Reply to Referee #3

Dear reviewer:

Thank you for reviewing our manuscript and providing constructive comments. Below, the comments are reproduced in black and our point-by-point answers are in blue. Each substantive response is followed by the corresponding revised text and/or figures and tables. Revised passages and figure captions are quoted from the revised manuscript and Supplement. Page and line numbers refer to the revised manuscript; S-numbered figures and tables refer to the revised Supplement.

Sincerely,

Yike Wang

On behalf of all co-authors

Wuhan University

General comment

This manuscript presents an intercomparison of DQ-1/ACDL XCO_2 with OCO-2, OCO-3, GOSAT, and GOSAT-2. The topic is relevant, and the comparison of active and passive XCO_2 observations is useful. My main concern is that several choices in sample selection and collocation are not sufficiently justified, which makes it difficult to assess the robustness of the reported agreement. I believe the paper makes a valuable contribution and is suitable for publication after revision.

ANSWER: We thank the reviewer for identifying the need to demonstrate robustness more explicitly. We have added spatial-threshold and temporal-window sensitivity tests, clarified the rationale for the 2°/4 h framework, quantified the CAMS-based column-definition correction, applied the official GOSAT-2 empirical correction, and expanded the TCCON sampling and uncertainty analysis. We have also narrowed the conclusions so that they apply to the coherent and well-sampled comparison regions tested here.

Specific comments

Comment 1

The analysis is restricted to the top 10% of grid cells based on the coherence and sampling score. Why was the 90th-percentile threshold chosen? Because this selection favors relatively homogeneous, well-sampled regions, it would be useful to show whether the reported bias and RMSE change when a less restrictive threshold is used.

ANSWER: P90 was adopted as a conservative reference criterion to prioritize background-field coherence and adequate multi-satellite sampling; it was not empirically optimized and is not intended to represent all global scenes. The percentile is calculated separately for eligible land and ocean grid cells. We therefore repeated the comparisons with P80 and P70 (cases S1 and S2), retaining approximately the upper 20 % and 30 % of eligible grid cells while keeping the ≤4 h temporal criterion unchanged. N increases for nearly all pairs, but the principal R, bias, and RMSE patterns remain close to the Main case. This indicates that the reported consistency statistics are relatively insensitive to the tested selection criteria, while not implying equivalence in strongly heterogeneous or poorly sampled scenes excluded by the framework.

Changes in the manuscript:

Sect. 2.2.1, pp. 10–11, lines 279–284.

> **Revised text:** "The P90 threshold was adopted as a conservative reference criterion. It prioritizes grid cells with relatively high background-field coherence and sufficient multi-satellite sampling. This choice was not intended to imply that P90 is an empirically optimized threshold. To test the sensitivity to this choice, we repeated the inter-satellite comparisons using P80 and P70 thresholds. These thresholds retain approximately the upper 20 % and 30 % of eligible land and ocean grid cells, respectively. All other processing and collocation procedures were kept unchanged. The sensitivity cases and results are presented in Sect. A2."

Sect. 3.3, p. 18, lines 483–489.

> **Revised text:** "We further assessed the robustness of the pairwise statistics using different spatial-selection thresholds and temporal-collocation windows (Sect. A2).

Using P80 and P70 instead of P90 increased the number of collocations (N) for nearly all satellite pairs because more comparison grid cells were included. However, the corresponding R, bias, and RMSE values remained close to those of the Main case. When the temporal window was narrowed from 4 h to 3, 2, and 1 h, N decreased, especially for satellite pairs with relatively sparse sampling. Despite these changes in sample size, R, bias, and RMSE changed only slightly. These results indicate that the main inter-satellite consistency statistics are robust to reasonable changes in the spatial-selection and temporal-collocation criteria."

Table A1. Sensitivity cases for the spatial-selection and temporal-collocation criteria used in the satellite XCO_2 intercomparison.

Case	Final-score threshold	Approx. retained fraction	Temporal window
Main	P90	Top 10%	≤4 h
S1	P80	Top 20%	≤4 h
S2	P70	Top 30%	≤4 h
S3	P90	Top 10%	≤3 h
S4	P90	Top 10%	≤2 h
S5	P90	Top 10%	≤1 h

P90, P80, and P70 denote the 90th, 80th, and 70th percentiles of the final score, respectively, calculated separately for land and ocean grid cells. S1–S2 test the spatial-selection criterion, whereas S3–S5 test the temporal-collocation criterion

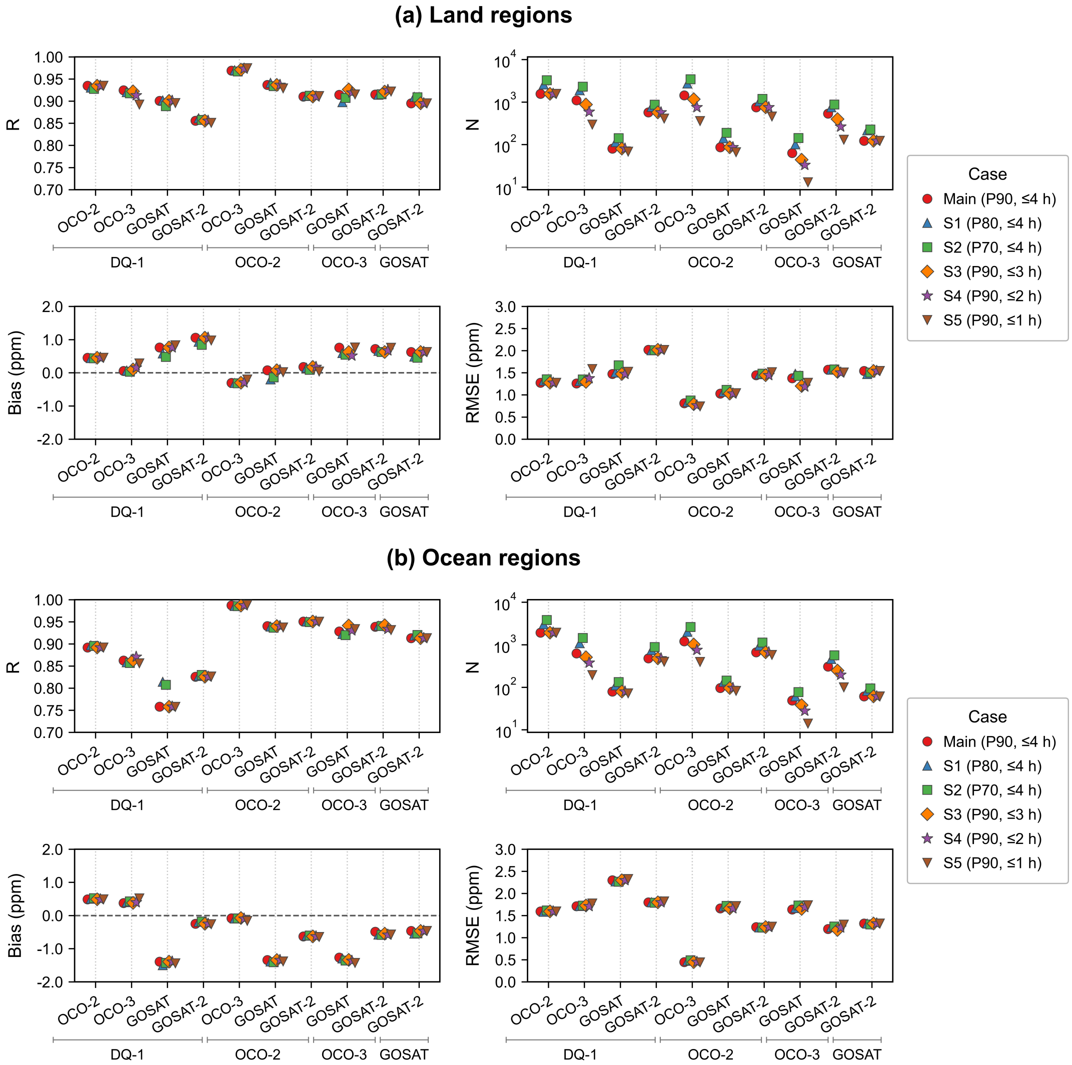

Figure A3. Sensitivity of satellite XCO_2 intercomparison statistics to the spatial-selection and temporal-collocation criteria over (a) land and (b) ocean regions. The four panels for each surface type show the correlation coefficient (R), number of collocations (N), mean bias, and RMSE for the six sensitivity cases defined in Table A1. Satellite pairs are organized by the horizontal brackets: the satellite name below each bracket denotes the first product in the pair, while the individual satellite labels above the bracket denote the second product. Bias is defined as XCO_2(first product) − XCO_2(second product). Main denotes the reference configuration (P90, ≤4 h); S1–S2 relax the spatial-selection threshold, whereas S3–S5 progressively tighten the temporal-collocation window.

Comment 2

The comparison uses 2° × 2° daily averages with a maximum time difference of 4 h. Please clarify the rationale for these choices, especially the 4 h threshold. A simple test using a smaller spatial or temporal window would help show how sensitive the results are to the collocation criteria.

ANSWER: The 2° grid provides a common aggregation scale for sensors with different footprints and sampling densities. It also contains up to four native 1° CAMS cells, which are needed for the within-cell coherence assessment. Simply reducing the grid to 1° would change that assessment rather than test only the collocation tolerance. We therefore retained 2° and performed the suggested temporal sensitivity test.

The 4 h threshold balances observation-time mismatch against the availability of collocations among sensors with different sampling patterns. Taylor et al. (2023) also used a 4 h time criterion for OCO-2/OCO-3 comparisons. In our implementation, the threshold applies to the difference between the two daily mean observation times; it is not a guarantee that every individual sounding in the two aggregates is within 4 h of every other sounding.

Cases S3–S5 tighten this threshold to 3, 2, and 1 h, respectively, with P90 fixed. Figure A3 shows that sample size decreases, particularly for sparsely sampled pairs, while R, bias, and RMSE change comparatively little. We have added this rationale and sensitivity test. We acknowledge that the selected aggregation and temporal-collocation framework may still introduce residual representativeness differences.

Changes in the manuscript :

Sect. 2.2.2, p. 11, lines 295–304.

> **Revised text:** "The 2° aggregation was chosen as a common analysis scale for sensors with different footprint sizes and sampling densities. It is also consistent with the CAMS-based coherence analysis, because each 2° grid cell contains up to four native 1° CAMS grid cells. A finer 1° grid would therefore not preserve the same within-cell coherence assessment. The 4 h temporal window was selected to balance temporal representativeness with the availability of valid collocations among satellites with different sampling characteristics. A comparable 4 h criterion has also been used in direct OCO-2/OCO-3 XCO_2 comparisons (Taylor et al., 2023). We therefore retained the 2° grid and tested temporal windows of 3, 2, and 1 h while keeping the P90 spatial criterion fixed. The sensitivity cases and results are presented in Sect. A2. Unless otherwise stated, the inter-satellite comparison results presented in the main text are based on the Main case, which uses the P90 spatial-selection threshold and a maximum temporal difference of 4 h."

Comment 3

CAMS is involved in the region screening, column-definition correction, and nighttime DQ-1 comparison. How large is the column-definition correction in practice? Showing the comparison statistics before and after this correction would make its influence on the results clearer.

ANSWER: We have added Fig. S1 to show the spatial distributions of the CAMS-derived DQ-1-minus-passive column-definition differences and Table S3 to report R, bias, and RMSE before and after correction for every satellite pair over land and ocean. The pair-mean DQ-1-minus-satellite definition differences are +0.44 ppm for OCO-2, +0.44 ppm for OCO-3, +1.30 ppm for GOSAT, and +0.49 ppm for GOSAT-2. The correction changes R by no more than about 0.01 in most comparisons but can materially change bias and RMSE, demonstrating that the observation-operator term is relevant without uniformly improving every comparison.

For example, DQ-1–OCO-2 bias changes from 0.92 to 0.45 ppm over land and from 0.91 to 0.49 ppm over ocean, while RMSE changes from 1.54 to 1.27 ppm and from 1.79 to 1.59 ppm, respectively. For DQ-1–GOSAT, land RMSE decreases from 2.30 to 1.47 ppm whereas ocean RMSE increases from 1.86 to 2.30 ppm. We therefore interpret the correction as a model-assisted harmonization of column definitions, not as a generic bias correction or guarantee of improved agreement.

We now distinguish CAMS-based comparison-region selection, column-definition harmonization, and model-reference diagnostics. The correction removes the modeled operator difference, rather than forcing observations toward CAMS or toward one another. Remaining differences still include retrieval, sampling, and reference-profile uncertainty; the results are therefore not fully independent of the model-assisted framework.

Changes in the manuscript:

Sect. 3.3, p. 17, lines 457–463.

Revised text: "Following the independent TCCON evaluation, we next examine the relative consistency among satellite XCO_2 products over the coherent and well-sampled comparison regions. Figure 4 summarizes the pairwise consistency statistics for the five satellite XCO_2 products over the selected comparison regions. Figure S1 shows the spatial distributions of the CAMS-derived column-definition differences, and Table S3 reports the pairwise statistics before and after the column-definition correction. The mean DQ-1-minus-OCO-2, OCO-3, GOSAT, and GOSAT-2 column-definition differences are 0.44, 0.44, 1.30, and 0.49 ppm, respectively. The correction changes the correlation coefficients by no more than 0.01 in most comparisons, while the bias and RMSE can change more noticeably."

Sect. 2.2.3, p. 12, lines 334–336.

Revised text: "This correction only accounts for the theoretical difference produced when the same CO_2 profile is represented under different satellite column definitions. CAMS is used only to estimate column-definition differences and is not treated as a truth reference for correcting satellite biases."

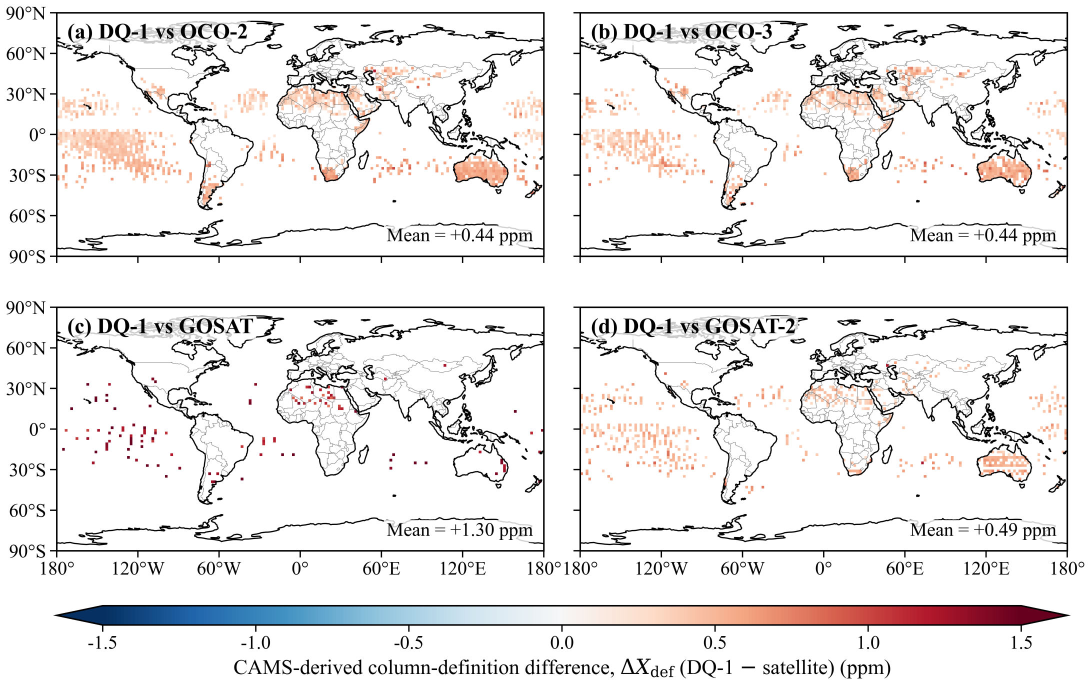

Figure S1. Spatial distributions of the CAMS-derived XCO_2 column-definition differences between DQ-1 and (a) OCO-2, (b) OCO-3, (c) GOSAT, and (d) GOSAT-2 for the collocated samples. The column-definition difference is defined as XCO_2 (DQ-1 definition) − XCO_2 (satellite definition), and the mean value for each satellite pair is shown in the lower-right corner of each panel.

Table S3. Pairwise consistency statistics for satellite XCO_2 products before and after CAMS-based column-definition correction over the selected comparison regions. Bias is defined as the y-axis product minus the x-axis product.

Y-axis product	X-axis product	Surface type	N	Before correction			After correction		
				R	Bias (ppm)	RMSE (ppm)	R	Bias (ppm)	RMSE (ppm)
DQ-1	OCO-2	Land	1559	0.93	0.92	1.54	0.93	0.45	1.27
		Ocean	1922	0.89	0.91	1.79	0.89	0.49	1.59
DQ-1	OCO-3	Land	1099	0.92	0.51	1.38	0.92	0.06	1.26
		Ocean	628	0.87	0.81	1.87	0.86	0.38	1.71
DQ-1	GOSAT	Land	80	0.91	1.97	2.30	0.90	0.76	1.47
		Ocean	79	0.77	0.01	1.86	0.76	-1.39	2.30
DQ-1	GOSAT-2 (original)	Land	569	0.87	-2.18	2.80	0.87	-2.66	3.19
		Ocean	479	0.83	-2.87	3.41	0.82	-3.36	3.84
DQ-1	GOSAT-2 (corrected)	Land	569	0.85	1.54	2.31	0.86	1.06	2.02
		Ocean	479	0.83	0.24	1.80	0.83	-0.25	1.80
OCO-2	OCO-3	Land	1435	0.97	-0.32	0.82	0.97	-0.30	0.81
		Ocean	1203	0.99	-0.09	0.45	0.99	-0.08	0.45
OCO-2	GOSAT	Land	86	0.94	0.58	1.14	0.94	0.08	1.03
		Ocean	96	0.93	-0.69	1.23	0.94	-1.34	1.66
OCO-2	GOSAT-2 (original)	Land	754	0.91	-3.42	3.78	0.91	-3.48	3.82
		Ocean	667	0.94	-3.69	3.92	0.94	-3.75	3.97
OCO-2	GOSAT-2 (corrected)	Land	754	0.91	0.23	1.45	0.91	0.17	1.45
		Ocean	667	0.95	-0.57	1.24	0.95	-0.63	1.24
OCO-3	GOSAT	Land	63	0.92	1.29	1.72	0.91	0.76	1.38
		Ocean	49	0.92	-0.62	1.27	0.93	-1.27	1.64
OCO-3	GOSAT-2 (original)	Land	534	0.93	-2.85	3.24	0.93	-2.90	3.29
		Ocean	307	0.91	-3.63	3.91	0.92	-3.70	3.97
OCO-3	GOSAT-2 (corrected)	Land	534	0.92	0.77	1.58	0.92	0.72	1.57

Y-axis product	X-axis product	Surface type	N	Before correction			After correction		
				R	Bias (ppm)	RMSE (ppm)	R	Bias (ppm)	RMSE (ppm)
		Ocean	307	0.94	-0.42	1.18	0.94	-0.49	1.20
GOSAT-2 (original)	GOSAT	Land	122	0.89	5.09	5.32	0.88	4.68	4.94
		Ocean	61	0.95	3.28	3.45	0.95	2.79	2.99
GOSAT-2 (corrected)	GOSAT	Land	122	0.90	1.04	1.74	0.89	0.63	1.54
		Ocean	61	0.91	0.02	1.26	0.91	-0.46	1.32

Comment 4

GOSAT-2 shows a large positive bias relative to the other products, but the manuscript also notes that this product has not been empirically bias corrected. This makes the direct comparison with bias-corrected OCO products somewhat difficult to interpret. The discussion should clarify this distinction, or include bias-corrected GOSAT-2 results if available.

ANSWER: We agree and have incorporated empirically corrected GOSAT-2 results. The official NIES V02.20/V02.21 land- and ocean-specific corrections are applied to each quality-controlled retrieval before daily 2° aggregation and collocation; the CAMS-based column-definition correction is applied afterward to the collocated samples. V02.20 is used through February 2024 and V02.21 from March 2024 onward.

For the same 880 GOSAT-2–TCCON collocations, the empirical correction changes R from 0.91 to 0.92, bias from +3.56 to +0.01 ppm, and RMSE from 3.98 to 1.54 ppm. The principal inter-satellite comparisons now use corrected GOSAT-2, while the original and corrected results are both retained in Fig. 3d and Tables S2–S3. We also now state that the official correction coefficients were derived using TCCON GGG2020 as the reference; consequently, the corrected GOSAT-2–TCCON comparison is not fully independent of TCCON. This calibration dependence is explicitly noted in Sect. 4.2.

We now interpret the large original offset as a systematic component addressed by the empirical correction, rather than as a direct measure of sensor performance.

Changes in the manuscript:

Sect. 2.1.2, p. 5, lines 132–147.

Revised text: "OCO-2 and OCO-3 are NASA Orbiting Carbon Observatory missions designed to retrieve XCO2 from reflected solar spectra using three-channel high-resolution grating spectrometers and full-physics retrieval algorithms (Crisp et al., 2017; Eldering et al., 2017; O'Dell et al., 2018; Taylor et al., 2023). GOSAT and GOSAT-2 carry the Thermal And Near-infrared Sensor for carbon Observation Fourier Transform Spectrometer (TANSO-FTS) and its successor TANSO-FTS-2, respectively. Their instrument characteristics, retrieval algorithms, and XCO2 products have been described in previous studies (Kuze et al., 2009, 2016; Noël et al., 2021, 2022; Suto et al., 2021; Yoshida et al., 2023). For GOSAT-2, the official land- and ocean-specific empirical bias corrections were applied to the individual XCO2 retrievals following the V02.20 and V02.21 release notes (NIES GOSAT-2 Project, 2026a, b). The official correction coefficients were derived using TCCON GGG2020 as the reference. GOSAT-2 V02.20 was used through February 2024 and V02.21 from March 2024 onward. The empirical bias correction was applied to each quality-controlled individual retrieval before daily 2° aggregation and collocation, whereas the CAMS-based column-definition correction was subsequently applied to the collocated samples. The passive satellite products are provided as individual soundings and provide the XCO2 observations for comparison with DQ-1. The corresponding quality-control criteria are summarized in Table 2, and the spatiotemporal collocation and column-definition correction are described in Sects. 2.2.2 and 2.2.3. Although all products report XCO2, their effective column definitions are not identical because the passive products use product-specific prior profiles, pressure weights, and column averaging kernels, whereas DQ-1 ACDL uses its weighting function and effective pressure layers. These differences are accounted for in Sect. 2.2.3."

Sect. 3.2, pp. 15–16, lines 422–428.

Revised text: "For GOSAT-2, the uncorrected XCO_2 product shows a pronounced positive bias relative to TCCON, with R = 0.91, a mean bias of 3.56 ppm, and an RMSE of 3.98 ppm (gray symbols in Fig. 3d). Similar positive biases of GOSAT-2 XCO_2 relative to TCCON have also been reported in previous studies (Ji et al., 2024; Yang et al., 2025). After applying the official empirical bias correction to the pixel-level GOSAT-2 XCO_2 data, the mean bias decreases to 0.01 ppm and the RMSE to 1.54 ppm, while R changes only slightly to 0.92 (colored symbols in Fig. 3d). The substantial reductions in bias and RMSE indicate that the pronounced overestimation in the uncorrected product is largely associated with systematic biases addressed by the empirical correction. After correction, GOSAT-2 shows good agreement with TCCON.”

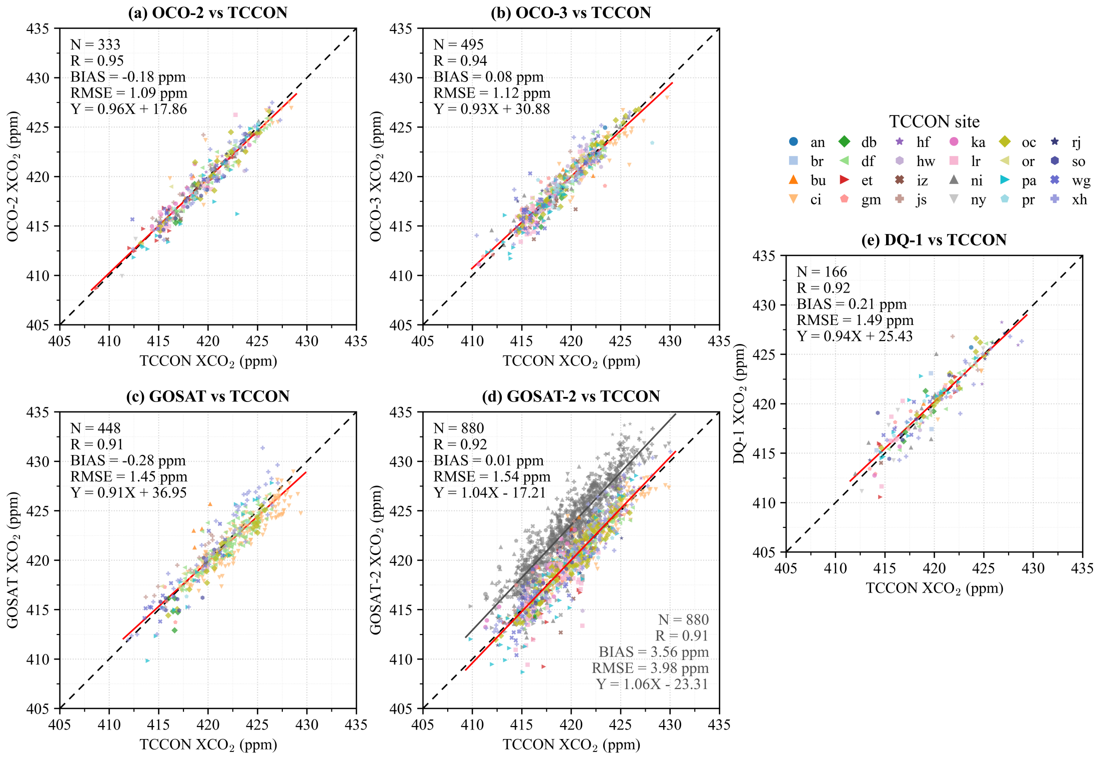

Figure 3. Comparison of satellite XCO_2 products with TCCON observations. (a–e) OCO-2, OCO-3, GOSAT, GOSAT-2, and DQ-1 XCO_2 compared with collocated TCCON XCO_2, respectively. Different colors and symbols denote individual TCCON sites. In panel (d), colored and gray symbols denote empirically bias-corrected and uncorrected GOSAT-2 XCO_2, respectively, with the corresponding linear least-squares fits shown by the red and gray solid lines. Red solid lines in panels (a–c, e) indicate linear least-squares fits, and black dashed lines indicate the 1:1 relationship. The TCCON site abbreviations are defined in Table A2

Table S2. Satellite–TCCON validation statistics, site-level variability, and 95 % confidence intervals. The 95 % confidence intervals (CIs) of the pooled bias and RMSE are defined by the 2.5th and 97.5th percentiles of 20 000 TCCON-site cluster-bootstrap realizations. Interstation bias is the standard deviation of the site-specific mean biases, and site-mean RMSE is the arithmetic mean of the site-specific RMSE values, with equal weight assigned to each contributing site.

Product	N	Bias (95 % CI) (ppm)	Interstation bias (ppm)	RMSE (95 % CI) (ppm)	Site-mean RMSE (ppm)
OCO-2	333	−0.18 [−0.35, 0.00]	0.60	1.09 [0.89, 1.31]	1.04
OCO-3	495	0.08 [−0.10, 0.28]	0.56	1.12 [0.98, 1.27]	1.12

GOSAT	448	−0.28 [−0.72, 0.60]	1.40	1.45 [1.21, 1.85]	1.60
GOSAT-2 original	880	3.56 [3.24, 3.88]	1.19	3.98 [3.75, 4.27]	4.03
GOSAT-2 corrected	880	0.01 [−0.20, 0.26]	0.97	1.54 [1.33, 1.81]	1.70
DQ-1	166	0.21 [−0.08, 0.48]	0.73	1.49 [1.25, 1.74]	1.40

Comment 5

All satellite products are further restricted to 380–480 ppm after the product-specific quality screening. What is the reason for applying this additional range filter? Please also report the number or fraction of observations removed for each product.

ANSWER: Thank you for requesting this clarification. After product-specific quality screening, the additional 380–480 ppm filter excluded no retrievals from any of the five products: DQ-1 ACDL, OCO-2, OCO-3, GOSAT, and GOSAT-2 each had 0 additional retrievals removed (0%). This range restriction was therefore redundant for the datasets used in this study. We have removed it from the revised analysis and deleted the "Valid XCO_2 range" column from Table 2. Removing this restriction leaves the retained observations, collocation counts, and comparison statistics unchanged.

Changes in the manuscript :

Table 2. Quality-control criteria for the satellite XCO_2 products.

Satellite product	Quality-control and surface-type screening
DQ-1 ACDL XCO_2	QAflag = 1 or 5; Asc/Desc processed separately
OCO-2 XCO_2	xco2_quality_flag = 0; land and ocean-glint observations retained
OCO-3 XCO_2	xco2_quality_flag = 0; land and ocean-glint observations retained
GOSAT XCO_2	totalPreScreeningResult = 0; totalPostScreeningResult = 0; Land: landFraction ≥ 80%; ocean-glint: landFraction ≤ 20% and sunglint flag = 1
GOSAT-2 XCO_2	xco2_quality_flag = 0; Land: landFraction ≥ 80%; ocean-glint: landFraction ≤ 20% and sunglintFlag = 1

Comment 6

The TCCON comparison provides an important independent evaluation, but more information on the collocations is needed. Please report the sample size for each satellite and indicate, if possible, whether a small number of stations dominates the pooled statistics. Some estimate of uncertainty for the reported bias and RMSE would also be useful.

ANSWER: We have expanded the TCCON evaluation to quantify both sample concentration and uncertainty. Tables S1 and S2 now report the number of collocations by site, the total number of collocations and contributing sites, the largest-site and top-three-site shares, pooled bias and RMSE, interstation bias, site-mean RMSE, and 95 % confidence intervals. The confidence intervals were estimated from 20,000 TCCON-site cluster-bootstrap realizations, in which sites were resampled with replacement and all collocations from each selected site were retained together. This approach preserves within-site dependence and quantifies sampling uncertainty associated with the finite station network rather than retrieval uncertainty for individual satellite observations.

For DQ-1, 166 collocations are distributed across 24 TCCON sites; the largest site contributes 11.4 % and the three largest sites 31.9 %, indicating that the pooled statistics are not dominated by a small number of stations. The pooled bias is 0.21 ppm (95 % CI: −0.08 to 0.48 ppm) and the RMSE is 1.49 ppm (95 % CI: 1.25–1.74 ppm). The site-mean RMSE is 1.40 ppm and the interstation bias is 0.73 ppm. Across all products, the site-mean RMSE differs from the corresponding pooled RMSE by no more than 0.16 ppm. GOSAT shows the most concentrated station sampling (38.6 % from the largest site and 76.6 % from the three largest sites) and the widest bias interval (−0.72 to 0.60 ppm), consistent with its greater site-to-site variability.

Because the official empirical correction for GOSAT-2 was calibrated using TCCON GGG2020, we now use TCCON as a common ground-based validation reference when discussing all satellite products collectively; the DQ-1–TCCON comparison itself remains an independent observational validation of DQ-1.

Changes in the manuscript :

Sect. 2.2.6, p. 13, lines 364–366.

Revised text: "TCCON is used as a common ground-based reference for satellite XCO2 validation. Following the validation approach for DQ-1/ACDL described by Han et al. (2025), we construct a 1° × 1° spatial matching window centered on each TCCON site and extract quality-controlled TCCON XCO2 observations within ±0.5 h of the satellite overpass time."

Sect. 3.2, p. 16, lines 437–446.

Revised text: "Additional information on the site-level sampling and validation statistics is provided in Tables S1 and S2. The site-mean RMSE differs from the corresponding overall RMSE by no more than 0.16 ppm across the satellite products, suggesting that the overall RMSE is not strongly affected by the unequal numbers of collocations among TCCON sites. The interstation bias ranges from 0.56 to 1.40 ppm, with the largest value found for GOSAT, indicating greater site-to-site variability. GOSAT also shows a more concentrated distribution of collocations among a small number of sites. In contrast, the DQ-1 collocations are distributed across 24 TCCON sites, and no individual site accounts for more than 11.4 % of the total, indicating that the DQ-1 validation statistics are not driven by a single site. The 95 % confidence intervals further quantify the uncertainty of the pooled validation statistics. For DQ-1, the bias is 0.21 ppm (95 % CI: −0.08 to 0.48 ppm) and the RMSE is 1.49 ppm (95 % CI: 1.25–1.74 ppm). GOSAT shows the widest bias interval (−0.72 to 0.60 ppm), consistent with its greater site-to-site variability and more concentrated station sampling."

Table S1. Distribution of valid satellite–TCCON collocations across TCCON sites. Values are reported as N (%), with percentages relative to the total number of collocations for each satellite.

TCCON site	OCO-2 N (%)	OCO-3 N (%)	GOSAT N (%)	GOSAT-2 N (%)	DQ-1 N (%)
an	–	1 (0.2)	–	–	1 (0.6)
br	1 (0.3)	–	–	3 (0.3)	3 (1.8)
bu	3 (0.9)	4 (0.8)	4 (0.9)	9 (1.0)	2 (1.2)
ci	33 (9.9)	57 (11.5)	173 (38.6)	152 (17.3)	7 (4.2)
db	6 (1.8)	8 (1.6)	5 (1.1)	4 (0.5)	4 (2.4)
df	58 (17.4)	63 (12.7)	122 (27.2)	99 (11.2)	19 (11.4)
et	16 (4.8)	–	–	27 (3.1)	7 (4.2)
gm	5 (1.5)	22 (4.4)	4 (0.9)	10 (1.1)	6 (3.6)
hf	4 (1.2)	2 (0.4)	–	24 (2.7)	5 (3.0)
hw	6 (1.8)	32 (6.5)	–	20 (2.3)	3 (1.8)
iz	4 (1.2)	14 (2.8)	–	5 (0.6)	1 (0.6)
js	9 (2.7)	17 (3.4)	20 (4.5)	30 (3.4)	3 (1.8)
ka	15 (4.5)	12 (2.4)	–	21 (2.4)	3 (1.8)
lr	21 (6.3)	47 (9.5)	1 (0.2)	77 (8.8)	10 (6.0)
ni	24 (7.2)	34 (6.9)	–	52 (5.9)	8 (4.8)
ny	4 (1.2)	–	–	–	10 (6.0)
oc	25 (7.5)	41 (8.3)	22 (4.9)	112 (12.7)	15 (9.0)
or	4 (1.2)	14 (2.8)	5 (1.1)	13 (1.5)	1 (0.6)
pa	24 (7.2)	39 (7.9)	16 (3.6)	39 (4.4)	16 (9.6)
pr	8 (2.4)	16 (3.2)	2 (0.4)	16 (1.8)	5 (3.0)
rj	9 (2.7)	7 (1.4)	–	17 (1.9)	7 (4.2)
so	10 (3.0)	–	9 (2.0)	13 (1.5)	6 (3.6)
wg	17 (5.1)	32 (6.5)	17 (3.8)	34 (3.9)	6 (3.6)
xh	27 (8.1)	33 (6.7)	48 (10.7)	103 (11.7)	18 (10.8)

TCCON site	OCO-2 N (%)	OCO-3 N (%)	GOSAT N (%)	GOSAT-2 N (%)	DQ-1 N (%)
Total	333	495	448	880	166

Table S2. Satellite–TCCON validation statistics, site-level variability, and 95 % confidence intervals. The 95 % confidence intervals (CIs) of the pooled bias and RMSE are defined by the 2.5th and 97.5th percentiles of 20 000 TCCON-site cluster-bootstrap realizations. Interstation bias is the standard deviation of the site-specific mean biases, and site-mean RMSE is the arithmetic mean of the site-specific RMSE values, with equal weight assigned to each contributing site.

Product	N	Bias (95 % CI) (ppm)	Interstation bias (ppm)	RMSE (95 % CI) (ppm)	Site-mean RMSE (ppm)
OCO-2	333	−0.18 [−0.35, 0.00]	0.60	1.09 [0.89, 1.31]	1.04
OCO-3	495	0.08 [−0.10, 0.28]	0.56	1.12 [0.98, 1.27]	1.12
GOSAT	448	−0.28 [−0.72, 0.60]	1.40	1.45 [1.21, 1.85]	1.60
GOSAT-2 original	880	3.56 [3.24, 3.88]	1.19	3.98 [3.75, 4.27]	4.03
GOSAT-2 corrected	880	0.01 [−0.20, 0.26]	0.97	1.54 [1.33, 1.81]	1.70
DQ-1	166	0.21 [−0.08, 0.48]	0.73	1.49 [1.25, 1.74]	1.40

Comment 7

Please clarify whether the daily grid-cell XCO_2 values are simple arithmetic means or uncertainty-weighted means.

ANSWER: The daily grid-cell XCO_2 values and mean observation times are simple arithmetic means of all quality-controlled observations within the corresponding selected grid cell and day. They are not uncertainty-weighted means. We have clarified this in Sect. 2.2.2.

Changes in the manuscript:

Sect. 2.2.2, p. 11, lines 289–294.

Revised text: "For each date and selected grid cell, the daily XCO_2 and mean observation time are calculated using simple arithmetic averaging over all valid observations within the grid cell. A sample pair is considered a candidate pair only when the two products fall within the same selected 2° comparison grid cell on the same date and both satisfy the valid-observation criterion defined in Sect. 2.2.1. Candidate pairs are retained as final paired samples only if the difference between their daily mean observation times does not exceed 4 h. For the pairwise statistics shown in Fig. 4 and Table S3, bias is defined as XCO_2(y-axis product) − XCO_2(x-axis product)."

Comment 8

The reduced OCO-3 coverage during the stowage/reconfiguration period should be separated from limitations caused by clouds, aerosols, or solar illumination.

ANSWER: We have separated the OCO-3 mission interruption from normal retrieval-availability effects. Appendix A1 now identifies reduced sampling in October and early November 2023 followed by a prolonged gap until mid-July 2024, associated with temporary storage on the ISS. We explicitly state that this was a mission-level interruption rather than an ocean-specific instrument malfunction. During normal operations, ocean sampling is additionally constrained by OCO-3 viewing/glint-pointing geometry and ISS operational requirements, while clouds, aerosols, illumination, and quality screening further affect retrieval availability.

To test whether the interruption materially changes the tropical-Pacific comparison, we repeated the analysis using only 21 common-operation full months. OCO-3 still provides only about 51 % of the OCO-2 sampling (8,734 versus 17,127 valid observations), and the DQ-1–OCO-2 bias/RMSE changes only from 0.58/1.27 to 0.61/1.26 ppm, while DQ-1–OCO-3 changes from 0.30/1.26 to 0.30/1.28 ppm. Thus, the interruption reduces sampling availability but has little effect on the regional residual statistics.

Changes in the manuscript:

Appendix A1, pp. 32–33, lines 755–776.

Revised text: “The monthly statistics are shown in Fig. A2. The mean monthly numbers of valid observations for DQ-1 Asc and Desc are 0.77 × 104 and 0.71 × 104, respectively, and the corresponding value for DQ-1 Total is approximately 1.48 × 104. This value is higher than the corresponding OCO-2 value of 0.99 × 104. OCO-3 has a mean monthly number of valid observations of approximately 0.61 × 104 but exhibits substantial month-to-month variability. Reduced sampling is evident in October and early November 2023, followed by a prolonged data gap until mid-July 2024. The gray-shaded region in Fig. A2 marks the OCO-3 data gap associated with its temporary storage on the ISS. This was a mission-level interruption rather than an ocean-specific instrument malfunction. During normal science operations, OCO-3 ocean sampling is additionally constrained by viewing and glint-pointing geometry and ISS operational requirements, while clouds, aerosols, solar-illumination conditions, and quality screening further affect retrieval availability (Taylor et al., 2023). The prolonged data gap nevertheless contributed to the lower overall number of valid OCO-3 observations relative to OCO-2 over the study period. To assess whether the reduced tropical-Pacific sampling was primarily caused by this interruption, we repeated the comparison using only common-operation full months, excluding October 2023–July 2024. During the remaining 21 months, OCO-2 and OCO-3 provided 17,127 and 8,734 valid observations, respectively, corresponding to monthly means of 815.6 and 415.9. OCO-3 sampling therefore remained approximately 51 % of OCO-2 sampling outside the interruption period, indicating that the interruption contributed to, but did not fully explain, the lower tropical-Pacific sampling. We further assessed whether the OCO-3 interruption affected the regional DQ-1–OCO comparison over the tropical Pacific by recalculating the bias and RMSE for the same 21 common-operation months. For DQ-1–OCO-2, the bias and RMSE changed from 0.58 and 1.27 ppm for the full period to 0.61 and 1.26 ppm, respectively; for DQ-1–OCO-3, the corresponding values changed from 0.30 and 1.26 ppm to 0.30 and 1.28 ppm. These small changes indicate that the OCO-3 interruption had little influence on the tropical-Pacific bias and RMSE. GOSAT-2 and GOSAT have much lower mean monthly numbers of valid observations, approximately 0.14 × 104 and 0.03 × 104, respectively. Overall, these results highlight the complementarity of multi-satellite observations in spatial coverage and temporal continuity.”

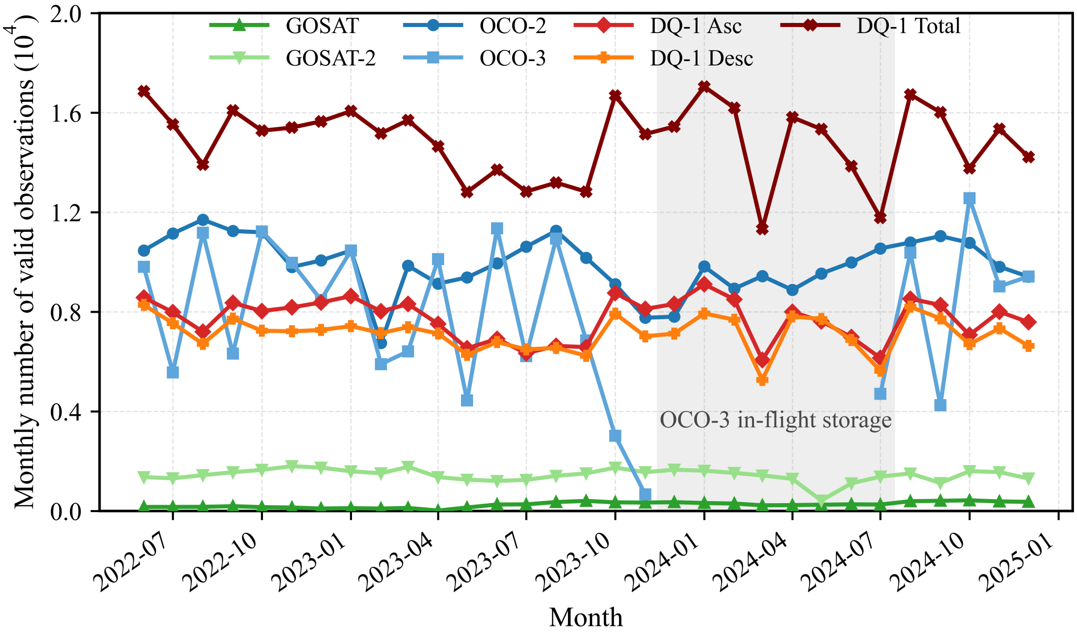

Figure A2. Monthly variations in the number of valid observations for the satellite XCO_2 products. The gray-shaded region indicates the OCO-3 data gap associated with its temporary storage on the ISS.

Comment 9

The title "First comparison" seems somewhat strong, since previous studies cited in the manuscript have already compared or validated DQ-1 against other XCO_2 datasets. The main novelty appears to be the unified multi-satellite comparison framework.

ANSWER: We agree and have replaced "First comparison" with a title that identifies the actual contribution: "Assessing the consistency of XCO_2 products from DQ-1 ACDL and passive satellites using a unified comparison framework". The final paragraphs of the Introduction and the Conclusions now distinguish this joint treatment of sampling and effective column definitions from earlier individual-product validation and passive-satellite intercomparisons.

Changes in the manuscript :

Introduction, p. 3, lines 87–91.

Revised text: "To our knowledge, no previous study has systematically compared DQ-1 ACDL with OCO-2, OCO-3, GOSAT, and GOSAT-2 within a unified active–passive framework that jointly accounts for background-field coherence, multi-satellite sampling availability, spatiotemporal collocation, and XCO_2 column-definition differences. This unified framework provides a consistent basis for interpreting inter-product differences between active and passive satellite XCO_2 observations."

Comment 10

Please check the manuscript for typographical errors.

ANSWER: We have checked the manuscript and corrected typographical and formatting errors, including duplicated wording, capitalization, and the author-contribution heading.