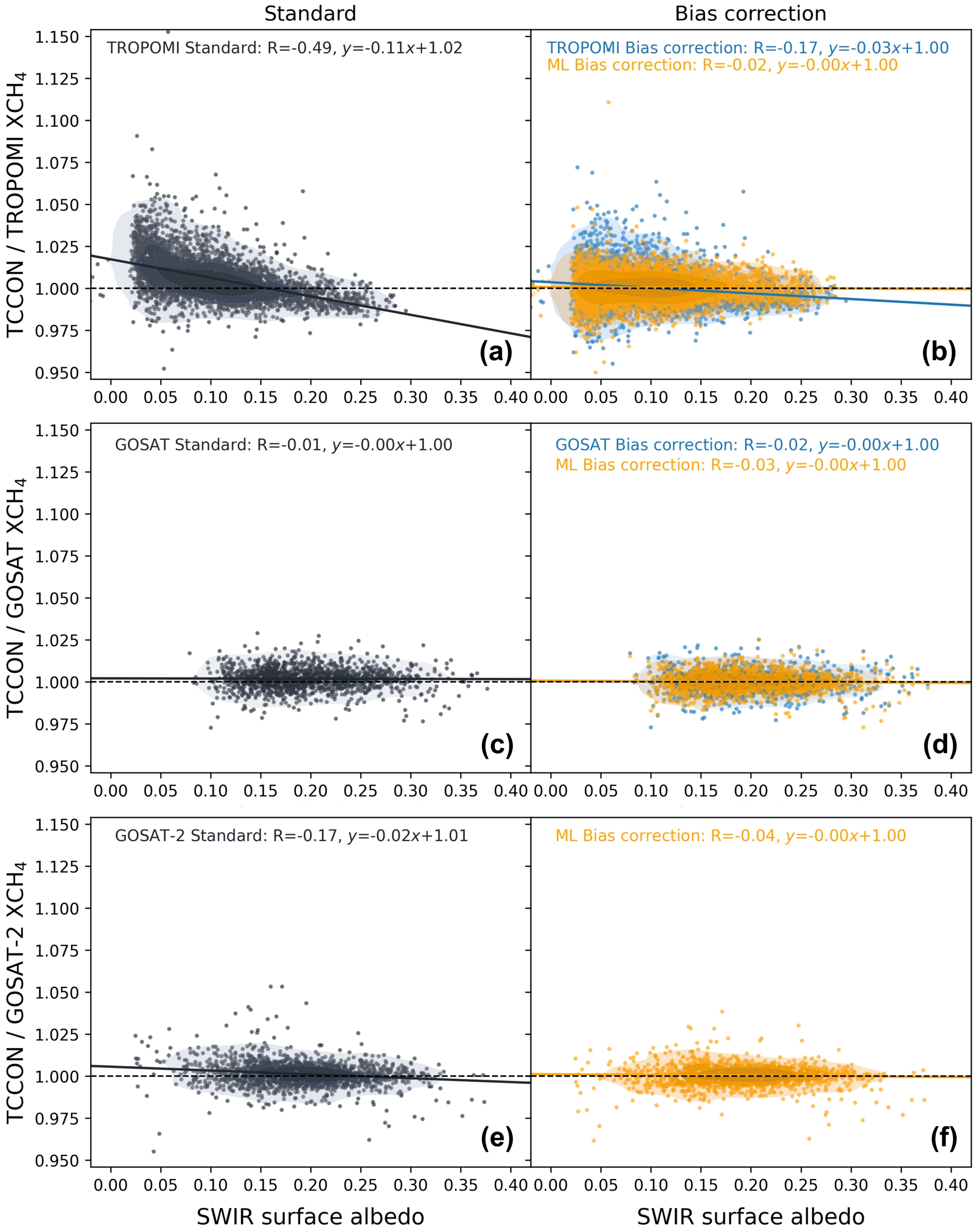

Figure R1. **SWIR surface albedo dependency of satellite XCH_4 products relative to TCCON XCH_4. Scatter density plots show the relationship between SWIR surface albedo and the ratio of satellite XCH_4 (TROPOMI, GOSAT, and GOSAT-2) to TCCON XCH_4. For each sensor, the left panel (a, c, e) shows the standard product, and the right panel (b, d, f) shows the bias-corrected product, including the operational bias-corrected product and the machine learning (ML)-based bias-corrected results for all sensors. The ML-based bias-corrected results are based on leave-one-site-out cross-validation, in which the validation site was excluded from model training.**

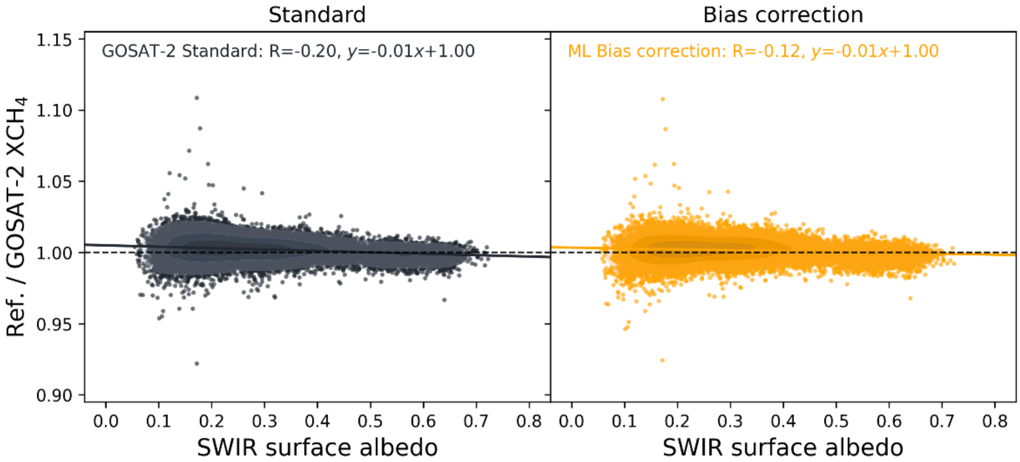

Figure R2. SWIR surface albedo dependency of GOSAT-2 XCH_4 relative to GOSAT bias-corrected XCH_4 (Reference). Scatter density plots show the ratio of GOSAT-2 XCH_4 to GOSAT bias-corrected XCH_4 as a function of SWIR surface albedo using GOSAT–GOSAT-2 satellite co-location samples.

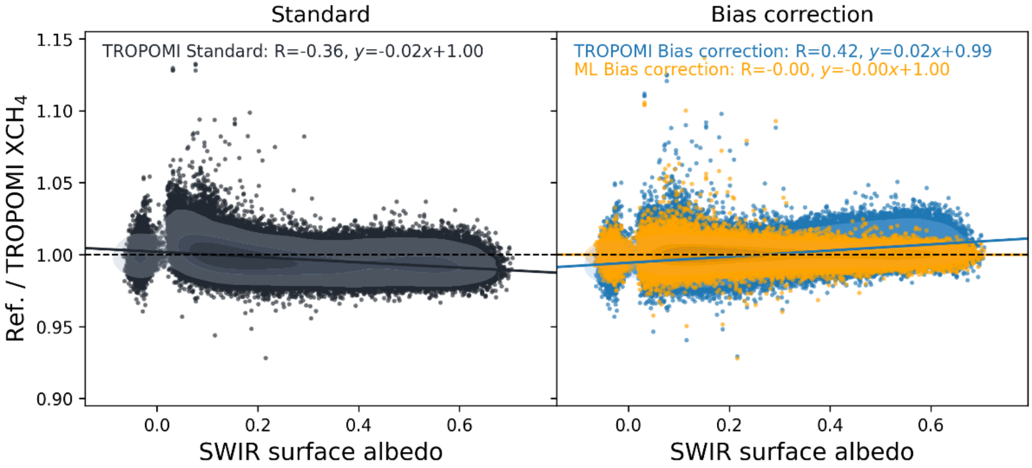

Figure R3. SWIR surface albedo dependency of TROPOMI XCH_4 relative to the final harmonization target. Scatter density plots show the ratio of TROPOMI XCH_4 to the final target, defined as the ML-based bias-corrected GOSAT-2 XCH_4, as a function of TROPOMI SWIR surface albedo. The comparison among the standard, operational bias-corrected, and ML-based harmonized TROPOMI products.

References:

Lorente, A., Borsdorff, T., Martinez-Velarte, M. C., & Landgraf, J. (2023). Accounting for surface reflectance spectral features in TROPOMI methane retrievals. *Atmospheric Measurement Techniques*, *16*(6), 1597-1608.

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.

Li, K., Bai, K., Jiao, P., Chen, H., He, H., Shao, L., ... & Chang, N. B. (2024). Developing unbiased estimation of atmospheric methane via machine learning and multiobjective programming based on TROPOMI and GOSAT data. *Remote Sensing of Environment*, *304*,

114039.

Fan, L., Wan, Y., & Dai, Y. (2024). Development of a multi-source satellite fusion method for XCH4 product generation in oil and gas production areas. *Applied Sciences*, *14*(23), 11100.

2. The authors have focused mostly on the delivery of the 0.1° × 0.1° product as the outcome of the study (as opposed to, for example, what their bias corrections tell us about shortcomings in any of the retrievals). Could the authors suggest some studies that could benefit from their data? In many cases, XCH_4 data alone is insufficient. In inverse modeling studies using chemical transport models, users would also need e.g. averaging kernel, prior profile, and pressure grid information (and might want data from all available sensors, not just the best one, in order to reduce random errors), though inverse modeling studies are not the only application of XCH_4 data.

➔ Thank you for this important comment. We agree that providing only the final gridded XCH_4 values may be limited for some downstream applications. In response, we expanded the publicly available dataset beyond the final daily 0.1° fused XCH_4 product to include the sensor-specific products used in the fusion. Specifically, the released dataset includes the TCCON-based bias-corrected GOSAT-2 product, the harmonized GOSAT-2-like TROPOMI and GOSAT products, and the source sensor identifier for the final fused product. These additional data allow users to trace which sensor contributed to each fused grid cell and to interpret sensor-specific differences and the fusion process more transparently.

Sections 4.1 and 4.2 also include analyses showing how the bias correction and harmonization results diagnose and mitigate residual bias structures in the individual satellite retrievals. For example, TROPOMI showed a strong dependency on SWIR surface albedo, while GOSAT showed a dependency on ΔPs. These retrieval-parameter-dependent biases were substantially reduced after ML-based correction and harmonization. Thus, the study does not only deliver a fused product but also provides bias-corrected and harmonized XCH_4 fields that address limitations of the operational satellite products.

We further added examples of downstream analyses that could benefit from the dataset. The TCCON-based bias-corrected GOSAT-2 product can serve as a reference-scale benchmark for future XCH_4 bias-correction or harmonization studies. The daily 0.1° fused XCH_4 product can support gap-filling studies, regional methane assessment, and hotspot identification. In addition, the corrected and harmonized sensor-specific products can be used to develop and evaluate more advanced fusion strategies. We clarified these dataset components and potential applications in the revised manuscript.

Lines 479-482: "*The bias-correction and harmonization analyses also revealed and mitigated sensor-specific residual bias structures, including the SWIR surface albedo dependency in TROPOMI and the ΔPs dependency in GOSAT. These results indicate that the proposed framework not only improves the final fused product but also provides diagnostic insight into retrieval-condition-dependent limitations of the individual satellite products.*"

Lines 489-492: "*The dataset can support downstream applications such as XCH_4 gap* *filling, multi-sensor intercomparison and harmonization, regional methane assessment, hotspot identification, and the development of advanced fusion strategies. The sensor-specific corrected and harmonized products can also serve as reference-scale benchmarks for future XCH_4 bias-correction studies"*

Lines 496–500: *"The publicly available dataset includes the daily globally harmonized fused XCH4 product at 0.1° spatial resolution for 2020–2023, individual products used in the fusion process (TCCON-based bias-corrected GOSAT-2, GOSAT-2-like harmonized TROPOMI and GOSAT), and source sensor identifiers indicating which satellite selected each valid observation in the fused product. All products are provided in HDF5 format and are available on Zenodo at* https://doi.org/10.5281/zenodo.20304047*."*

3. Have the authors found it necessary to account for differences in prior profiles and vertical sensitivities of the different instruments (TCCON, TROPOMI, GOSAT, GOSAT-2) when making comparisons?

➔ We appreciate this important methodological point. The final objective of this study is to generate a daily 0.1° fused XCH_4 product by integrating GOSAT, GOSAT-2, and TROPOMI. To achieve this, we first needed to evaluate the relative performance of the three satellite products against a common external reference, select a reference scale for fusion, and then harmonize the other sensors to that scale.

In this study, TCCON was used as a common reference for bias correction and reference-scale selection. We did not explicitly apply averaging-kernel smoothing in Step 1−2. Applying satellite-specific averaging kernels can make each satellite–TCCON comparison more profile-aware, but it also smooths the TCCON reference differently according to the prior profile and vertical sensitivity of each satellite. This can make the effective reference comparison sensor-specific, which is not fully consistent with our objective of comparing the relative performance of the three sensors within a common TCCON reference framework.

Nevertheless, as the reviewer noted, differences in prior profiles and vertical sensitivities can affect satellite–TCCON comparisons. We therefore performed an additional sensitivity test to quantify their potential impact. In this analysis, we applied the TCCON pressure-weighting and averaging-kernel smoothing approach described in Appendix A of Balasus et al. (2023) and the TCCON data-comparison guidance. Before comparison, the TCCON profile was adjusted to the satellite-specific prior profile and averaging-kernel sensitivity, and the resulting validation statistics were compared with those from the original comparison.

The results show that averaging-kernel smoothing produced only limited changes in the validation statistics across all sensors. Across all TCCON stations, RMSE changes remained within ±0.6 ppb and ΔR^2 within ±0.02 for all datasets (Fig. R4). Similar results were obtained for the independent test stations used in this study, with RMSE differences within ±0.2 ppb and ΔR^2 within ±0.01 across all sensors (Table R1).

Therefore, although prior profiles and vertical sensitivities clearly differ among TCCON, TROPOMI, GOSAT, and GOSAT-2, our sensitivity test indicates that their impact on the validation statistics is relatively small compared with the systematic retrieval biases addressed by the ML-based bias correction and harmonization framework.

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.

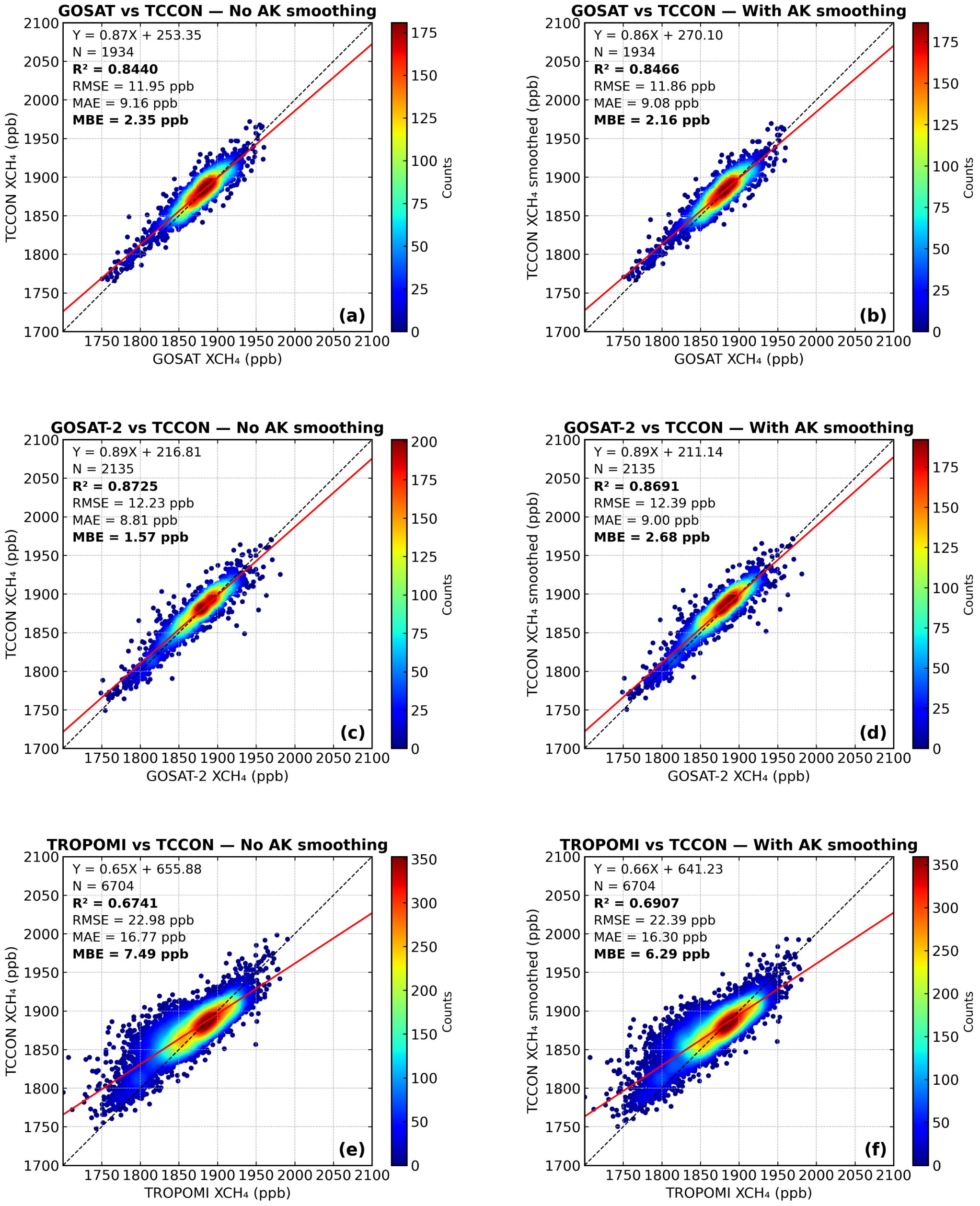

Figure R4. Comparison of satellite XCH_4 products against TCCON observations with and without averaging kernel (AK) smoothing. Panels (a), (c), and (e) show direct comparisons without AK smoothing,

while panels (b), (d), and (f) show comparisons after applying AK smoothing to TCCON observations. Results are shown for GOSAT, GOSAT-2, and TROPOMI, respectively. Each point represents an individual collocation pair and is colored by point density. The red line denotes the least-squares linear regression fit and the dashed black line indicates the 1:1 reference. Validation statistics including R^2, RMSE, MAE, and MBE are shown in each panel.

Table R1. Validation statistics (RMSE and MAE) for standard satellite XCH_4 products against TCCON at the three independent test stations (Edwards01, Garmisch01, Xianghe01), comparing direct comparison (no average kernel (AK) smoothing) and AK-smoothed TCCON results

	RMSE(Direct) ppb	RMSE(AK) ppb	MAE(Direct) ppb	MAE(AK) ppb	N
GOSAT	11.20	11.15	8.53	8.46	629
GOSAT-2	11.05	11.20	7.95	8.17	567
TROPOMI	17.48	17.43	13.48	13.49	1598

Minor Comments:

1. Line 39: misspelled reference?

   ➔ We thank the reviewer for pointing this out. The reference spelling has been corrected to "Saunois et al. (2025)" in line 36.

   Lines 35-36: "*Recent observations show that CH_4 levels have been rising at an unprecedented rate, with 2020–2022 marking the fastest growth since systematic monitoring began (Saunois et al., 2025).*"

2. Line 46: consider defining XCH_4

   ➔ We thank the reviewer for this suggestion. XCH_4 has been defined at its first occurrence in the revised manuscript (Lines 43–44) as 'the column-averaged dry-air mole fraction of atmospheric methane (XCH_4)

   Lines 40-41: "*Accurate quantification of the column-averaged dry-air mole fraction of atmospheric methane (XCH_4) is therefore essential for identifying emission sources and evaluating progress toward climate goals.*"

3. Line 61: Lorente et al. (2021) uses a small-area approximation, not TCCON

   ➔ We thank the reviewer for this correction. The original text incorrectly implied that Lorente et al. (2021) used TCCON as a reference for bias correction. The revised text now accurately distinguishes between TCCON-based statistical regression approaches and the small-area approximation method.

   Lines 57-59: "*Previous bias correction studies have relied on statistical regression approaches that account for factors such as surface albedo, aerosol loading, and viewing* *geometry (Inoue et al., 2016), as well as small-area approximation methods that derive surface-albedo-related biases directly from satellite observations (Lorente et al., 2021)*"

4. Line 99: please check the 30 times number (cf. Table 1 in Jacob et al., 2022)

   ➔ We thank the reviewer for this insightful reference. Upon checking Table 1 in Jacob et al. (2022) and Noël et al. (2021), we confirmed that the spectral resolution is 0.06 nm for GOSAT and GOSAT-2 and 0.25 nm for TROPOMI. We found that the "~30 times" statement was incorrectly calculated. We therefore revised the text to state approximately fourfold difference and explicitly added the spectral resolution of each sensor to Table 2.

   Lines 105-108: "*GOSAT and GOSAT-2 provide high-precision measurements with spectral resolution approximately four times finer than that of TROPOMI (Kuze et al., 2009, 2016; Suto et al., 2021; Jacob et al., 2022; Noël et al., 2021), whereas TROPOMI offers daily global mapping, which increases data density despite higher susceptibility to atmospheric interference (Hu et al., 2018).*"

   Reference:
   Jacob, D. J., Varon, D. J., Cusworth, D. H., Dennison, P. E., Frankenberg, C., Gautam, R., ... & Duren, R. M. (2022). Quantifying methane emissions from the global scale down to point sources using satellite observations of atmospheric methane. *Atmospheric Chemistry and Physics*, *22*(14), 9617-9646.

   Noël, S., Reuter, M., Buchwitz, M., Borchardt, J., Hilker, M., Bovensmann, H., ... & Warneke, T. (2021). XCO 2 retrieval for GOSAT and GOSAT-2 based on the FOCAL algorithm. *Atmospheric Measurement Techniques*, *14*(5), 3837-3869.

5. Line 224: extra "and" at the end of the sentence

   ➔ Thank you for pointing this out. We have removed the extra "and" at the end of the sentence.

   Lines 231-232: "*Model performance was evaluated using three standard metrics: coefficient of determination (R^2), root mean square error (RMSE) and mean absolute error (MAE) (Equations S1–S3).*"

6. Line 253: specify northern-hemisphere/boreal summer and autumn

   ➔ Thank you for the suggestion. To avoid ambiguity and ensure consistency with the figure, we revised the seasonal terms to the corresponding month ranges.

   Lines 261-263: "*Across sensors, the standard products showed negative biases and broad interquartile ranges (IQRs) throughout the year, with strong negative shifts from June to September (Fig. 5). The seasonal aggregation confirmed that this negative bias was most pronounced in June-July-August (JJA).*"

7. Figure 4: metrics (e.g. RMSE in subplots c,f,h) look much more optimistic than the LOSOCV column of Table S6

    ➔ Thank you for pointing this out. Figure 4 and Table S5 are both based on the leave-one-site-out cross-validation (LOSOCV) results, but the metrics were calculated at different aggregation levels and were intended to evaluate different aspects of performance. The LOSOCV metrics in Table S5 were calculated using individual satellite–TCCON co-location samples, and therefore reflect sample-level accuracy, including the effects of observation-level errors and variability, as well as uneven sample numbers across sites, as shown in the Figure 4 legend.

    In contrast, Figure 4 was constructed after aggregating the LOSOCV predictions and TCCON observations at each TCCON site. Each point represents the site-mean XCH_4, and the error bars indicate the standard deviation of the co-location samples within each site, representing within-site variability. Thus, Figure 4 was intended to evaluate how effectively the bias correction reduces mean satellite–TCCON differences at the site level and how consistently this improvement is achieved across different concentration ranges and sites, rather than to assess sample-level prediction accuracy.

    Because random sample-level errors and short-term variability are averaged during site-level aggregation, the performance metrics in Figure 4 can appear more optimistic than the sample-level LOSOCV metrics in Table S6. Therefore, the two results are not contradictory but represent different aspects of performance. We have clarified this distinction in the revised manuscript text.

    Lines 236-237: "*In Figure 4, LOSOCV results and TCCON observations were aggregated at each TCCON site to examine site-level mean bias, within-site variability, and station-to-station consistency.*"

    Line 243: "*In contrast, the ML-based correction tightened site-level agreement for all sensors, reducing both RMSE and MAE.*"

8. Figure 5: why not leave one season out in your cross validation if you are going to plot like this?

    ➔ Thank you for pointing this out. The validation was performed using leave-one-month-out cross-validation (LOMOCV), but the previous version of Figure 5 presented only seasonally aggregated results, which may have obscured the distinction between the validation unit and the visualization unit. We revised Fig. 5 to first show the monthly bias distributions from the LOMOCV results and then the seasonally aggregated bias distributions from the same LOMOCV outputs. This revision clarifies the monthly validation basis while also allowing the seasonal bias patterns to be interpreted more intuitively. The corresponding text and figure caption were revised accordingly.

    Lines 261-269: "*Across sensors, the standard products showed negative biases and broad interquartile ranges (IQRs) throughout the year, with strong negative shifts from June to September. The seasonal aggregation confirmed that this negative bias was most pronounced in June-July-August (JJA). This seasonal pattern is consistent with known* *sensitivities of SWIR-based XCH_4 retrievals to surface albedo and atmospheric scattering by aerosols and cirrus (Inoue et al., 2016; Hu et al., 2016; Lorente et al., 2021; Oshio et al., 2020). These factors can modify the effective light path and vary seasonally with vegetation, humidity, cloud conditions, and solar geometry. Operational bias correction shifted the median bias toward zero but still exhibited larger seasonal dispersion than the ML-based correction. The ML-based correction maintained seasonal medians near zero and reduced both the IQRs and non-outlier spread, with most IQRs remaining within ±8 ppb, indicating more stable performance under seasonal variability.*"

References:

Inoue, M., Morino, I., Uchino, O., Nakatsuru, T., Yoshida, Y., Yokota, T., ... & Tanaka, T. (2016). Bias corrections of GOSAT SWIR XCO 2 and XCH 4 with TCCON data and their evaluation using aircraft measurement data. *Atmospheric Measurement Techniques*, *9*(8), 3491-3512.

Hu, H., Hasekamp, O., Butz, A., Galli, A., Landgraf, J., Aan de Brugh, J., ... & Aben, I. (2016). The operational methane retrieval algorithm for TROPOMI. *Atmospheric Measurement Techniques*, *9*(11), 5423-5440.

Lorente, A., Borsdorff, T., Butz, A., Hasekamp, O., Schneider, A., Wu, L., ... & Landgraf, J. (2021). Methane retrieved from TROPOMI: improvement of the data product and validation of the first 2 years of measurements. *Atmospheric Measurement Techniques*, *14*(1), 665-684.

Oshio, H., Yoshida, Y., Matsunaga, T., Deutscher, N. M., Dubey, M., Griffith, D. W., ... & Wunch, D. (2020). Bias correction of the ratio of total column ch4 to co2 retrieved from gosat spectra. *Remote Sensing*, *12*(19), 3155.

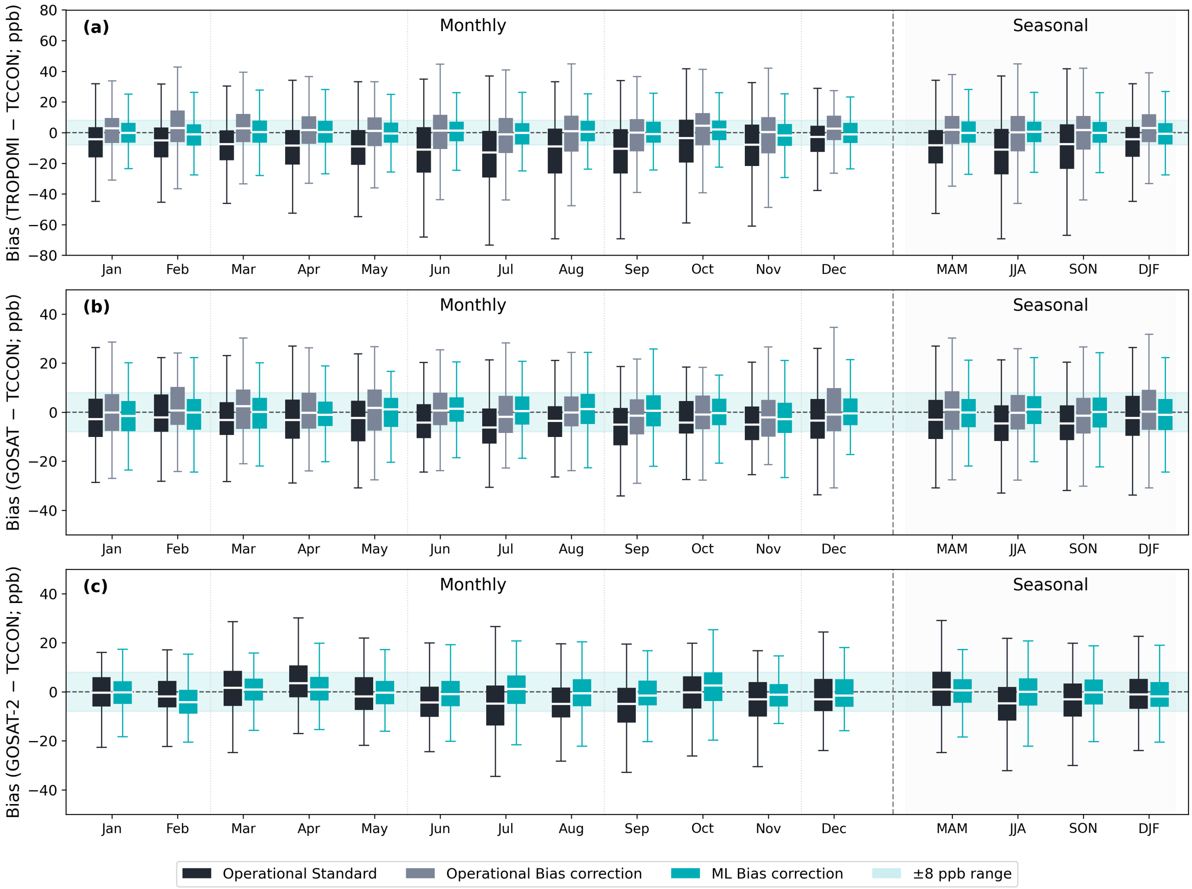

Figure 5. Monthly and seasonally aggregated XCH_4 bias distributions under leave-one-month-out cross-validation (LOMOCV). Box-and-whisker plots show bias distributions (Satellite − TCCON) for each calendar month and their seasonal aggregations for (a) TROPOMI, (b) GOSAT, and (c) GOSAT-2. The compared datasets include the uncorrected operational standard products, operational bias-corrected products, and ML-based bias-corrected results. Boxes indicate the interquartile range (IQR) with the median shown as a horizontal white line. The shaded band marks the ±8 ppb range around zero bias.

9. SI: double-check (e.g., no bolds in Tables S6 and S7, no Table S8, etc.)

   ➔ Thank you for pointing this out. We checked the Supplementary Information, added the missing bold formatting in Tables S5 and S6, and corrected the table numbering from Table S6 to Table S7.

10. Line 525: Jacob et al. (2022) no longer is in Discussion

    ➔ Thank you for checking this. Jacob et al. (2022) is still cited in the Introduction and Datasets sections. We have checked the citation locations accordingly.