Reply to Referee #1

Dear Artem Feofilov:

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

The reviewed manuscript discusses the XCO_2 product retrieved from the DQ-1 ACDL space-borne lidar and compares it both with other space-borne XCO_2 measurements and with the reference ground-based stations. The need for accurate CO_2 measurements is obvious, its mean value characterizes global changes, whereas its sources and sinks tell us about carbon cycle, emissions, and transport effects.

Compared to existing and past space-borne passive measurements (SCIAMACHY, GOSATs, OCOs, TanSat), the space-borne lidar has certain advantages. It uses a controlled source of probing light whereas the passive sounders depend on the incident solar radiance that is not pointed the same way throughout the globe. In particular, this changes the observability of the nighttime part of the globe that is completely inaccessible for the passive XCO_2 sounders relying on reflected sunlight (OCO-2, GOSAT, TanSat).

The thermal-infrared emission sounders like AIRS, IASI, or CrIS do observe CO_2 emissions, but their channels are characterized by altitude-dependent contribution

functions related to CO_2 concentrations, whereas the differential absorption method probes the whole atmospheric column down to the ground. For this method, the absence of solar photons contaminating the signals makes the nighttime observations preferable to daytime ones.

From this point of view, ACDL might serve either as a reference for other sounders or as a strong supplement to existing XCO_2 data.

The manuscript introduces the problem, discusses an extensive list of comparison datasets that includes not only spaceborne passive remote sensors like OCOs or GOSATs, but also a network of ground-based Fourier-spectrometers TCCON that retrieve XCO_2 from solar near-infrared absorption spectra and serve as a reference for space-borne observations.

Throughout the manuscript, the authors show different cross-comparisons between the space-borne passive, space-borne active, and ground-based observations, and conclude that the overall agreement between the compared instruments is astoundingly good, with bias less than 0.3 ppmv and RMSE less than 1.5ppmv for all instruments except GOSAT-2. The correlation coefficients between the datasets are on the order of 0.91-0.95.

The manuscript is topical, it fits well the scope of the journal, is well organized and well-written.

I believe it can be published in the journal with minor revisions, which I propose for the discussion below.

ANSWER: We thank the reviewer for the positive assessment and constructive suggestions. We have revised the manuscript and Supplement extensively, with particular attention to the accessibility of the Abstract, the IPDA optical path and day–night behavior, the hard-target limitation, the role and ordering of TCCON validation, sampling diagnostics, and the ocean wind-speed analysis.

Specific comments

Comment 1

The abstract relies heavily on acronyms (XCO_2, IPDA, TCCON, and the various instrument names) without expanding any of them at first use. This is fine for a specialist audience already familiar with the history of spaceborne XCO_2 observations, but makes the abstract nearly unreadable for a general reader outside this subfield.

Since the abstract is often the only part of the paper read by non-specialists, could the authors, please, either expand all acronyms at first use or rewrite the abstract in plain language?

ANSWER: We have rewritten the Abstract so that the measurement concept and motivation are clear to readers outside the immediate subfield. All major acronyms are expanded at first use, including DQ-1, ACDL, IPDA, XCO2, OCO, GOSAT, CAMS, TCCON, and RMSE. The opening now states that ACDL uses an active laser, can observe at night and high latitudes, and nevertheless differs from passive products in sampling and effective column definition; this motivates the unified comparison before the quantitative results are presented.

Changes in the manuscript:

Abstract, p. 1, lines 14–28.

Revised text: "Abstract. Da Qi-1 (DQ-1) carries the Aerosol and Carbon Dioxide Lidar (ACDL), a spaceborne integrated path differential absorption (IPDA) lidar that retrieves the column-averaged dry-air mole fraction of carbon dioxide (XCO_2) using active laser measurements. Unlike passive near-infrared/shortwave-infrared sensors that rely on reflected sunlight, ACDL can observe at night and high latitudes, but its sampling and effective column definition differ from passive XCO_2 products. Evaluation is therefore needed before active and passive observations can be interpreted jointly. We compare DQ-1 ACDL with the Orbiting Carbon Observatory (OCO-2 and OCO-3) and Greenhouse gases Observing SATellite (GOSAT and GOSAT-2) missions from June 2022 to December 2024 using a unified framework that combines Copernicus Atmosphere Monitoring Service (CAMS)-based coherence assessment, sampling availability, daily 2° aggregation, spatiotemporal collocation, and correction for XCO_2 column-definition differences. Validation against the Total Carbon Column Observing

Network (TCCON) gives R = 0.92, a mean bias of 0.21 ppm, and a root-mean-square error (RMSE) of 1.49 ppm for daytime DQ-1, comparable to GOSAT. Over coherent and well-sampled regions, DQ-1 is broadly consistent with OCO-2 and OCO-3, with mean differences generally below 0.5 ppm and similar spatial and meridional structures. Over ocean, the DQ-1–OCO differences vary with surface reflectance and wind speed, with different reflectance relationships for OCO-2 and OCO-3. Nighttime DQ-1 is also broadly consistent with CAMS-derived XCO_2. Regional analyses show complementary DQ-1 coverage where passive observations are limited. These results support DQ-1 ACDL as a quantitatively comparable and complementary component of the satellite XCO_2 observing system."

Comment 2

Lines 55-72 and/or lines 102-113: it makes sense to add some information about the optical path of ACDL to explain why the vertically resolved lidar signal does not carry the information about the CO_2 and only the column value is retrievable. The references given in this section explain this, but adding a couple of sentences regarding the method will simplify the flow.

ANSWER: We have added an explanation of the two-way optical path. The online and offline pulses travel through the atmosphere, are reflected by a hard target, and return to the receiver. Their differential attenuation measures absorption integrated along that path. Range resolution identifies the reflecting target; it does not supply independent CO_2 absorption measurements for each atmospheric layer in this IPDA product. Accordingly, ACDL retrieves a weighted column, whose vertical sensitivity is described by its weighting function. This distinction is now stated explicitly and linked to the column-definition treatment in Sect. 2.2.3.

Changes in the manuscript:

Introduction, pp. 2–3, lines 65–71.

Revised text: "For each CO_2-IPDA measurement, the online and offline laser pulses propagate through the atmosphere and are reflected by a hard target, typically the surface or an optically thick cloud top, before the return signals are collected by the telescope. Their differential attenuation therefore provides an integrated differential absorption optical depth along the two-way path. Although the lidar return is range resolved for locating the reflecting target, the CO_2 absorption measurement is path

integrated and does not provide independent layer-by-layer CO_2 information; instead, the ACDL weighting function describes the vertical sensitivity of the integrated column measurement, from which XCO_2 is retrieved (Shi et al., 2023; Han et al., 2025)."

Sect. 2.1.2, pp. 5–6, lines 145–147.

Revised text: "Although all products report XCO_2, 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."

Comment 3

Neither of these paragraphs discusses the differences of daytime and nighttime observations. Since the online and offline channels share nearly the same additive solar background but different signal levels, an imperfectly subtracted background biases the online/offline ratio toward unity and, therefore, biases XCO_2 low – an effect that should vanish at night.

Could the authors show a daytime-vs-nighttime comparison over similar scenes to bound this residual bias, if not already addressed in the cited references? Maybe it makes sense to provide a theoretical estimate for exactly the same atmospheric column to exclude any effects caused by diurnal change.

ANSWER: We agree that imperfect solar-background subtraction is a potential mechanism that may affect daytime ACDL retrievals. We therefore added a day–night diagnostic using the ascending (daytime) and descending (nighttime) DQ-1 observations. Because exact daytime/nighttime repetition of the same atmospheric column is sparse, we compare the zonal-mean Asc-minus-Desc signal with the corresponding CAMS-derived signal after applying the DQ-1 observation operator. This avoids treating unmatched columns as identical scenes.

Over 60° S–60° N, the DQ-1 minus CAMS-derived day–night signal difference is +0.09 ppm in June 2022 and −0.05 ppm in December 2022, with zonal diagnostic RMSE

values of 0.17 and 0.18 ppm, respectively (Fig. S8). The small magnitudes and opposite signs do not indicate a persistent systematic daytime-low bias at the zonal-mean scale in these two months.

We emphasize that this is a diagnostic constraint rather than a same-column theoretical estimate or a direct calibration of the solar-background subtraction. Zonal averaging and the model reference cannot fully separate residual background effects from sampling differences and unresolved diurnal variability, so we do not claim that the analysis excludes such effects for individual retrievals.

Changes in the manuscript:

Sect. 3.5, p. 25, lines 617–623.

Revised text: "To examine whether a residual daytime retrieval effect is evident at the zonal scale, we compared the ascending-minus-descending (daytime-minus-nighttime) XCO2 differences in DQ-1 with the corresponding CAMS-derived differences for June and December 2022 (Fig. S8). Over 60° S–60° N, the differences between the DQ-1 and CAMS-derived day–night signals are +0.09 ppm in June and −0.05 ppm in December, with RMSE values of 0.17 and 0.18 ppm, respectively. The small magnitudes and opposite signs do not indicate a persistent systematic daytime-low bias at the zonal-mean scale. This diagnostic does not fully isolate residual solar-background effects at the individual-column level, but it provides no evidence of a large systematic day–night offset in these two months."

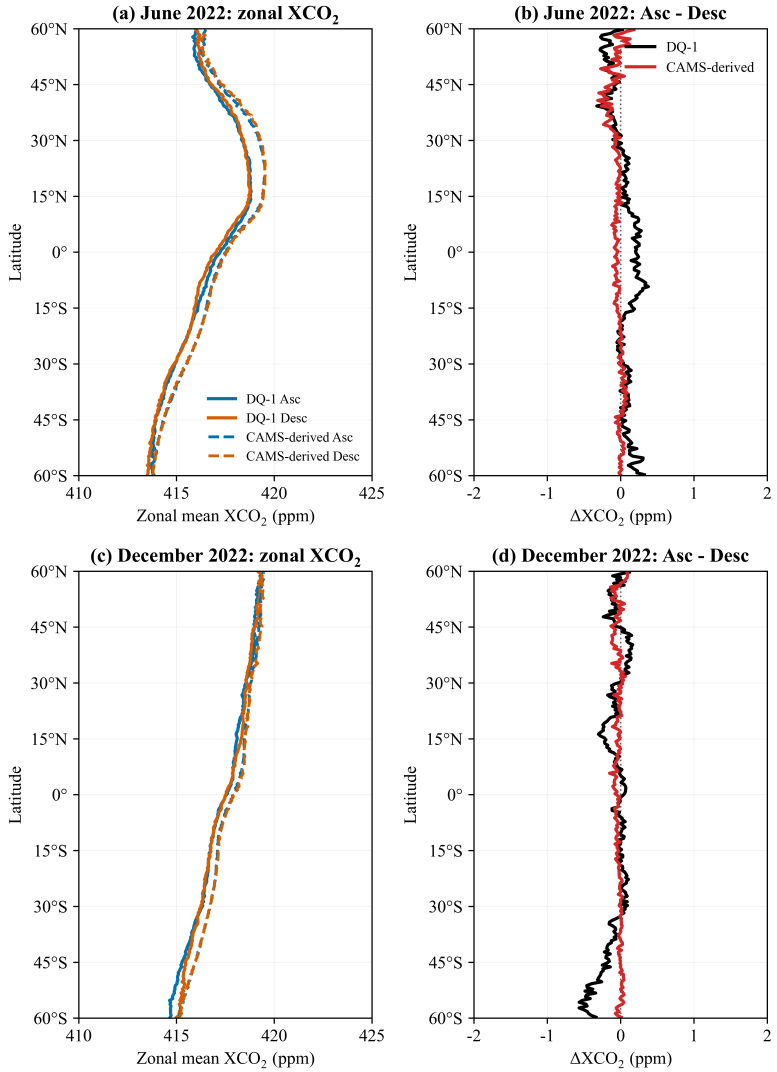


Figure S8. Zonal-mean comparison of ascending (daytime) and descending (nighttime) DQ-1 XCO2 and the corresponding CAMS-derived XCO2 under the DQ-1 column definition for (a, b) June 2022 and (c, d) December 2022. Panels (a) and (c) show the zonal-mean XCO2 distributions for the Asc and Desc observations, and panels (b) and (d) show the corresponding Asc-minus-Desc (daytime-minus-nighttime) differences for DQ-1 and CAMS-derived XCO2.

Comment 4

A related point: the differential absorption method requires a valid hard target (surface or opaque cloud top) to close the two-way path. Could the authors please note the fraction of shots lost to a missing hard target and whether this loss is geographically or seasonally structured?

ANSWER: The released ACDL L2A product has no dedicated flag that uniquely identifies a missing hard target. We therefore examined QAflag = 3, which denotes abnormally low echo intensity, as an operational proxy for insufficient return signal. Over June 2022–December 2024, 222,287,781 of 1,424,855,515 examined L2A records have QAflag = 3, corresponding to 15.60 %. The monthly fraction ranges from 13.80 % to 17.68 %, without a clear recurring seasonal cycle.

These figures describe low-echo records and should not be interpreted as a measured fraction of shots lost specifically because a hard target was absent. Moreover, these records do not contain valid footprint coordinates in the released product, so their geographical distribution cannot be quantified reliably from L2A. We have added the available fraction, monthly variability, and both limitations to Sect. 2.1.1.

Changes in the manuscript:

Sect. 2.1.1, pp. 4–5, lines 113–118.

> **Revised text:** "Because the released ACDL L2A product does not provide a dedicated flag for missing hard targets, QAflag = 3, which denotes abnormal retrievals associated with low echo intensity, is used here only as an operational proxy for insufficient return signal rather than as a direct identification of missing hard targets. During June 2022–December 2024, QAflag = 3 accounted for 15.60 % of the examined L2A records, with monthly fractions ranging from 13.80 % to 17.68 % and no clear recurring seasonal cycle. These records lack valid footprint geolocation in the released L2A product, preventing a robust assessment of their geographical distribution."

Comment 5

The validation with TCCON observation appears only after the diagnosis of residuals in passive spaceborne instrument comparison, whereas I'd say that TCCON observations should have more weight and therefore precede them. See the other comment regarding the order of presentation and TCCON measurements below.

ANSWER: We agree that the ground-based TCCON comparison should precede the product-to-product residual diagnostics. We therefore moved the TCCON evaluation from the original Sect. 3.7 to revised Sect. 3.2, immediately after the definition of the

comparison regions and before the inter-satellite analysis in Sect. 3.3 and the regional residual diagnostics in Sect. 3.4. Figure 3 and Tables S1–S2 now provide the collocation counts, site distribution, uncertainty intervals, and validation statistics at this earlier stage.

Changes in the manuscript:

Sect. 3.2, pp. 15–16, lines 415–449.

Revised text: "We first compare the five satellite XCO_2 products with TCCON observations to assess their overall agreement with a common ground-based reference. Figure 3 shows that OCO-2 and OCO-3 have the strongest agreement with TCCON, with correlation coefficients of R = 0.95 and 0.94, mean biases of −0.18 and 0.08 ppm, and RMSE values of 1.09 and 1.12 ppm, respectively. These results indicate small systematic biases and random scatter for the OCO products, consistent with the validation results for OCO-2 and OCO-3 reported by Taylor et al. (2023). GOSAT also agrees well with TCCON, with R = 0.91, a mean bias of −0.28 ppm, and an RMSE of 1.45 ppm. Overall, OCO-2, OCO-3, and GOSAT show close agreement with TCCON, with absolute mean biases below 0.3 ppm and RMSE values below 1.5 ppm.

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.

The comparison between DQ-1 and TCCON shows that, although the number of valid collocations is smaller than that for the passive satellites, DQ-1 still has R = 0.92, a mean bias of 0.21 ppm, and an RMSE of 1.49 ppm. These statistics are comparable to

those of GOSAT and empirically corrected GOSAT-2, although the OCO-2 and OCO-3 comparisons show slightly lower RMSE values, indicating that daytime DQ-1 ACDL XCO_2 is consistent with TCCON observations. This result is consistent with the initial validation of ACDL XCO_2 by Han et al. (2025), who reported a near-zero mean bias and an RMSE of about 1.4 ppm. The relatively limited number of DQ-1–TCCON collocations is mainly constrained by the point-sampling geometry of the active lidar observations. Unlike passive sensors with wider swaths or denser sampling, DQ-1/ACDL samples only along its ground track, which reduces the probability of close spatiotemporal coincidence with TCCON sites.

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.

Overall, the TCCON comparisons show that OCO-2, OCO-3, GOSAT, and empirically corrected GOSAT-2 are in good agreement with the ground-based reference. Daytime DQ-1 ACDL XCO_2 also agrees well with TCCON, providing independent validation of the DQ-1 product and supporting its role as a complement to passive satellite XCO_2 observations.”

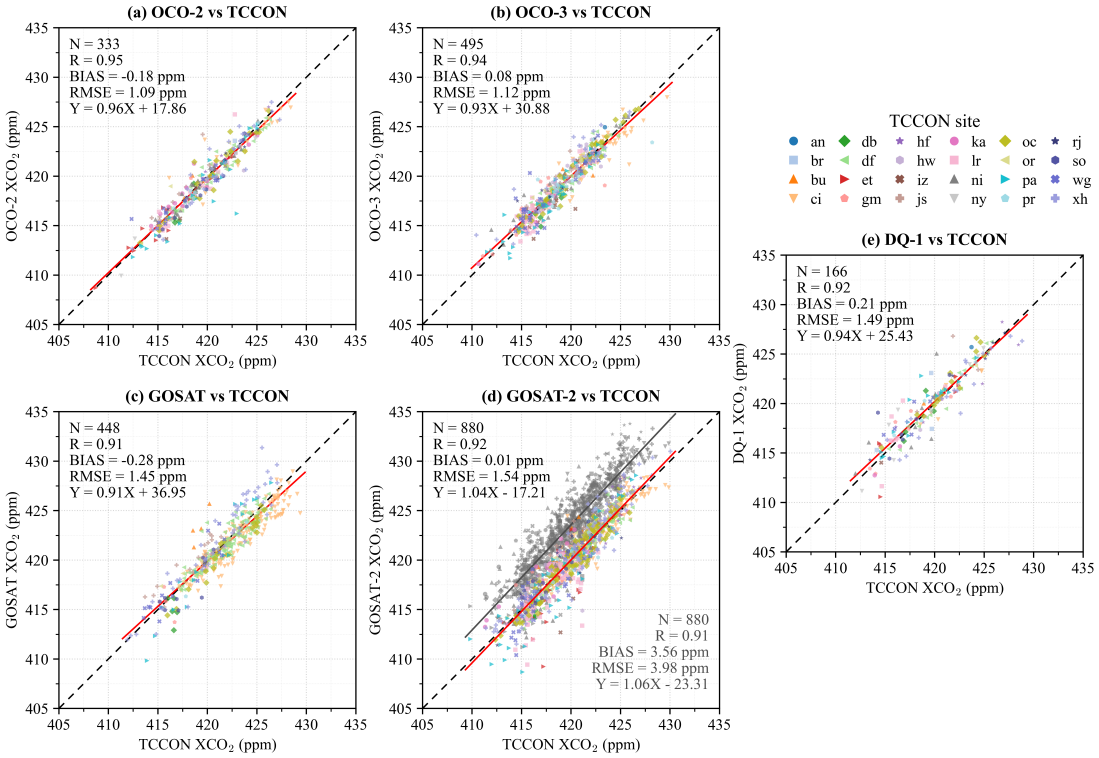


Figure 3. Comparison of satellite XCO_2 products with TCCON observations. (a–e) OCO-2, OCO-3, GOSAT, GOSAT-2, and DQ-1 XCO_2 compared with collocated TCCON XCO_2, respectively. Different colors and symbols denote individual TCCON sites. In panel (d), colored and gray symbols denote empirically bias-corrected 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 6

Section 3.1 is dedicated to sampling and takes a lot of space – I am not sure that it belongs to the Results. In principle, it could have been reduced or moved to Appendix without losing the main message of the section.

ANSWER: We have moved the detailed spatial sampling coverage and monthly availability analysis from the original Results section to Appendix A1, with Figs. A1 and A2. The main Results retain only the comparison-region selection needed to define the domain of the subsequent evaluation.

Changes in the manuscript:

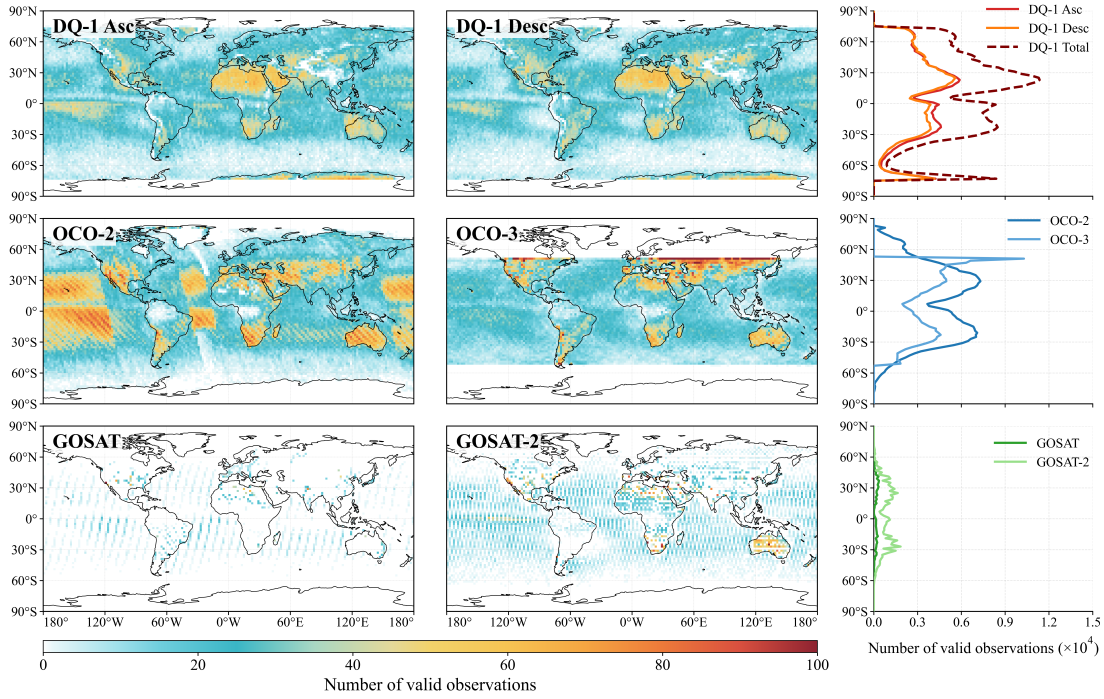


Figure A1. Spatial distributions and zonal totals of cumulative number of valid observations for the satellite XCO_2 products at 2° × 2° resolution.

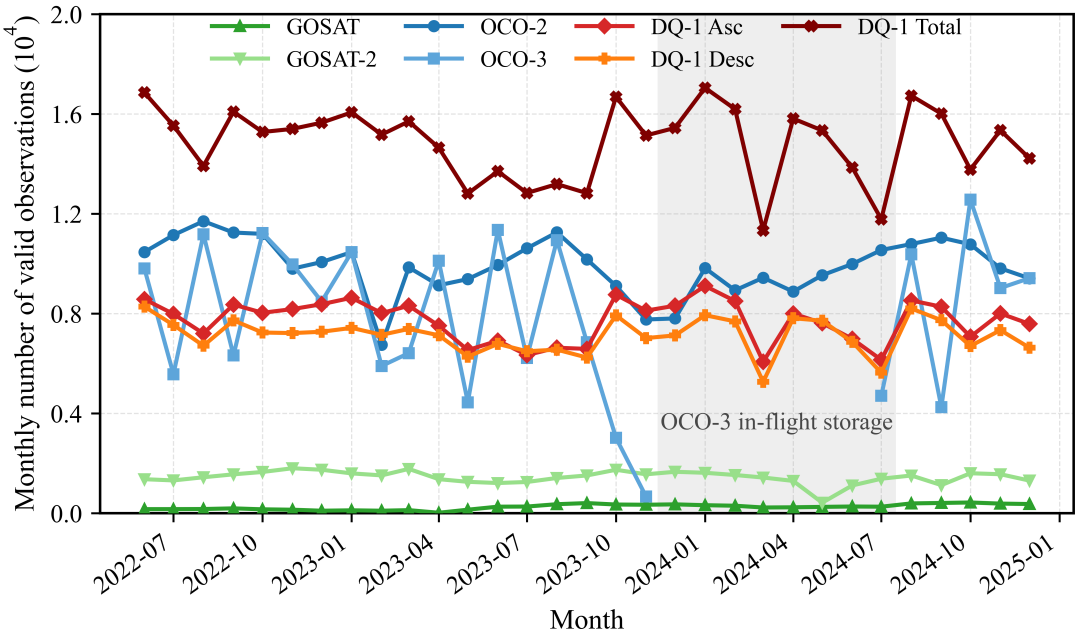


Figure A2. Monthly variations in the number of valid observations for the satellite XCO_2 products. The gray-shaded region indicates the OCO-3 data gap associated with its temporary storage on the ISS.

Comment 7

Throughout the Results section, some of the differences between the datasets are explained whereas the other ones are just reported. Maybe it makes sense to give the priorities or weights to these differences: for example, one chooses TCCON as a reference as in Section 3.7 and discusses the differences only with respect to these observations. Again, this will require a slight reorganization of the order of sections.

If the authors believe that TCCON comparisons are of lesser importance, I'd like to understand, why – for me, they are a good candidate for the reference ones because they depend on smaller number of unknowns.

ANSWER: We thank the reviewer for this important comment. We agree that a clear reference hierarchy is essential for interpreting the consistency among different XCO_2 products. In the revised manuscript, we have clarified the distinct roles of TCCON and CAMS in the evaluation framework and organized the analyses into a three-level evaluation hierarchy.

First, TCCON observations serve as the absolute observational reference for evaluating satellite XCO_2 products against independent ground-based measurements. Second, satellite-to-satellite comparisons are used to assess the relative consistency among different XCO_2 products after accounting for sampling differences and differences in effective column definitions. Third, CAMS-based analyses are used as model-based diagnostic analyses of remaining inter-product differences rather than as independent observational validation.

Following this framework, the revised manuscript clearly distinguishes observational validation, satellite consistency assessment, and model-based diagnostic analyses. Specifically, Sect. 3.2 presents the validation of satellite XCO_2 products against TCCON observations, Sect. 3.3 evaluates satellite-to-satellite consistency over coherent and well-sampled comparison regions, and Sect. 3.4 provides CAMS-based and regional diagnostic analyses of remaining differences.

For the Australia-related differences identified in the satellite comparisons, we have added an independent regional TCCON diagnostic using Darwin and Wollongong observations in Sect. 3.4.2. Because the number of available DQ-1–TCCON collocations is limited in this region, this analysis is presented as a regional diagnostic to examine spatially dependent differences rather than as a primary validation result for DQ-1.

Changes in the manuscript :

Sect. 2.1.3, p. 6, lines 161–163.

Revised text: “Throughout this study, TCCON is used as the ground-based observational reference for daytime validation, whereas CAMS is used as a model-based common reference for comparison-region selection, column-definition harmonization, and diagnostic analyses. CAMS is not treated as observational truth. The specific uses of CAMS are described in Sects. 2.2.1, 2.2.3, and 2.2.7.”

Sect. 3.4.2, pp. 21–22, lines 530-544.

Revised text: “To further diagnose these regional differences, DQ-1 and OCO-2/OCO-3 were independently compared with CAMS-derived XCO2 using exactly the same collocated samples (Figs. S3–S4 and Table S5). Over Australia, DQ-1 shows small positive biases of +0.22 and +0.23 ppm relative to CAMS-derived XCO2 for the OCO-2 and OCO-3 collocations, respectively, whereas OCO-2 and OCO-3 show negative biases of −0.53 and −0.35 ppm. Thus, the positive DQ-1–OCO differences over Australia are accompanied by opposite-signed departures of DQ-1 and the OCO products from the common CAMS-derived reference. A site-based comparison with TCCON at Darwin and Wollongong provides an independent regional check and likewise shows a positive DQ-1 bias, although only 10 DQ-1 collocations are available (Fig. S5). Over the tropical Pacific, the DQ-1–CAMS-derived XCO2 RMSE values are 1.09 and 1.08 ppm, compared with 0.41 and 0.42 ppm for OCO-2 and OCO-3, respectively. In contrast, over the Sahara, both DQ-1 and the OCO products show RMSE values below 1 ppm relative to CAMS-derived XCO2, and the contrast between the two products is much smaller than that over the tropical Pacific (Table S5). Over the tropical Pacific, the substantially larger DQ-1 RMSE relative to the common CAMS-derived reference indicates that the enhanced scatter is more evident on the DQ-1 side under these collocated conditions. This differs from Australia, where DQ-1 and the OCO products show opposite-signed mean departures from CAMS. CAMS is used here only as a common diagnostic reference rather than as observational truth; these comparisons therefore help identify relative product behavior but do not uniquely attribute retrieval errors.”

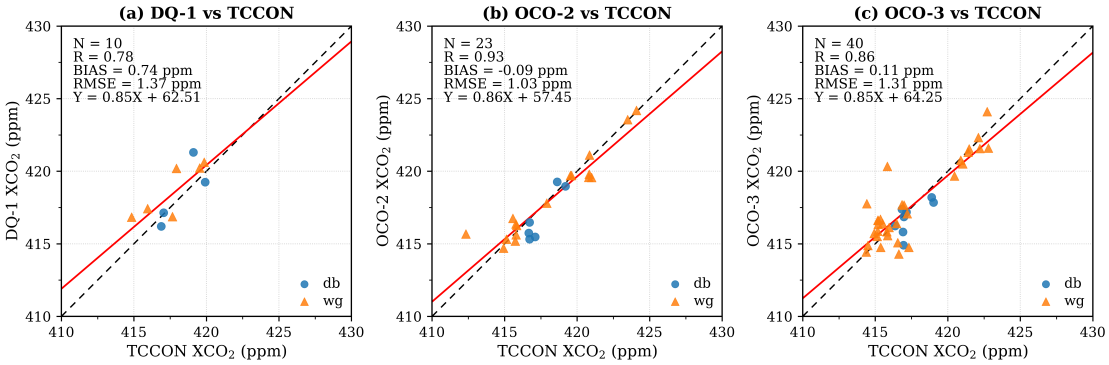


Figure S5. Scatterplot comparisons of DQ-1, OCO-2, and OCO-3 XCO_2 with TCCON observations at the Darwin (db) and Wollongong (wg) sites in Australia. Panels (a–c) show the comparisons for DQ-1, OCO-2, and OCO-3, respectively. The dashed black line denotes the 1:1 line, and the red line shows the linear regression fit.

Comment 8

For the figures showing geographical distribution of XCO_2 differences, did the authors try to subset the datasets over the oceans using the surface winds values? Does the smoother sea produce better results?

ANSWER: Yes. We added an ERA5 hourly 10 m wind-speed diagnostic for the oceanic DQ-1–OCO-2 and DQ-1–OCO-3 samples. We evaluate both the individual 2° grid-cell statistics and 1 m s−1 wind-speed-bin means, because the two levels answer different questions: the grid-cell analysis measures spatial scatter, whereas the bin means diagnose systematic changes in the conditional mean residuals.

At the individual-grid-cell level, the correlations with wind speed are weak: R = −0.16 and −0.29 for bias and R = −0.03 and −0.08 for RMSE for DQ-1–OCO-2 and DQ-1–OCO-3, respectively. After 1 m s−1 binning, however, the mean bias and RMSE both decrease with increasing wind speed, with R = −0.97 and −0.81 for DQ-1–OCO-2 and R = −0.94 and −0.89 for DQ-1–OCO-3. Thus, the binned statistics do not support the simple expectation that a lower wind speed (a smoother-sea proxy) necessarily gives better agreement. The large within-bin spread nevertheless shows that wind speed alone explains only a limited fraction of grid-cell-to-grid-cell variability.

The tropical Pacific also has slightly higher mean wind speeds than the selected ocean samples as a whole (6.48/6.73 versus 5.93/5.96 m s−1), while its mean grid-cell RMSE is slightly higher rather than lower. The elevated tropical-Pacific RMSE therefore cannot be explained simply by stronger near-surface winds. We further emphasize that ERA5 10 m wind speed is only an indirect proxy for sea-surface roughness and does not represent the instantaneous sun–sensor glint geometry of an individual OCO sounding; the analysis is therefore diagnostic rather than causal.

Changes in the manuscript:

Sect. 2.2.5, p. 13, lines 356–362.

Revised text: "To further examine the potential influence of ocean surface conditions on the DQ-1–OCO-2/OCO-3 differences, ERA5 10 m zonal and meridional wind components (u10 and v10) are collocated in space and time with the oceanic paired samples and used to derive 10 m wind speed. For each selected 2° ocean grid cell, the mean wind speed, mean bias, and RMSE are calculated, and the resulting grid-cell statistics are further grouped into wind-speed bins. The mean bias and RMSE within each bin are used to assess whether the active–passive XCO2 differences vary systematically with near-surface wind conditions. Wind-speed bins have a width of 1 m s−1. Linear least-squares fits and Pearson correlation coefficients are calculated separately for the individual 2° grid-cell statistics and for the wind-speed-bin means shown in Fig. S6."

Sect. 3.4.3, pp. 22–23, lines 572–584.

Revised text: "The ERA5 wind-speed analysis reveals different behavior at the individual-grid-cell and binned levels (Fig. S6). At the individual 2° grid-cell scale, the correlations with wind speed are weak, with R = −0.16 and −0.29 for bias and R = −0.03 and −0.08 for RMSE for DQ-1–OCO-2 and DQ-1–OCO-3, respectively. After grouping the grid cells into 1 m s−1 wind-speed bins, however, both the bin-mean bias and RMSE decrease with increasing wind speed, with R = −0.97 and −0.81 for DQ-1–OCO-2 and R = −0.94 and −0.89 for DQ-1–OCO-3, respectively. The strong binned relationships, together with the large spread of individual grid cells within the bins, indicate a systematic change in the mean residual statistics with wind speed, while substantial grid-cell variability remains. The tropical Pacific has slightly higher mean wind speeds than the selected ocean samples as a whole, but its mean grid-cell RMSE is also slightly higher (Table S4). Thus, the elevated tropical-Pacific RMSE cannot be explained simply by stronger near-surface winds. Moreover, ERA5 10 m wind speed is only an indirect indicator of sea-surface roughness and does not describe the

instantaneous sun–sensor glint geometry of individual OCO observations. The present diagnostics therefore indicate that ocean-surface conditions are associated with the residual statistics, but they do not identify a single reflectance- or wind-related mechanism that explains the tropical-Pacific differences."

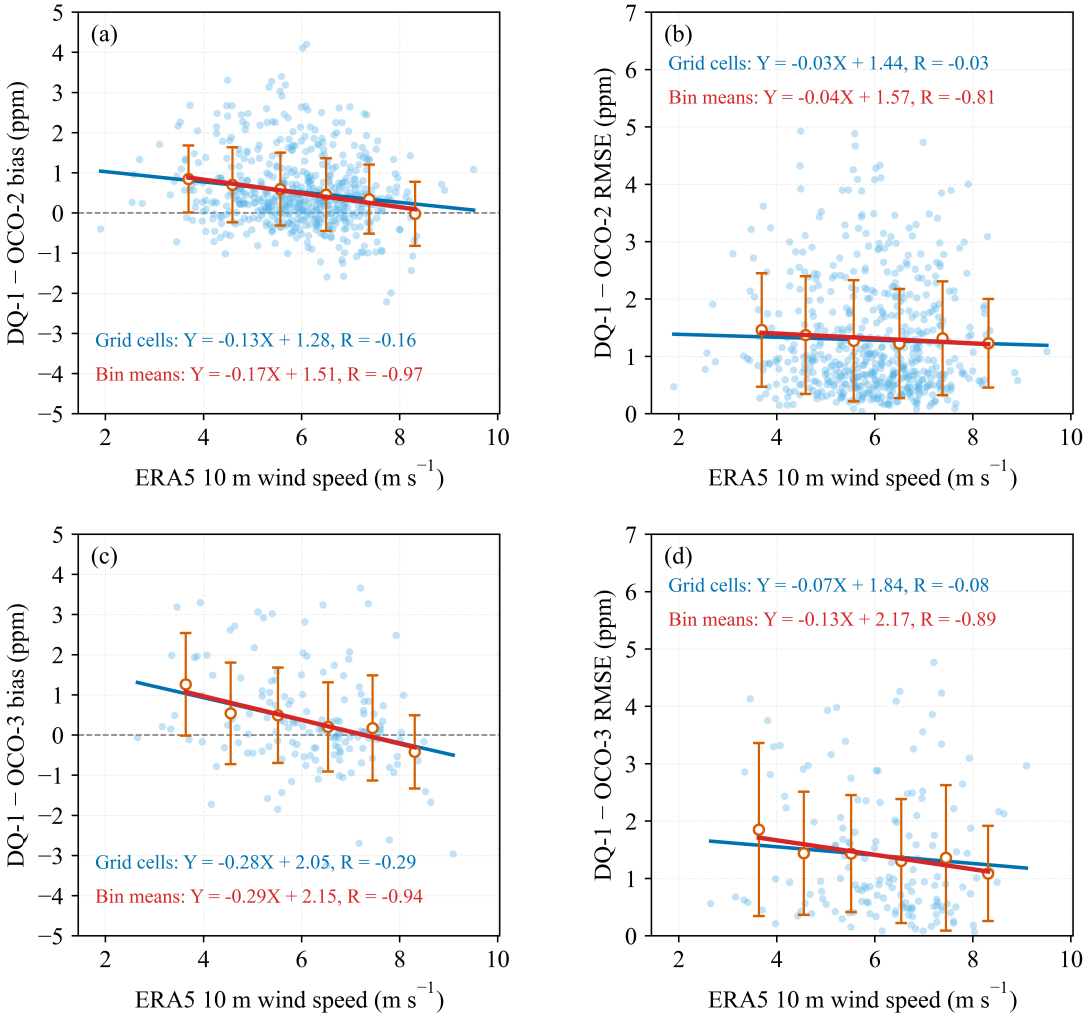


Figure S6. Dependence of oceanic DQ-1–OCO-2 and DQ-1–OCO-3 bias and RMSE on ERA5 10 m wind speed. Panels (a, b) show DQ-1–OCO-2 and panels (c, d) show DQ-1–OCO-3. Light blue dots denote individual oceanic 2° grid cells, while orange circles and error bars show 1 m s−1 wind-speed-bin means ±1 standard deviation. Blue and red lines show linear least-squares fits to the individual grid-

cell values and displayed bin means, respectively; the corresponding regression equations and Pearson correlation coefficients are annotated in each panel. Bias is defined as DQ-1 minus OCO.

Table S4. Regional and surface-type grid-cell statistics, MODIS Band 6 reflectance, and ERA5 10 m wind speed for the DQ-1–OCO-2 and DQ-1–OCO-3 comparisons.

Region	Pair	Mean grid-cell Band 6 reflectance	Mean 10 m wind speed (m s^{-1})	Mean grid-cell bias (ppm)	Median grid-cell bias (ppm)	Mean grid-cell RMSE (ppm)
Australia	DQ-1–OCO-2	0.3306	—	+0.76	+0.72	1.17
	DQ-1–OCO-3	0.3270	—	+0.57	+0.48	1.04
Tropical Pacific	DQ-1–OCO-2	0.0508	6.48	+0.58	+0.40	1.27
	DQ-1–OCO-3	0.0348	6.73	+0.30	+0.28	1.26
Sahara	DQ-1–OCO-2	0.5834	—	+0.51	+0.47	1.11
	DQ-1–OCO-3	0.5768	—	+0.22	+0.19	0.96
All selected land	DQ-1–OCO-2	0.3970	—	+0.46	+0.46	1.10
	DQ-1–OCO-3	0.4007	—	+0.15	+0.11	1.04
All selected ocean	DQ-1–OCO-2	0.0374	5.93	+0.51	+0.36	1.21
	DQ-1–OCO-3	0.0313	5.96	+0.40	+0.21	1.20

Note: Regional statistics were calculated by equally weighting the contributing 2° × 2° grid cells. MODIS Band 6 reflectance was averaged over grid cells with valid same-day MYD09CMG.061 retrievals. ERA5 10 m wind speed is reported only for the Tropical Pacific and all selected ocean samples. Bias is defined as DQ-1 minus OCO.

Comment 9

Fig. 1: the panels share the same color bar and OX/OY axes. In this case it would be logical to merge them and keep only 2xOX axis labels, 3xOY axis labels, and 1 color bar.

ANSWER: Revised. The original Fig. 1 is now Fig. A1. The six maps use a single shared color bar, longitude labels along the bottom row, and latitude labels along the left column. The added zonal-total panels provide a compact comparison of latitudinal sampling.

Reply to Referee #2

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 showed the validation results of XCO_2 from the DQ-1 satellite instrument, compared with other satellite XCO_2 and reanalysis data. All analyses look sincerely conducted, results are well organized, and writing quality seems fine. However, there are some questions and comments about suggested results as below. Final decision for this manuscript will be executed after checking authors' response and revised manuscript.

ANSWER: We thank the reviewer for the positive assessment and for the detailed questions. We have revised the manuscript and Supplement substantially, with particular attention to the Abstract motivation, satellite sampling differences, the OCO-3 operational interruption, the official GOSAT-2 bias correction, the regional Australia/tropical-Pacific diagnostics, ocean reflectance and wind-speed relationships, and the same-sample CAMS comparisons requested below.

Specific comments

Comment 1

Readers often read abstract first, sometimes only to see the main lesson of that paper. Thus the completeness of abstract is very important. In that context, the abstract of this manuscript should be improved more. What is DQ-1 ACDL? How does it relate to the IPDA? Why do we need to evaluate the DQ-1 ACDL data? Overall, the research motivation was not well transferred from this abstract.

ANSWER: We have rewritten the Abstract to explain what DQ-1 ACDL is, how it relates to IPDA, and why the active product needs to be evaluated against passive XCO2 products. The opening now identifies ACDL as a spaceborne IPDA lidar that retrieves XCO2 using active laser measurements, notes its nighttime/high-latitude sampling capability, and states that active and passive products still differ in sampling and effective column definition. This provides the motivation for the unified comparison framework before the main quantitative results are introduced.

Changes in the manuscript:

Abstract.

Revised text: "Abstract. Da Qi-1 (DQ-1) carries the Aerosol and Carbon Dioxide Lidar (ACDL), a spaceborne integrated path differential absorption (IPDA) lidar that retrieves the column-averaged dry-air mole fraction of carbon dioxide (XCO_2) using active laser measurements. Unlike passive near-infrared/shortwave-infrared sensors that rely on reflected sunlight, ACDL can observe at night and high latitudes, but its sampling and effective column definition differ from passive XCO_2 products. Evaluation is therefore needed before active and passive observations can be interpreted jointly. We compare DQ-1 ACDL with the Orbiting Carbon Observatory (OCO-2 and OCO-3) and Greenhouse gases Observing SATellite (GOSAT and GOSAT-2) missions from June 2022 to December 2024 using a unified framework that combines Copernicus Atmosphere Monitoring Service (CAMS)-based coherence assessment, sampling availability, daily 2° aggregation, spatiotemporal collocation, and correction for XCO2 column-definition differences. Validation against the Total Carbon Column Observing Network (TCCON) gives R = 0.92, a mean bias of 0.21

ppm, and a root-mean-square error (RMSE) of 1.49 ppm for daytime DQ-1, comparable to GOSAT. Over coherent and well-sampled regions, DQ-1 is broadly consistent with OCO-2 and OCO-3, with mean differences generally below 0.5 ppm and similar spatial and meridional structures. Over ocean, the DQ-1–OCO differences vary with surface reflectance and wind speed, with different reflectance relationships for OCO-2 and OCO-3. Nighttime DQ-1 is also broadly consistent with CAMS-derived XCO_2. Regional analyses show complementary DQ-1 coverage where passive observations are limited. These results support DQ-1 ACDL as a quantitatively comparable and complementary component of the satellite XCO_2 observing system."

Comment 2

In Figure 1, the number of available GOSAT-2 data number looks larger than that of GOSAT-1 data, which is OK because the GOSAT-2 is recent one. Then why the data number of OCO-3 is smaller than that of OCO-2? Additionally, the number of available OCO-3 data decreases over the tropical Pacific, but increases over the 40-50 °N in the North America and Asia. How to explain this? I did not see meaningful authors' statements about this issue.

This seems important in terms of sampling bias factor related to provided validation results.

ANSWER: We have expanded Appendix A1 to explain the spatial and temporal sampling differences. The original Fig. 1 is now Fig. A1, and Fig. A2 shows monthly availability. OCO-3 is carried by the ISS, whose precessing orbit restricts its coverage to approximately 52° S–52° N and concentrates sampling toward the orbital turning latitudes near 50°. This is consistent with the enhanced coverage over North America and Asia at 40–50° N. Its ocean-glint sampling is also more restricted by ISS pointing geometry and operations than that of OCO-2 (Taylor et al., 2023).

The lower total OCO-3 count additionally reflects the 2023–2024 operational interruption discussed in our response to Comment 5. Outside the interruption period, tropical-Pacific OCO-3 sampling remains approximately 51 % of OCO-2 sampling. Thus, the gap alone does not explain the spatial contrast. GOSAT-2 provides more valid

gridded observations than GOSAT in this analysis, but both remain much more sparsely sampled than the OCO products.

We now explicitly acknowledge the consequence for interpretation: same-date, same-grid, time-limited collocation reduces mismatch, but differing coverage still affects the number and representativeness of retained pairs. Here, a "valid observation" is one sensor–day–grid record meeting the sampling-length criterion, not one raw satellite sounding.

Changes in the manuscript:

Appendix A1, p. 31, lines 738–744.

Revised text: "In contrast, the precessing ISS orbit of OCO-3 restricts its coverage to approximately 52° S–52° N and produces a strongly nonuniform latitudinal sampling pattern. Sampling density increases toward the orbital turning latitudes near 50°, consistent with the enhanced OCO-3 sampling over parts of North America and Asia in the 40–50° N latitude band seen in Fig. A1 (Taylor et al., 2023). GOSAT and GOSAT-2 exhibit much sparser point-sampling patterns, with GOSAT-2 providing more valid gridded observations than GOSAT, although both remain substantially sparser than the OCO series."

Appendix A1, p. 31, lines 746–751.

Revised text: "Compared with OCO-2, OCO-3 provides fewer valid observations over much of the ocean, including the tropical Pacific. This pattern is consistent with the lower sampling rate of OCO-3 in glint mode, which is constrained by the pointing geometry and operational requirements of the ISS platform (Taylor et al., 2023). During normal science operations, clouds, aerosols, solar-illumination conditions, and quality screening can further modulate the number of valid passive retrievals, but these retrieval-related effects are distinct from the sampling patterns imposed by orbital and viewing geometry."

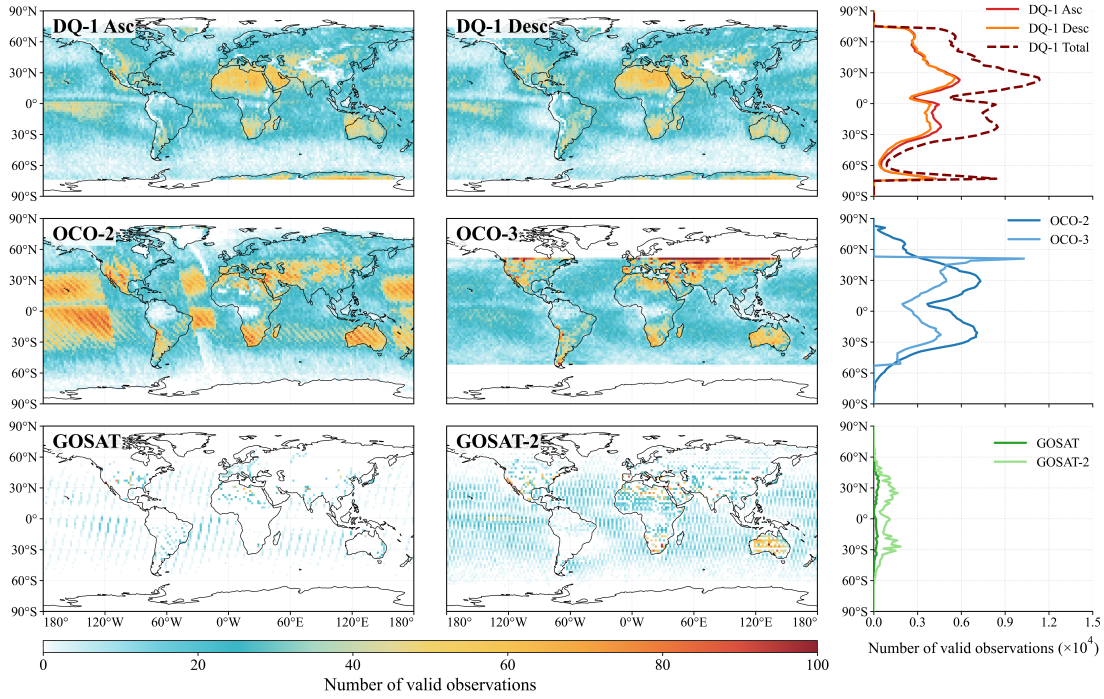


Figure A1. Spatial distributions and zonal totals of cumulative number of valid observations for the satellite XCO_2 products at 2° × 2° resolution.

Comment 3

Overall, GOSAT-2 XCO_2 overestimation looks obvious. What is the reason? I know that the main topic of this manuscript is DQ-1. Once the GOSAT-2 overestimation issue was suggested in this manuscript, however, authors need to explain its reason.

I only found a single statement about this issue in Line 397-398 as next "This systematic positive bias may be related to the fact that the GOSAT-2 SWIR L2 product used here has not been empirically bias corrected", however, I could not figure out the meaning of this statement.

ANSWER: We agree that the original statement was insufficient. We have now applied the official NIES land- and ocean-specific empirical bias corrections to each quality-controlled GOSAT-2 retrieval before daily 2° aggregation and collocation. GOSAT-2 V02.20 is used through February 2024 and V02.21 from March 2024 onward. The subsequent CAMS-based column-definition correction is applied only after collocation

and serves a different purpose: harmonizing the effective XCO2 column definition rather than removing an empirical product bias.

For the same 880 GOSAT-2–TCCON collocations, the empirical correction changes the mean bias from +3.56 to +0.01 ppm and the RMSE from 3.98 to 1.54 ppm, while R changes only slightly from 0.91 to 0.92. Corrected GOSAT-2 is therefore used in the principal inter-satellite comparisons, while Fig. 3d and Tables S2–S3 retain the original values for transparency. The reduction in bias indicates that a substantial fraction of the original offset is associated with the systematic component represented by the official empirical correction. The correction improves product consistency but does not identify the physical origin of the original retrieval offset.

We also added an important qualification from the NIES release notes: the official correction coefficients were derived using TCCON GGG2020 as the reference. Accordingly, the corrected GOSAT-2–TCCON comparison is not fully independent of TCCON. This dependence is now stated in Sect. 4.2, while the DQ-1–TCCON comparison remains an independent observational validation of DQ-1.

Changes in the manuscript :

Sect. 2.1.2, p. 5, lines 137–147.

Revised text: “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. 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."

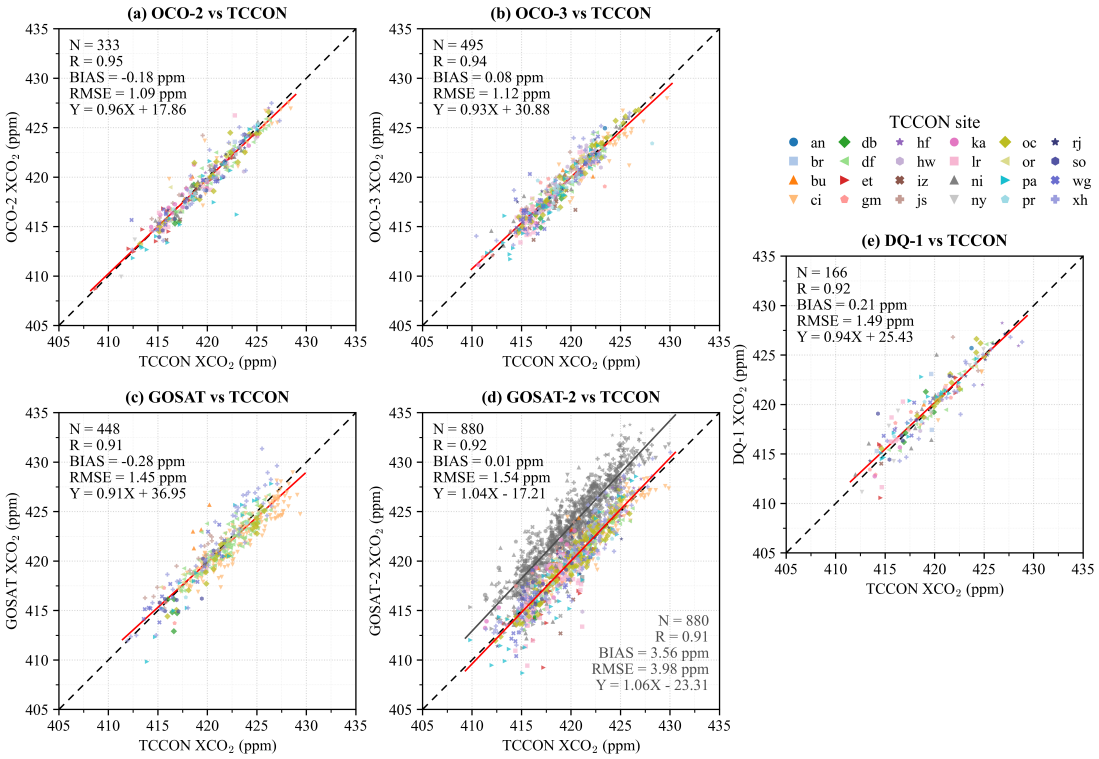


Figure 3. Comparison of satellite XCO_2 products with TCCON observations. (a–e) OCO-2, OCO-3, GOSAT, GOSAT-2, and DQ-1 XCO_2 compared with collocated TCCON XCO_2, respectively. Different colors and symbols denote individual TCCON sites. In panel (d), colored and gray symbols denote empirically bias-corrected 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 4

In Figure 5, the bias looks large over the land of Australia and tropical Pacific. What is the reason to explain this? There was authors' statement in Line 425 "This pattern suggests that the residual differences are more regional and sampling dependent than globally systematic". Again, I could not figure out the meaning of this statement. This statement is consistent with results and discussions in next paragraphs?

ANSWER: We agree that the original sentence stating that the residuals were ‘regional and sampling dependent’ was too vague and inferred more than Fig. 5 alone could establish. We have removed that interpretation. Revised Sect. 3.4.1 now reports the regional statistics first and deliberately postpones physical interpretation to the subsequent CAMS-, reflectance-, and wind-based diagnostics.

Over Australia, the mean DQ-1–OCO-2 and DQ-1–OCO-3 biases are +0.76 and +0.57 ppm, compared with +0.46 and +0.15 ppm over all selected land grid cells. The tropical Pacific shows a different pattern: DQ-1–OCO-2 has a regional mean bias of +0.58 ppm versus +0.51 ppm over all selected ocean grid cells, whereas DQ-1–OCO-3 has +0.30 ppm versus +0.40 ppm. Thus, an enhanced tropical-Pacific mean bias is not common to both OCO comparisons. In contrast, the mean grid-cell RMSE values over the tropical Pacific are 1.27 and 1.26 ppm, both 0.06 ppm above the corresponding all-ocean values of 1.21 and 1.20 ppm (Table S4).

The regional diagnostics are now interpreted separately. Over Australia, DQ-1 and the OCO products show opposite-signed mean departures from the common CAMS-derived reference. Over the tropical Pacific, the same-sample CAMS comparison shows substantially larger DQ-1 scatter (RMSE 1.09/1.08 ppm) than OCO-2/OCO-3 scatter (0.41/0.42 ppm). The reflectance and wind analyses then test surface-condition relationships without treating either factor as a demonstrated cause. This sequence removes the ambiguity of the original sentence and keeps the observation, diagnosis, and limitation distinct.

Changes in the manuscript:

Sect. 3.4.1, p. 20, lines 504–516.

Revised text: “Figure 5 further shows the spatial distributions of the mean DQ-1–OCO-2 and DQ-1–OCO-3 differences, together with the corresponding zonal-mean results. Spatially, most DQ-1–OCO-2 and DQ-1–OCO-3 differences are within approximately ±1 ppm, with no clear global-scale systematic bias. Local positive differences occur mainly over several oceanic sampling bands, Australia, and some low- and middle-latitude land regions, whereas negative differences are scattered over North Africa, Eurasia, and parts of the tropical oceans. Although most grid-cell differences remain within approximately ±1 ppm, several regional features are evident (Fig. 5). Over Australia, the mean DQ-1–OCO-2 and DQ-1–OCO-3 biases are +0.76 and +0.57 ppm, respectively, compared with +0.46 and +0.15 ppm over all selected land grid cells. The tropical Pacific shows a different pattern. For DQ-1–OCO-2, the regional mean bias is +0.58 ppm, slightly higher than the all-ocean value of +0.51 ppm, whereas for DQ-1–OCO-3 the regional mean bias is +0.30 ppm, lower than the corresponding all-ocean value of +0.40 ppm. Thus, an enhanced regional mean bias is not common to both tropical-Pacific comparisons. In contrast, the mean grid-cell RMSE values

are 1.27 and 1.26 ppm, respectively, both 0.06 ppm higher than the corresponding all-ocean values of 1.21 and 1.20 ppm (Table S4). These regional differences are examined further in Sects. 3.4.2 and 3.4.3 using CAMS-derived XCO2, surface reflectance, and ocean wind-speed diagnostics."

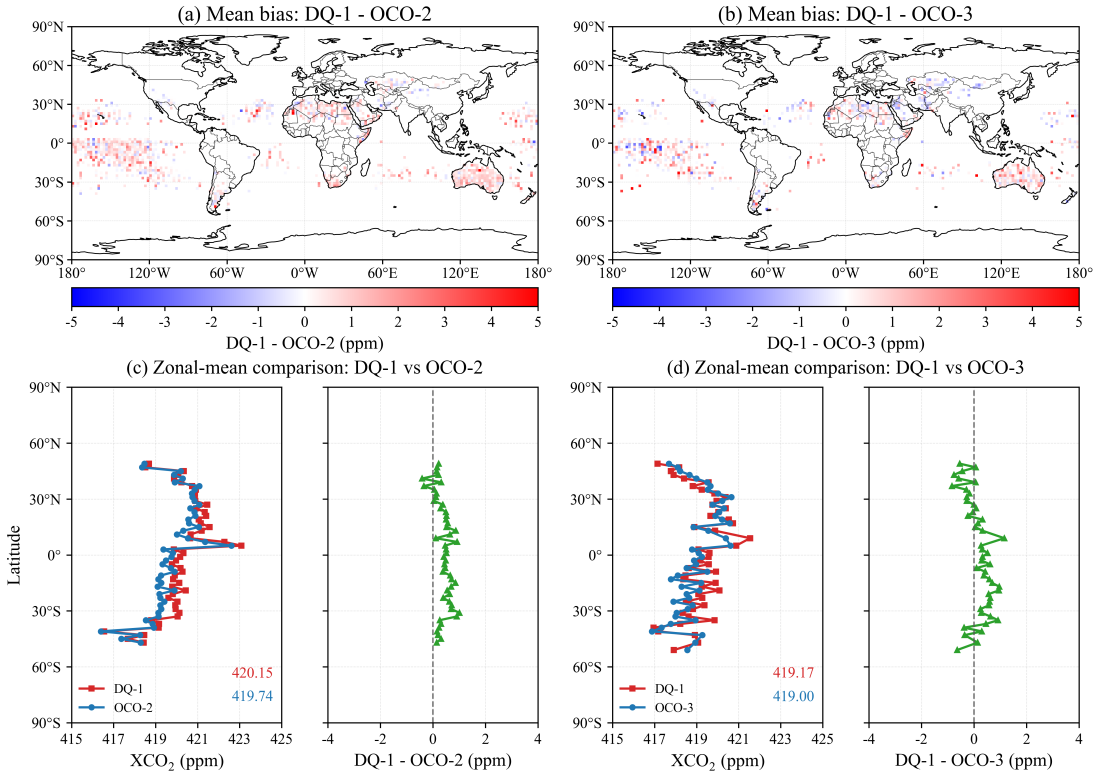


Figure 5. Spatial and zonal differences between DQ-1 and OCO-2/OCO-3 XCO_2 over the selected comparison regions. (a, b) Spatial distributions of the mean DQ-1–OCO-2 and DQ-1–OCO-3 differences, respectively. (c, d) Corresponding zonal-mean XCO_2 values and zonal-mean differences.

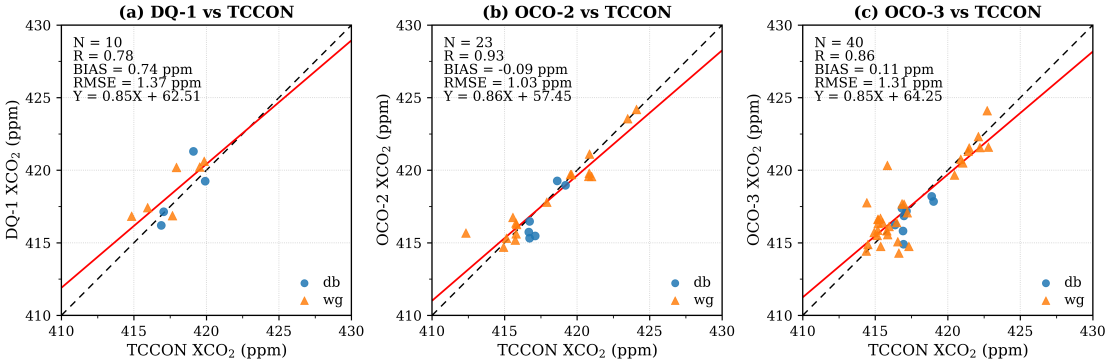


Figure S5. Scatterplot comparisons of DQ-1, OCO-2, and OCO-3 XCO_2 with TCCON observations at the Darwin (db) and Wollongong (wg) sites in Australia. Panels (a–c) show the comparisons for DQ-1, OCO-2, and OCO-3, respectively. The dashed black line denotes the 1:1 line, and the red line shows the linear regression fit.

Table S4. Regional and surface-type grid-cell statistics, MODIS Band 6 reflectance, and ERA5 10 m wind speed for the DQ-1–OCO-2 and DQ-1–OCO-3 comparisons.

Region	Pair	Mean grid-cell Band 6 reflectance	Mean 10 m wind speed (m s^{-1})	Mean grid-cell bias (ppm)	Median grid-cell bias (ppm)	Mean grid-cell RMSE (ppm)
Australia	DQ-1–OCO-2	0.3306	—	+0.76	+0.72	1.17
	DQ-1–OCO-3	0.3270	—	+0.57	+0.48	1.04
Tropical Pacific	DQ-1–OCO-2	0.0508	6.48	+0.58	+0.40	1.27
	DQ-1–OCO-3	0.0348	6.73	+0.30	+0.28	1.26
Sahara	DQ-1–OCO-2	0.5834	—	+0.51	+0.47	1.11
	DQ-1–OCO-3	0.5768	—	+0.22	+0.19	0.96
All selected land	DQ-1–OCO-2	0.3970	—	+0.46	+0.46	1.10
	DQ-1–OCO-3	0.4007	—	+0.15	+0.11	1.04
All selected ocean	DQ-1–OCO-2	0.0374	5.93	+0.51	+0.36	1.21
	DQ-1–OCO-3	0.0313	5.96	+0.40	+0.21	1.20

Note: Regional statistics were calculated by equally weighting the contributing 2° × 2° grid cells. MODIS Band 6 reflectance was averaged over grid cells with valid same-day MYD09CMG.061 retrievals. ERA5 10 m wind speed is reported only for the Tropical Pacific and all selected ocean samples. Bias is defined as DQ-1 minus OCO.

Comment 5

Figure 6 (XCO_2 comparison between DQ-1 and OCO-2) looks similar Figure 7 (XCO_2 comparison between DQ-1 and OCO-2). In tropical Pacific, the validation data number in OCO-3 (Fig. 7) is smaller than those in OCO-2 (Fig. 6), probably due to the data number difference (Fig. 1). Based on Figure 2, I can notice that the data number difference between OCO-2 and OCO-3 suddenly becomes larger after August 2023 for a year.

What happen in this period? This malfunction of OCO-3 can make the data difference in tropical Pacific between OCO-2 and OCO-3? (Over Australia, RMSE data points look similar between OCO-2 and OCO-3). Is there any special issue of OCO-3 for only ocean surface monitoring?

ANSWER: The reduced OCO-3 availability in late 2023–mid-2024 reflects a mission-level operational interruption associated with temporary storage on the ISS; it is not evidence of an ocean-specific instrument malfunction. Appendix A1 now separates this interruption from normal-operation retrieval losses and from the persistent ocean-sampling constraints associated with ISS viewing and glint-pointing geometry (Taylor et al., 2023).

We also performed a common-operation test excluding October 2023–July 2024. Across the remaining 21 full months, OCO-2 and OCO-3 provide 17,127 and 8,734 valid tropical-Pacific observations, respectively, so OCO-3 still provides only about 51 % of the OCO-2 sampling outside the interruption. This confirms that the persistent sampling difference is not created solely by the gap.

More importantly, the tropical-Pacific residual statistics are essentially unchanged when the interruption period is excluded: DQ-1–OCO-2 bias/RMSE changes from 0.58/1.27 to 0.61/1.26 ppm, and DQ-1–OCO-3 changes from 0.30/1.26 to 0.30/1.28 ppm. Thus, the interruption reduces OCO-3 sampling availability but is not the main cause of the regional DQ-1–OCO-3 residual characteristics.

Changes in the manuscript:

Appendix A1, p. 32, lines 755–764.

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."

Appendix A1, pp. 32–33, lines 765–776.

Revised text: "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."

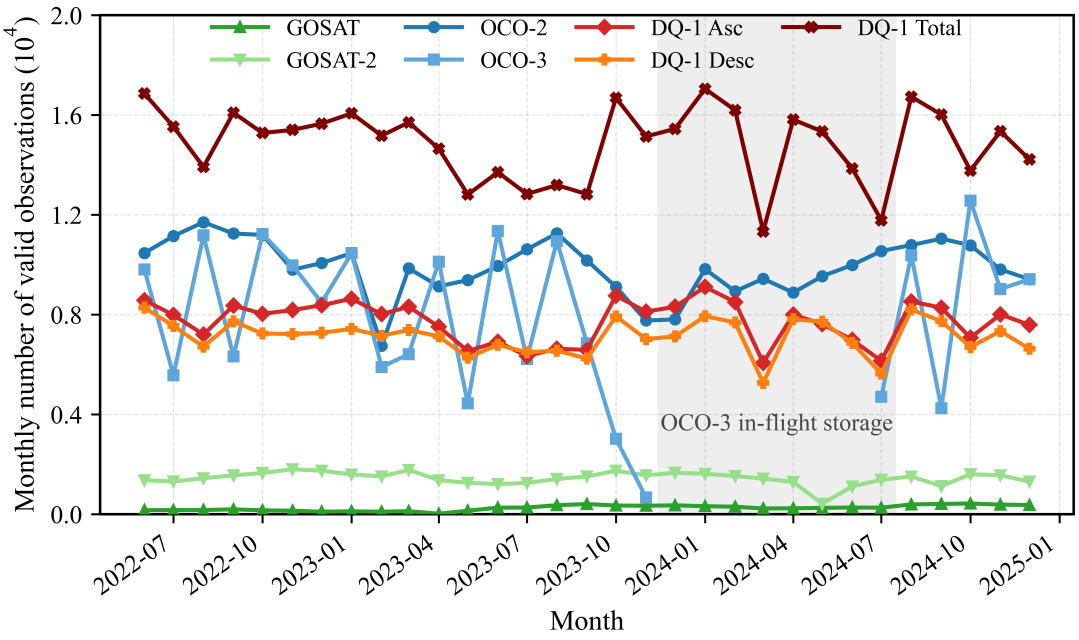


Figure A2. Monthly variations in the number of valid observations for the satellite XCO_2 products. The gray-shaded region indicates the OCO-3 data gap associated with its temporary storage on the ISS.

Comment 6

In Line 447, there is a statement as next "These patterns suggest that the spatial characteristics of DQ-1–OCO-2/OCO-3 bias and RMSE cannot be explained by surface reflectance alone, and may also be related to ocean glint observation conditions and the distribution of valid collocated samples".

How the ocean glint differs? Is there any clue that can be applied to high RMSE in tropical Pacific ocean? Since the bias is especially large over this tropical Pacific (different from other ocean area), authors need to add more detailed statement about this part, at least based on previous study results.

ANSWER: We expanded this discussion and added an explicit sea-surface diagnostic. The revised analysis distinguishes two issues. First, MODIS Band 6 reflectance is a contextual surface indicator rather than the actual OCO glint radiance or sun–sensor geometry. Over ocean, its relationship with the residual statistics is pronounced but pair dependent: DQ-1–OCO-2 bias and RMSE increase with reflectance ($R = +0.77$ and $+0.64$), whereas DQ-1–OCO-3 bias and RMSE decrease ($R = -0.77$ and -0.78). Thus,

reflectance is associated with the oceanic residuals, but not through a common response shared by OCO-2 and OCO-3.

Second, the ERA5 wind-speed diagnostic shows weak correlations at the individual 2° grid-cell level (bias R = −0.16/−0.29; RMSE R = −0.03/−0.08), but the 1 m s−1 bin means decrease systematically with increasing wind speed (DQ-1–OCO-2: R = −0.97/−0.81 for bias/RMSE; DQ-1–OCO-3: R = −0.94/−0.89). The strong binned trends together with the large within-bin spread indicate a shift in the conditional mean with wind speed, while substantial grid-cell variability remains. This is why the revised text no longer characterizes the wind relationship simply as 'weak.'

These diagnostics also argue against explaining the tropical-Pacific RMSE simply by stronger winds: mean wind speeds there are slightly higher than the selected-ocean averages (6.48/6.73 versus 5.93/5.96 m s−1), yet the RMSE is also slightly higher. ERA5 10 m wind is only an indirect roughness proxy and does not reconstruct the instantaneous sun–sensor glint geometry. The same-sample CAMS comparison provides an additional clue: tropical-Pacific RMSE is 1.09/1.08 ppm for DQ-1 versus 0.41/0.42 ppm for OCO-2/OCO-3. We therefore state that the enhanced scatter is more evident on the DQ-1 side under these collocated conditions, without uniquely attributing it to glint, surface roughness, or a specific retrieval mechanism.

Changes in the manuscript :

Sect. 3.4.3, pp. 22–23, lines 546–584.

Revised text: "To further diagnose the influence of surface reflectance on the residuals, we calculate reflectance-binned statistics for the bias and RMSE of DQ-1–OCO-2 and DQ-1–OCO-3. For all samples, the linear relationship between bias and reflectance is weak, with correlation coefficients of approximately R = −0.22 for DQ-1–OCO-2 and R = −0.24 for DQ-1–OCO-3. This indicates that, when all samples are considered, reflectance is not the dominant factor controlling the mean bias. For samples over land, the DQ-1–OCO-2 bias increases slightly with reflectance, with R = 0.42, whereas the DQ-1–OCO-3 bias shows a weaker relationship with reflectance, with R = −0.21. For samples over ocean, the correlations are stronger, but the two comparison pairs show opposite trends: the DQ-1–OCO-2 bias increases with reflectance, with R = 0.77, whereas the DQ-1–OCO-3 bias decreases with reflectance, with R = −0.77. Thus, surface reflectance is clearly associated

with the oceanic bias statistics, but the direction of the association is pair dependent: positive for DQ-1–OCO-2 and negative for DQ-1–OCO-3.

RMSE shows a similar pattern. For all samples, the correlations between RMSE and reflectance are weak for both DQ-1–OCO-2 and DQ-1–OCO-3, with correlation coefficients of approximately R = −0.24 and R = −0.35, respectively. The correlations remain limited for samples over land, with R = −0.17 and R = −0.42, respectively. By contrast, the RMSE–reflectance relationships are more pronounced for samples over ocean, but again show opposite trends for the two comparison pairs: the RMSE of DQ-1–OCO-2 over ocean increases with reflectance, with R = 0.64, whereas that of DQ-1–OCO-3 decreases with reflectance, with R = −0.78. The corresponding RMSE relationships show the same pair dependence. Taken together, the ocean results indicate pronounced but opposite-signed reflectance associations rather than a common reflectance response across the two OCO comparisons.

The ERA5 wind-speed analysis reveals different behavior at the individual-grid-cell and binned levels (Fig. S6). At the individual 2° grid-cell scale, the correlations with wind speed are weak, with R = −0.16 and −0.29 for bias and R = −0.03 and −0.08 for RMSE for DQ-1–OCO-2 and DQ-1–OCO-3, respectively. After grouping the grid cells into 1 m s−1 wind-speed bins, however, both the bin-mean bias and RMSE decrease with increasing wind speed, with R = −0.97 and −0.81 for DQ-1–OCO-2 and R = −0.94 and −0.89 for DQ-1–OCO-3, respectively. The strong binned relationships, together with the large spread of individual grid cells within the bins, indicate a systematic change in the mean residual statistics with wind speed, while substantial grid-cell variability remains.

The tropical Pacific has slightly higher mean wind speeds than the selected ocean samples as a whole, but its mean grid-cell RMSE is also slightly higher (Table S4). Thus, the elevated tropical-Pacific RMSE cannot be explained simply by stronger near-surface winds. Moreover, ERA5 10 m wind speed is only an indirect indicator of sea-surface roughness and does not describe the instantaneous sun–sensor glint geometry of individual OCO observations. The present diagnostics therefore indicate that ocean-surface conditions are associated with the residual statistics, but they do not identify a single reflectance- or wind-related mechanism that explains the tropical-Pacific differences.”

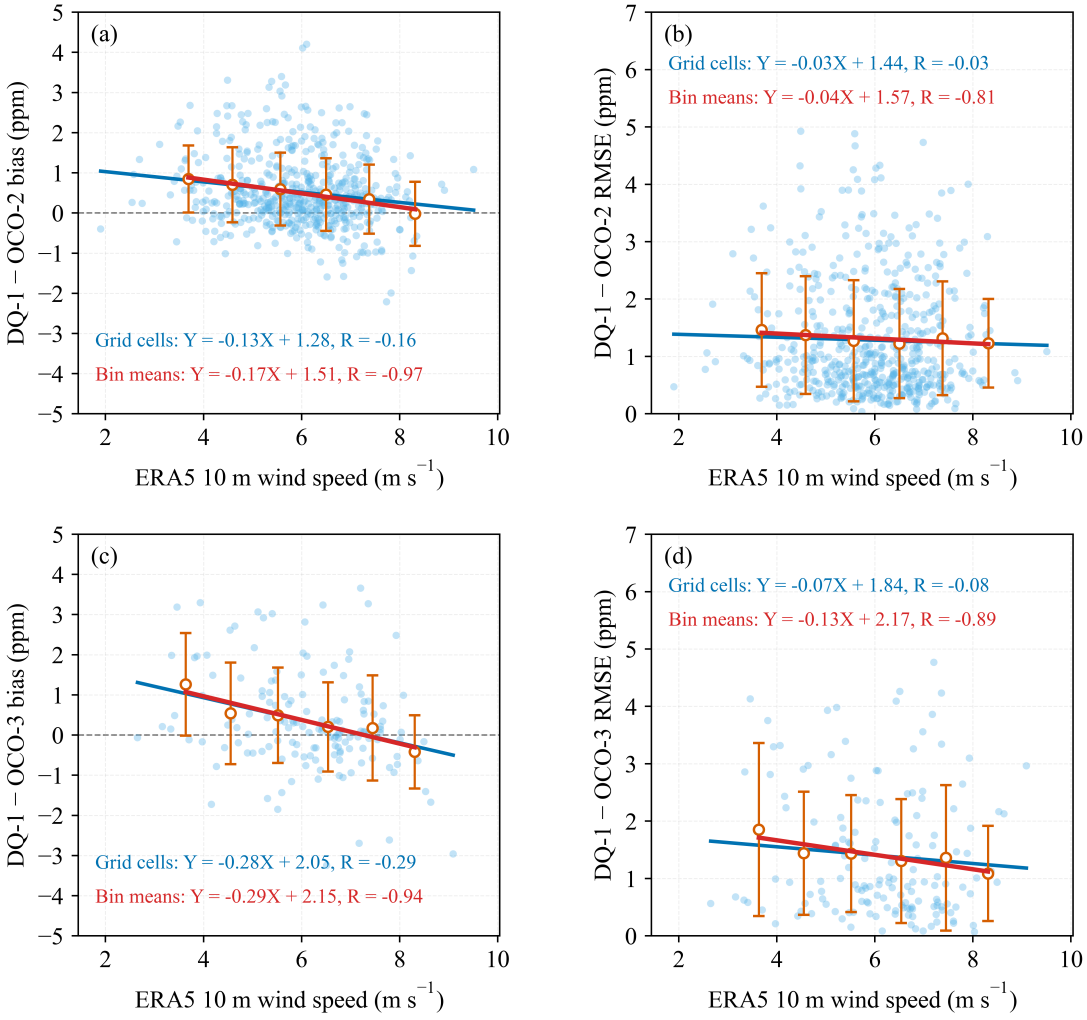


Figure S6. Dependence of oceanic DQ-1–OCO-2 and DQ-1–OCO-3 bias and RMSE on ERA5 10 m wind speed. Panels (a, b) show DQ-1–OCO-2 and panels (c, d) show DQ-1–OCO-3. Light blue dots denote individual oceanic 2° grid cells, while orange circles and error bars show 1 m s−1 wind-speed-bin means ±1 standard deviation. Blue and red lines show linear least-squares fits to the individual grid-cell values and displayed bin means, respectively; the corresponding regression equations and Pearson correlation coefficients are annotated in each panel. Bias is defined as DQ-1 minus OCO.

Comment 7

Please cite figure number appropriately after important statements (especially in chapter 3.4). I cannot make good connection between written statements and suggested figures.

ANSWER: Revised. We have checked the cross-references throughout Sect. 3.4. Figure 5 accompanies the spatial/zonal residual discussion; Figs. S3–S5 and Table S5

accompany the CAMS and regional TCCON diagnostics; Figs. 6–7 accompany the reflectance analysis; and Fig. S6 and Table S4 accompany the wind-speed analysis. This makes each quantitative statement traceable to the corresponding figure or table.

Changes in the manuscript:

Sect. 3.4.1, p. 20, lines 504–516.

Revised text: "Figure 5 further shows the spatial distributions of the mean DQ-1–OCO-2 and DQ-1–OCO-3 differences, together with the corresponding zonal-mean results. Spatially, most DQ-1–OCO-2 and DQ-1–OCO-3 differences are within approximately ±1 ppm, with no clear global-scale systematic bias. Local positive differences occur mainly over several oceanic sampling bands, Australia, and some low- and middle-latitude land regions, whereas negative differences are scattered over North Africa, Eurasia, and parts of the tropical oceans. Although most grid-cell differences remain within approximately ±1 ppm, several regional features are evident (Fig. 5). Over Australia, the mean DQ-1–OCO-2 and DQ-1–OCO-3 biases are +0.76 and +0.57 ppm, respectively, compared with +0.46 and +0.15 ppm over all selected land grid cells. The tropical Pacific shows a different pattern. For DQ-1–OCO-2, the regional mean bias is +0.58 ppm, slightly higher than the all-ocean value of +0.51 ppm, whereas for DQ-1–OCO-3 the regional mean bias is +0.30 ppm, lower than the corresponding all-ocean value of +0.40 ppm. Thus, an enhanced regional mean bias is not common to both tropical-Pacific comparisons. In contrast, the mean grid-cell RMSE values are 1.27 and 1.26 ppm, respectively, both 0.06 ppm higher than the corresponding all-ocean values of 1.21 and 1.20 ppm (Table S4). These regional differences are examined further in Sects. 3.4.2 and 3.4.3 using CAMS-derived XCO2, surface reflectance, and ocean wind-speed diagnostics."

Comment 8

Figure 8 (DQ-1 vs. CAMS-derived DQ-1) is interesting. The RMSE is also high over Australia, Sahara desert, and tropical Pacific, similar to FIgure 6 and 7. Then my question is how about the comparison results between OCO and CAMS-derived XCO_2? If they are closer, then DQ-1 research team can achieve some tips about bias factors.

ANSWER: We added the requested OCO–CAMS comparisons in Figs. S3–S4 and Table S5. DQ-1 and each OCO product are evaluated against their corresponding CAMS-derived XCO2 on exactly the same DQ-1–OCO collocated sample set, which

controls sample-set differences when comparing the two products with the model reference.

The result is region dependent. Over the tropical Pacific, OCO-2/OCO-3 are substantially closer to CAMS in RMSE (0.41/0.42 ppm) than DQ-1 (1.09/1.08 ppm), so the enhanced scatter is more evident on the DQ-1 side under these collocated conditions. Over Australia, DQ-1 RMSEs (0.63/0.56 ppm) are lower than those of OCO-2/OCO-3 (0.80/0.69 ppm), and DQ-1 and OCO show opposite-signed mean departures from CAMS. Over the Sahara, all corresponding RMSEs are below 1 ppm. Therefore, the analysis does not support a uniform statement that OCO is closer to CAMS everywhere.

These additions are daytime diagnostics. The nighttime DQ-1–CAMS analysis is retained separately because passive NIR/SWIR instruments cannot provide an equivalent nighttime comparison. CAMS is used only as a common diagnostic reference rather than as observational truth, so closer agreement with CAMS does not uniquely assign an error to either satellite retrieval.

Changes in the manuscript:

Sect. 3.4.2, pp. 21–22, lines 530–544.

Revised text: “To further diagnose these regional differences, DQ-1 and OCO-2/OCO-3 were independently compared with CAMS-derived XCO2 using exactly the same collocated samples (Figs. S3–S4 and Table S5). Over Australia, DQ-1 shows small positive biases of +0.22 and +0.23 ppm relative to CAMS-derived XCO2 for the OCO-2 and OCO-3 collocations, respectively, whereas OCO-2 and OCO-3 show negative biases of −0.53 and −0.35 ppm. Thus, the positive DQ-1–OCO differences over Australia are accompanied by opposite-signed departures of DQ-1 and the OCO products from the common CAMS-derived reference. A site-based comparison with TCCON at Darwin and Wollongong provides an independent regional check and likewise shows a positive DQ-1 bias, although only 10 DQ-1 collocations are available (Fig. S5). Over the tropical Pacific, the DQ-1–CAMS-derived XCO2 RMSE values are 1.09 and 1.08 ppm, compared with 0.41 and 0.42 ppm for OCO-2 and OCO-3, respectively. In contrast, over the Sahara, both DQ-1 and the OCO products show RMSE values below 1 ppm relative to CAMS-derived XCO2, and the contrast between the two products is much smaller than that over the tropical Pacific (Table S5). Over the

tropical Pacific, the substantially larger DQ-1 RMSE relative to the common CAMS-derived reference suggests that the enhanced scatter is more pronounced in DQ-1 under these collocated conditions. This differs from Australia, where DQ-1 and the OCO products show opposite-signed mean departures from CAMS. CAMS is used here only as a common diagnostic reference rather than as observational truth; these comparisons therefore help identify relative product behavior but do not uniquely attribute retrieval errors."

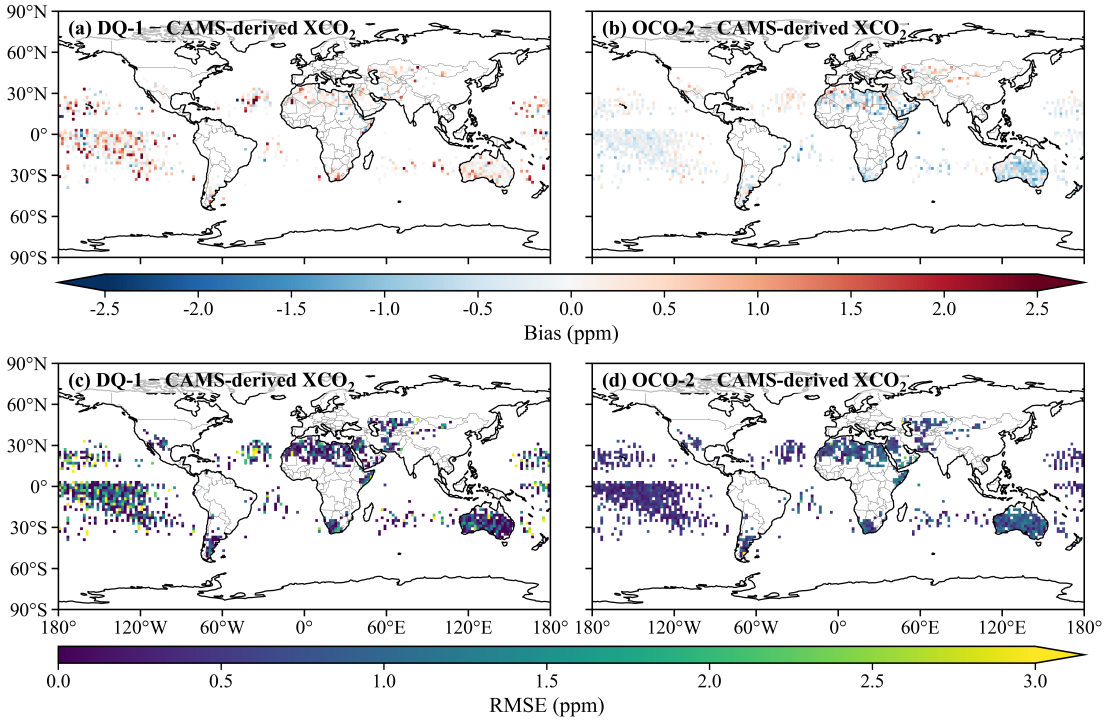


Figure S3. Spatial distributions of daytime bias and RMSE relative to CAMS-derived XCO_2 for DQ-1 and OCO-2, based on the same DQ-1–OCO-2 collocated samples. Panels (a, b) show the biases for DQ-1 and OCO-2, respectively, and panels (c, d) show the corresponding RMSEs. Bias is defined as satellite XCO_2 − CAMS-derived XCO_2.

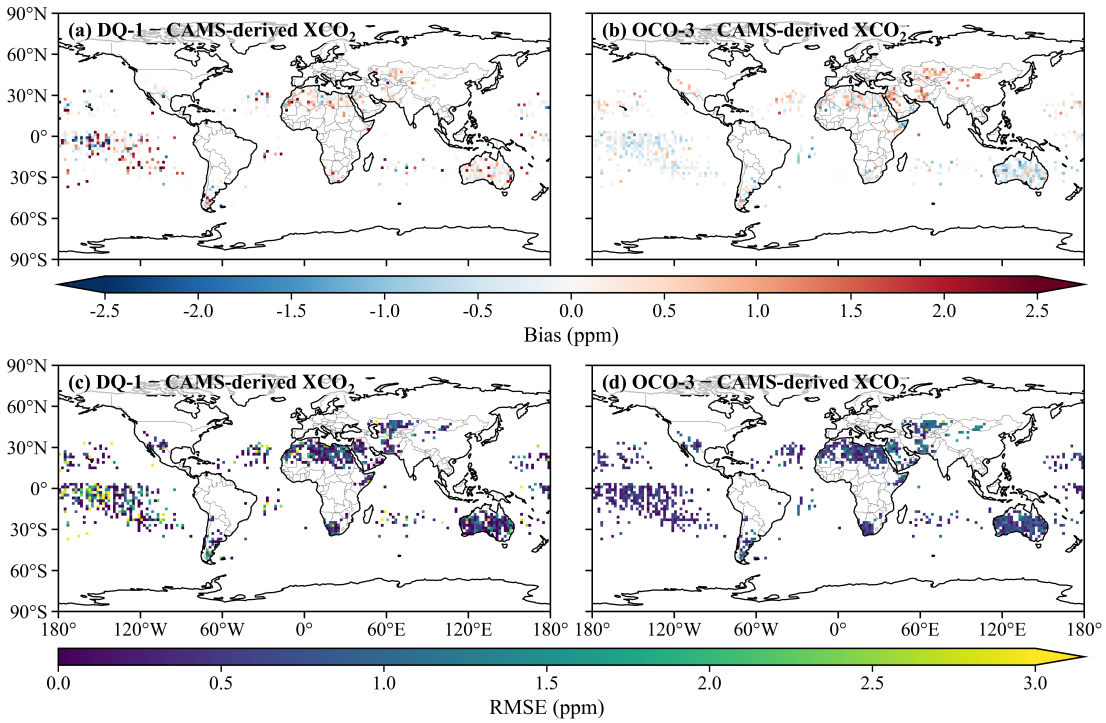


Figure S4. Spatial distributions of daytime bias and RMSE relative to CAMS-derived XCO_2 for DQ-1 and OCO-3, based on the same DQ-1–OCO-3 collocated samples. Panels (a, b) show the biases for DQ-1 and OCO-3, respectively, and panels (c, d) show the corresponding RMSEs. Bias is defined as satellite XCO_2 − CAMS-derived XCO_2.

Table S5. Regional daytime statistics relative to CAMS-derived XCO_2 for the DQ-1–OCO-2 and DQ-1–OCO-3 comparisons.

Region	Pair	DQ-1–CAMS-derived XCO_2 mean bias (ppm)	OCO–CAMS-derived XCO_2 mean bias (ppm)	DQ-1–CAMS-derived XCO_2 mean RMSE (ppm)	OCO–CAMS-derived XCO_2 mean RMSE (ppm)
Australia	DQ-1–OCO-2	+0.22	-0.53	0.63	0.80
	DQ-1–OCO-3	+0.23	-0.35	0.56	0.69
Tropical Pacific	DQ-1–OCO-2	+0.37	-0.20	1.09	0.41
	DQ-1–OCO-3	+0.05	-0.25	1.08	0.42
Sahara	DQ-1–OCO-2	+0.16	-0.36	0.51	0.80
	DQ-1–OCO-3	+0.22	+0.00	0.60	0.62
All selected land	DQ-1–OCO-2	+0.16	-0.29	0.58	0.79
	DQ-1–OCO-3	+0.18	+0.03	0.60	0.71
All selected ocean	DQ-1–OCO-2	+0.37	-0.14	1.00	0.45
	DQ-1–OCO-3	+0.29	-0.11	0.96	0.43

Note: Values are equally weighted across contributing 2° × 2° grid cells. CAMS-derived XCO2 is used as a common diagnostic reference rather than as truth. Bias is defined as satellite XCO2 minus CAMS-derived XCO2.

Comment 9

Please add the site name in the caption of Figure 10 (there are many abbreviation in the legend, but I do now know what those are).

ANSWER: The original Fig. 10 is now Fig. 3. Its revised caption explicitly directs readers to Table A2, which gives the full name, abbreviation, coordinates, altitude, and valid observation period for every TCCON site. This makes the abbreviations identifiable without repeating the complete 24-site list in the caption.

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

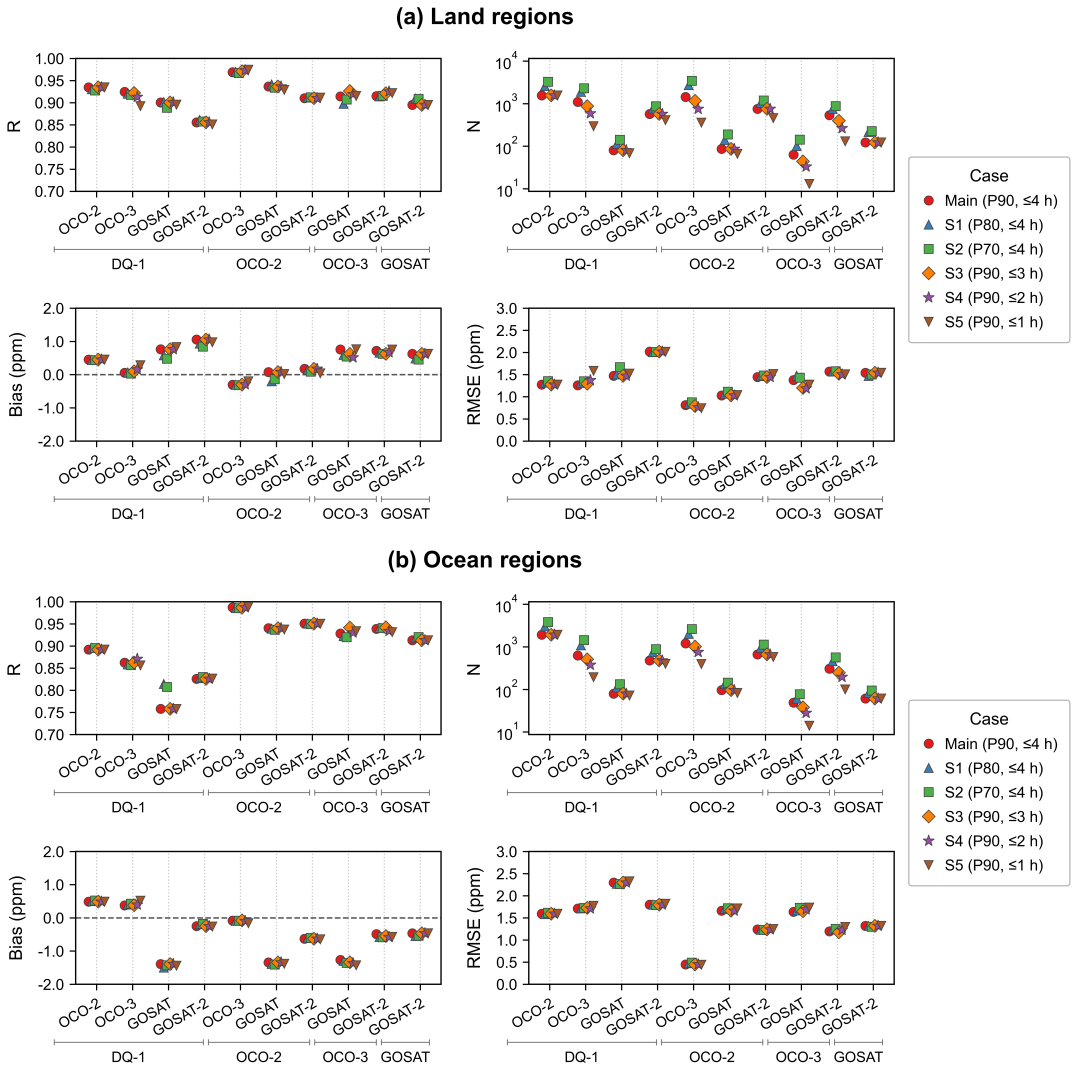


Figure A3. Sensitivity of satellite XCO_2 intercomparison statistics to the spatial-selection and temporal-collocation criteria over (a) land and (b) ocean regions. The four panels for each surface type show the correlation coefficient (R), number of collocations (N), mean bias, and RMSE for the six 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."

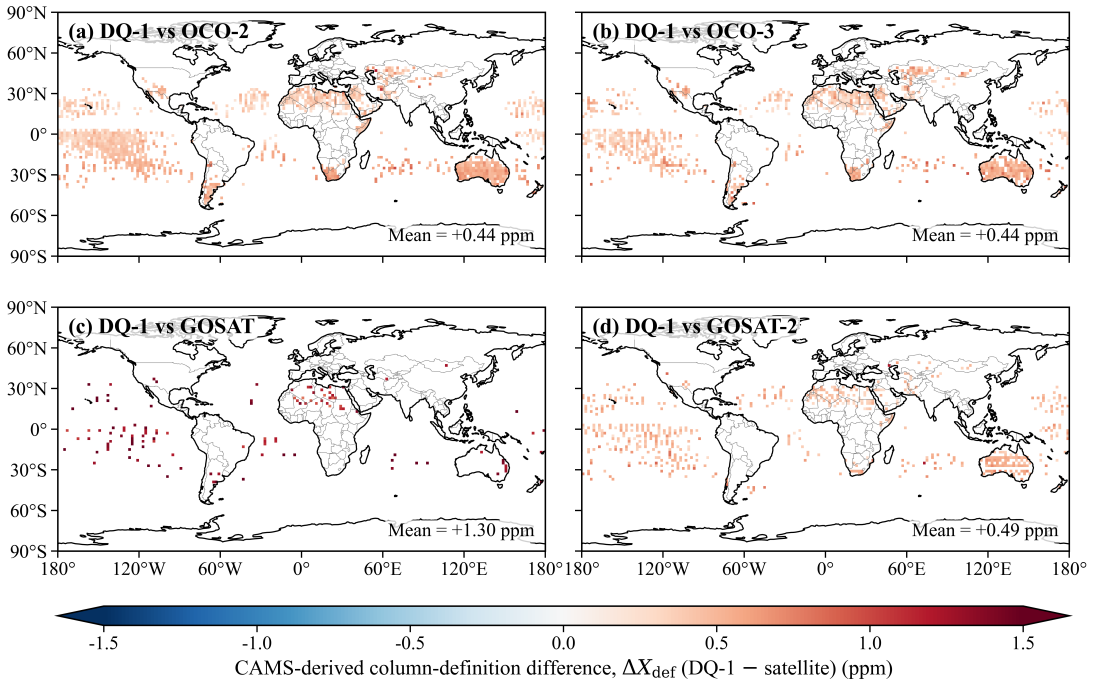


Figure S1. Spatial distributions of the CAMS-derived XCO_2 column-definition differences between DQ-1 and (a) OCO-2, (b) OCO-3, (c) GOSAT, and (d) GOSAT-2 for the collocated samples. The column-definition difference is defined as XCO_2 (DQ-1 definition) − XCO_2 (satellite definition), and the mean value 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)
GOSAT-2 (original)	GOSAT	Ocean	307	0.94	-0.42	1.18	0.94	-0.49	1.20
		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."

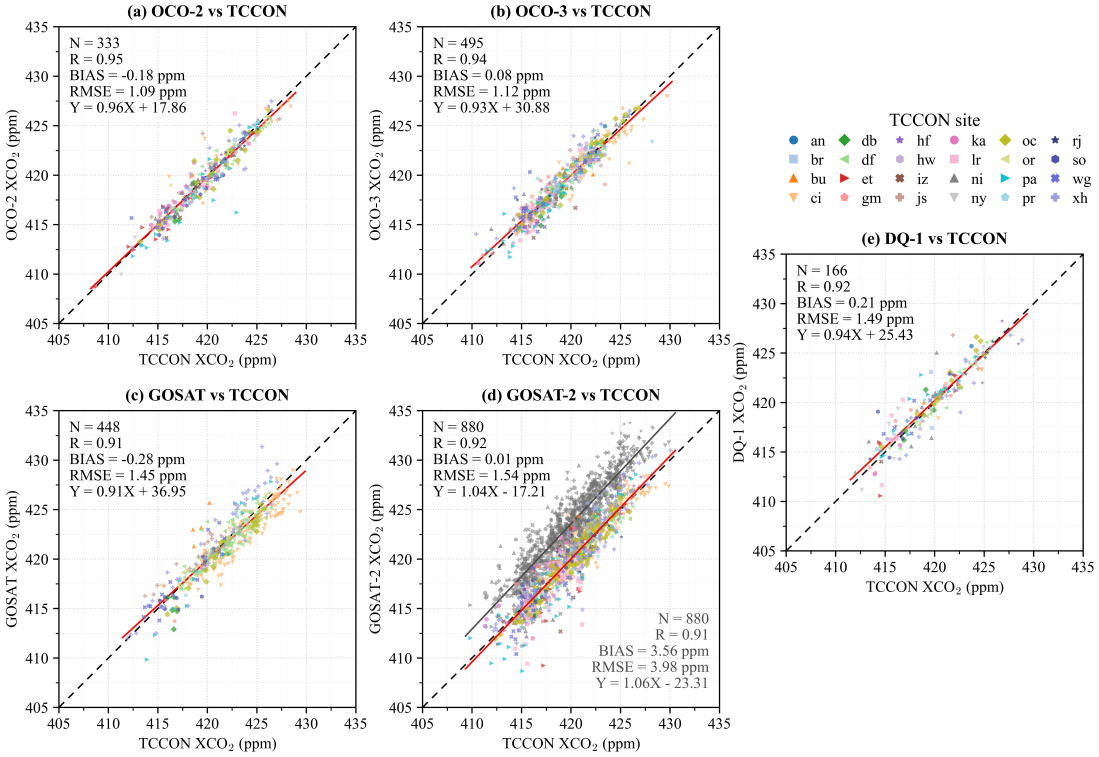


Figure 3. Comparison of satellite XCO_2 products with TCCON observations. (a–e) OCO-2, OCO-3, GOSAT, GOSAT-2, and DQ-1 XCO_2 compared with collocated TCCON XCO_2, respectively. Different colors and symbols denote individual TCCON sites. In panel (d), colored and gray symbols denote empirically bias-corrected 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.”

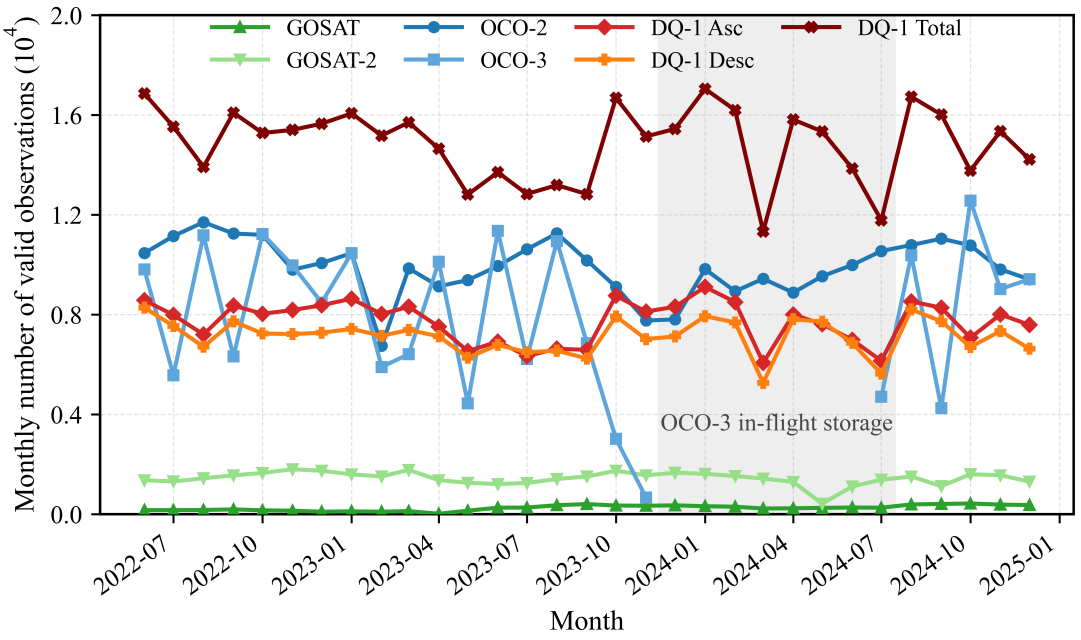


Figure A2. Monthly variations in the number of valid observations for the satellite XCO_2 products. The gray-shaded region indicates the OCO-3 data gap associated with its temporary storage on the ISS.

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.

Reply to Community Comment #1

Dear Kai Qin,

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 study presents a systematic comparison of DQ-1 ACDL XCO_2 with passive satellite products from OCO-2, OCO-3, GOSAT, and GOSAT-2. Wang et al. develop a unified comparison framework that combines CAMS-based assessment of background-field spatiotemporal coherence, multi-satellite sampling availability, daily 2° spatial aggregation, spatiotemporal collocation, and correction for differences in XCO_2 column definitions among the satellite products.

The study further investigates the remaining inter-product differences through surface-reflectance analysis, validation against TCCON observations, nighttime comparisons with an external model reference, and regional sampling analyses.

The datasets used in this study are comprehensive, and the proposed framework provides a useful basis for assessing the comparability and potential joint use of active and passive satellite XCO_2 observations. The manuscript is generally well structured and relevant to the scope of Atmospheric Chemistry and Physics. However, several

methodological details and some of the principal conclusions require further clarification and quantitative support.

I recommend publication after the following comments have been adequately addressed.

ANSWER: We thank Dr. Qin for the careful and constructive assessment. We have revised the manuscript and Supplement substantially to clarify the active/passive observation operators and novelty, add an overall workflow, quantify the effect of the CAMS-based column-definition harmonization, apply the official GOSAT-2 empirical correction, strengthen TCCON uncertainty/sampling information, and address the technical and presentation issues listed below.

Main comments

Comment Main 1

Although both DQ-1 ACDL and passive satellite products provide XCO_2 retrievals, the two approaches differ in their retrieval formulations and effective vertical weighting. Please add a brief comparison in Sects. 2.1.1–2.1.2, noting that passive XCO_2 retrievals depend on prior CO_2 profiles and are characterized by column averaging kernels, whereas ACDL derives a weighted column from differential absorption measurements and its weighting functions.

This clarification would help readers understand that the original active and passive XCO_2 products should not be interpreted as directly equivalent without accounting for differences in their effective column definitions, and would provide the necessary context for the column-definition correction in Sect. 2.2.3.

ANSWER: We have clarified that the reported active and passive XCO_2 quantities have different effective vertical sensitivities. Passive retrievals use product-specific prior profiles, pressure weights, and column averaging kernels. DQ-1 ACDL instead derives a path-integrated weighted column from online/offline differential absorption, with vertical sensitivity represented by its weighting function and effective pressure layers.

The revised instrument descriptions introduce this distinction, and Sect. 2.2.3 expresses it through the corresponding observation operators. A common CAMS profile is passed through each operator to estimate the column-definition difference, which is then used for harmonization. This treatment addresses the modeled difference in column definitions.

Changes in the manuscript:

Sect. 2.1.2, pp. 5–6, lines 145–147.

Revised text: "Although all products report XCO_2, 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. 2.2.3, p. 12, lines 307–312.

Revised text: "The active and passive satellite XCO_2 products do not necessarily share the same effective column definition because they differ in prior profiles, pressure weighting functions, column averaging kernels, and vertical sensitivities. If the retrieved XCO_2 values are compared directly, part of the difference may arise from differences in the observation operators, rather than from real atmospheric CO_2 differences. To reduce this effect, we use CAMS 3-hourly CO_2 profiles as a common external reference profile and apply the observation operator of each satellite to these profiles to estimate the theoretical difference caused by XCO_2 column-definition differences."

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."

Comment Main 2

The novelty of the manuscript could be stated more clearly relative to previous DQ-1 validation studies and intercomparisons among passive satellite products.

Please clarify in the final part of the Introduction, and briefly reinforce in the Conclusions, that the principal contribution of this study is the systematic comparison of DQ-1 with OCO-2, OCO-3, GOSAT, and GOSAT-2 within a common framework that jointly accounts for background-field coherence, multi-sensor sampling availability, daily gridded collocation, observation-time differences, and differences in XCO_2 column definitions among the satellite products.

The authors should also explain briefly how this framework extends beyond previous validation of individual DQ-1 products and conventional intercomparisons among passive satellite products.

ANSWER: We have revised the final part of the Introduction and the Conclusions to identify the contribution more precisely. The contribution is the coordinated comparison of DQ-1 with OCO-2, OCO-3, GOSAT, and GOSAT-2 while jointly addressing background-field coherence, sampling availability, daily aggregation, observation-time differences, and effective column definitions.

Earlier DQ-1 studies established retrieval performance and validation, and earlier passive-satellite intercomparisons established important product-consistency methods. Our extension is to bring the distinct active and passive observation operators and sampling characteristics into the same comparison sequence. We therefore changed the title from "First comparison" to "Assessing the consistency of XCO_2 products from DQ-1 ACDL and passive satellites using a unified comparison framework". We also state that the resulting consistency applies primarily to the selected coherent and well-sampled regions, rather than presenting it as a global absolute-accuracy claim.

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 Main 3

The study contains several interconnected processing steps, including comparison-region selection, satellite-data aggregation and collocation, column-definition correction, inter-satellite comparisons, TCCON validation, nighttime DQ-1 assessment, surface-reflectance analysis, and regional sampling evaluation. However, the overall methodological structure is difficult to follow from the text alone.

I suggest adding a workflow figure showing the input datasets, principal processing steps, and relationships among the different analyses. In particular, the different roles of CAMS in comparison-region selection, column-definition correction, and nighttime DQ-1 assessment should be clearly distinguished.

ANSWER: We have added a workflow diagram as Fig. 1 showing the input datasets, product-specific quality control, comparison-region selection, daily aggregation and collocation, column-definition harmonization, TCCON validation, inter-satellite comparisons, regional surface diagnostics, nighttime DQ-1 assessment, and regional sampling analysis. The figure explicitly separates the reference roles of TCCON and CAMS. The workflow figure is intended to clarify the logical sequence and relationships among the analyses rather than only summarize data-processing steps.

We also revised Sect. 2.1.3 to distinguish the three uses of CAMS: (1) CAMS column XCO2 supports the background-field coherence assessment; (2) CAMS profiles provide a common atmospheric input to the satellite-specific observation operators for column-definition harmonization; and (3) CAMS-derived XCO2 serves as an external model reference for diagnostic comparisons, including nighttime DQ-1. TCCON is identified separately as the ground-based observational reference for daytime validation. This avoids interpreting the different CAMS uses as independent validations or as observational truth.

Changes in the manuscript:

Introduction, p. 3, lines 92–99.

Revised text: "In this study, we compare DQ-1 ACDL with OCO-2, OCO-3, GOSAT, and GOSAT-2 XCO_2 products from June 2022 to December 2024 within a unified active–passive framework. We first identify coherent and well-sampled comparison regions, evaluate the satellite products against TCCON, and quantify inter-satellite consistency after daily 2° aggregation, collocation, and correction for XCO_2 column-definition differences. We then diagnose the remaining regional differences, assess nighttime DQ-1 against CAMS-derived XCO_2, and examine regional monthly variability and sampling continuity. The overall workflow is summarized in Fig. 1. Our objective is to evaluate the comparability and complementarity of DQ-1 active IPDA XCO_2 observations relative to passive NIR/SWIR products and to support future active–passive joint XCO_2 analyses, flux inversions, and emission monitoring."

Sect. 2.1.3, p. 6, lines 154–160.

Revised text: "CAMS is used in three main ways in this study. First, CAMS column-averaged XCO_2 is used to evaluate the spatial homogeneity and short-term temporal coherence within each 2° candidate grid cell, thereby identifying background regions that are more suitable for inter-satellite comparison. Second, CAMS model-level CO_2 profiles are used as a common external profile and are passed through the observation operators of different satellites to construct CAMS-derived XCO_2 values consistent with each satellite's column definition. This allows us to estimate the theoretical differences caused by differences in vertical sensitivity and effective column definition. Third, for nighttime DQ-1 observations, CAMS-derived XCO_2 is used as an external model reference to evaluate the consistency between nighttime DQ-1 XCO_2 and the model reference field."

Sect. 2.1.3, p. 6, lines 161–164.

Revised text: "Throughout this study, TCCON is used as the ground-based observational reference for daytime validation, whereas CAMS is used as a model-based common reference for comparison-region selection, column-definition harmonization, and diagnostic analyses. CAMS is not treated as observational truth. The specific uses of CAMS are described in Sects. 2.2.1, 2.2.3, and 2.2.7."

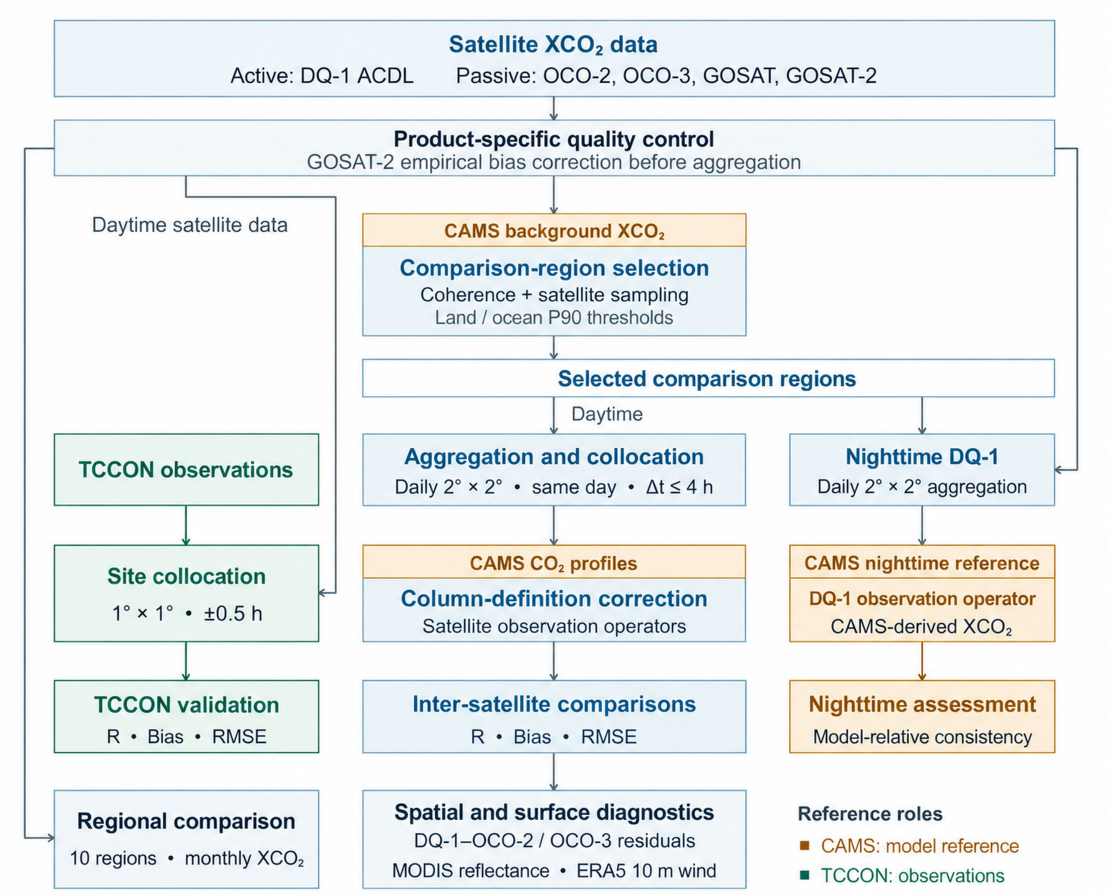


Figure 1. Workflow of the unified active–passive XCO_2 comparison framework.

Comment Main 4

The CAMS-based treatment of XCO_2 column-definition differences is a central component of the proposed framework, but its magnitude and influence on the final results are not sufficiently demonstrated. The authors should present the spatial distributions of the column-definition differences between DQ-1 and OCO-2, OCO-3, GOSAT, and GOSAT-2 for the collocated samples or selected comparison regions, and briefly discuss where these differences are relatively pronounced.

In addition, the correlation coefficient, mean difference, and RMSE should be reported before and after column-definition correction for each satellite pair. This comparison would quantify the actual influence of the correction and show whether and to what extent the correction affects the inter-product statistics and conclusions.

ANSWER: We have added Fig. S1 to show the spatial distributions of the CAMS-derived DQ-1-minus-passive column-definition differences for the collocated samples, and Table S3 to report R, bias, and RMSE before and after harmonization for every satellite pair, separately over land and ocean.

The pair-mean differences in Fig. S1 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 largest pair-mean term is therefore for GOSAT, with pronounced positive values among its comparatively sparse collocations, including sampled land and ocean locations. The OCO and GOSAT-2 maps contain mostly smaller positive differences, with local variation across the sampled regions. These maps describe the collocations and are not continuous global correction fields.

For 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; RMSE decreases 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. Thus, the correction has a material effect on some absolute offsets and does not uniformly improve agreement. R changes by no more than 0.01 in most comparisons.

The conclusions now describe consistency after model-assisted harmonization. We distinguish this operator correction from empirical bias correction and retain uncertainties associated with the CAMS profiles, retrievals, and sampling.

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."

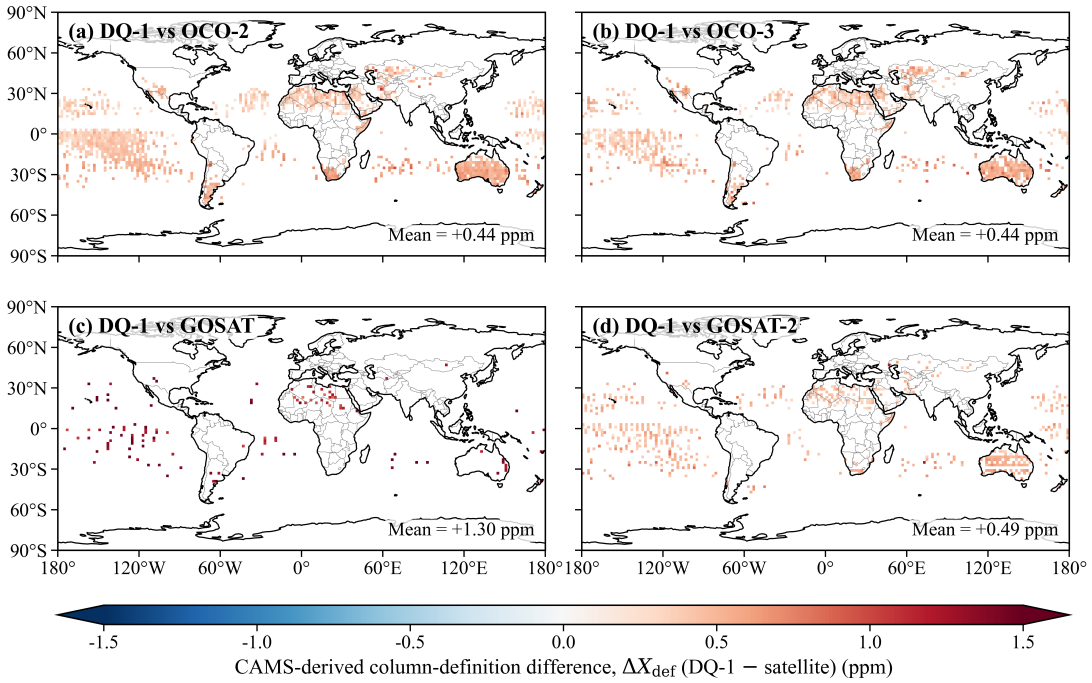


Figure S1. Spatial distributions of the CAMS-derived XCO_2 column-definition differences between DQ-1 and (a) OCO-2, (b) OCO-3, (c) GOSAT, and (d) GOSAT-2 for the collocated samples. The column-definition difference is defined as XCO_2 (DQ-1 definition) − XCO_2 (satellite definition), and the mean value 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		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)
	GOSAT-2 (corrected)	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 Main 5

The manuscript reports that GOSAT-2 is systematically higher than the other satellite products and TCCON and attributes this mainly to the absence of empirical bias correction. However, differences in the bias-correction status of the products may confound differences related to sensors and retrieval algorithms with those caused by product post-processing.

The official NIES release notes for the GOSAT-2 SWIR L2 Full-Physics V02.20 product provide land- and ocean-specific empirical bias-correction formulas and coefficients based on TCCON GGG2020 data. The authors should apply the corresponding correction to the pixel-level GOSAT-2 XCO_2 retrievals and update the inter-satellite and TCCON comparisons. At minimum, results before and after empirical correction should be presented as a sensitivity analysis.

The product version used during the study period and the order in which empirical bias correction and column-definition correction are applied should also be stated explicitly.

ANSWER: We have applied the official NIES empirical bias correction to the individual quality-controlled GOSAT-2 retrievals and updated both the inter-satellite and TCCON comparisons. GOSAT-2 V02.20 is used through February 2024 and V02.21 from March 2024 onward. The version-specific land/ocean empirical correction is applied before daily 2° aggregation and collocation, and the CAMS-based column-

definition correction is applied afterward to the collocated samples. Table 1 and Sect. 2.1.2 now specify both the versions and this processing order.

For the same 880 GOSAT-2–TCCON collocations, empirical correction changes R from 0.91 to 0.92, mean bias from +3.56 to +0.01 ppm, and RMSE from 3.98 to 1.54 ppm. Corrected GOSAT-2 is used in the principal comparisons, while Fig. 3d and Tables S2–S3 retain original and corrected results. We also state explicitly 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 limitation is now stated in Sect. 4.2. This qualification avoids conflating the value of TCCON as a common validation reference with statistical independence for the corrected GOSAT-2 product.

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. 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."

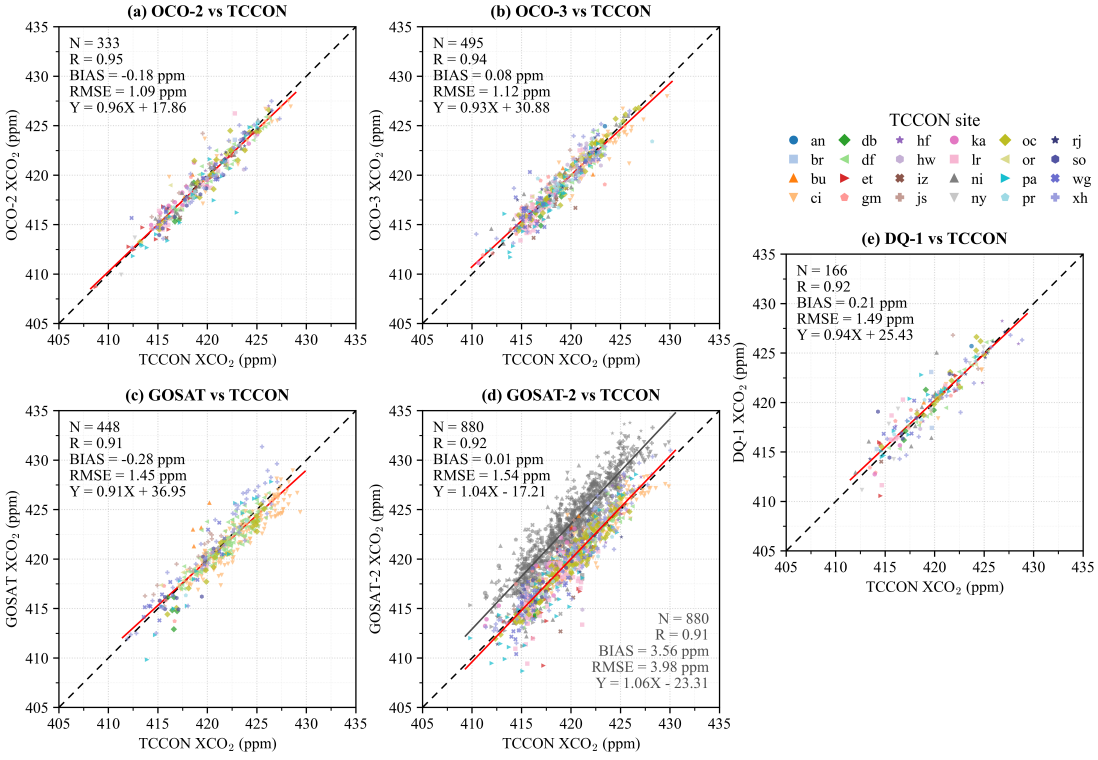


Figure 3. Comparison of satellite XCO_2 products with TCCON observations. (a–e) OCO-2, OCO-3, GOSAT, GOSAT-2, and DQ-1 XCO_2 compared with collocated TCCON XCO_2, respectively. Different colors and symbols denote individual TCCON sites. In panel (d), colored and gray symbols denote empirically bias-corrected 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

Minor comments

Comment Minor 1

Notation throughout the manuscript: Please check the entire manuscript, including the title, Abstract, main text, equations, tables, figure captions, and axis labels, and consistently replace "CO_2" and "XCO_2" containing the Unicode subscript character "₂" with "CO_2" and "XCO_2" produced using standard subscript formatting.

ANSWER: Revised. CO_2 and XCO_2 have been standardized using subscript formatting in the manuscript text and equations.

Comment Minor 2

P3, L66: The word “characteristics” is repeated in the phrase “global XCO_2 observation characteristics characteristics”. Please remove the duplicated word.

ANSWER: Revised. The duplicated “characteristics” has been removed.

Comment Minor 3

Table 1: The column heading “Native resolution/scale” combines temporal resolution, spatial resolution, and sampling mode, which are not directly comparable across the listed datasets. Please either separate this information into “Temporal resolution” and “Spatial resolution/sampling” columns or revise the heading and entries so that the dataset characteristics are described consistently.

ANSWER: Revised. Table 1 now separates “Temporal resolution” from “Spatial resolution/sampling”, and the entries have been revised accordingly.

Comment Minor 4

(1)–(16): Please format all displayed equations consistently following the ACP/Copernicus manuscript template. The equations should be left-aligned, with the equation numbers aligned on the right and numbered consecutively in order of appearance.

ANSWER: Revised. The displayed equations have been reformatted with left-aligned expressions, right-aligned equation numbers, and consecutive numbering.

Comment Minor 5

P7, L180 and L189: “Where” should be changed to lowercase “where”, as the text following the equation continues the preceding sentence. In addition, please remove the extra space before the comma after “d” on L180.

ANSWER: Revised. “Where” has been changed to “where”, and the extra space before the comma has been removed.

Comment Minor 6

P8, Eq. (6): The text following the equation states that represents “the overall sampling sufficiency of grid cell ”. The latter expression should be replaced by “grid cell ”.

ANSWER: Revised. The text following Eq. (6) now consistently refers to grid cell g.

Comment Minor 7

The colorbar in Figure 1 needs to be adjusted. The present colorbar cannot clearly distinguish regions with no valid observations from those where the number of valid observations is lower than 20. The color corresponding to zero valid observations could be changed to white or another color clearly distinct from dark blue.

ANSWER: Revised. The original Fig. 1 is now Fig. A1. Its color scheme uses white for zero observations and colored shading for positive counts, improving the distinction between absent and sparse observations.

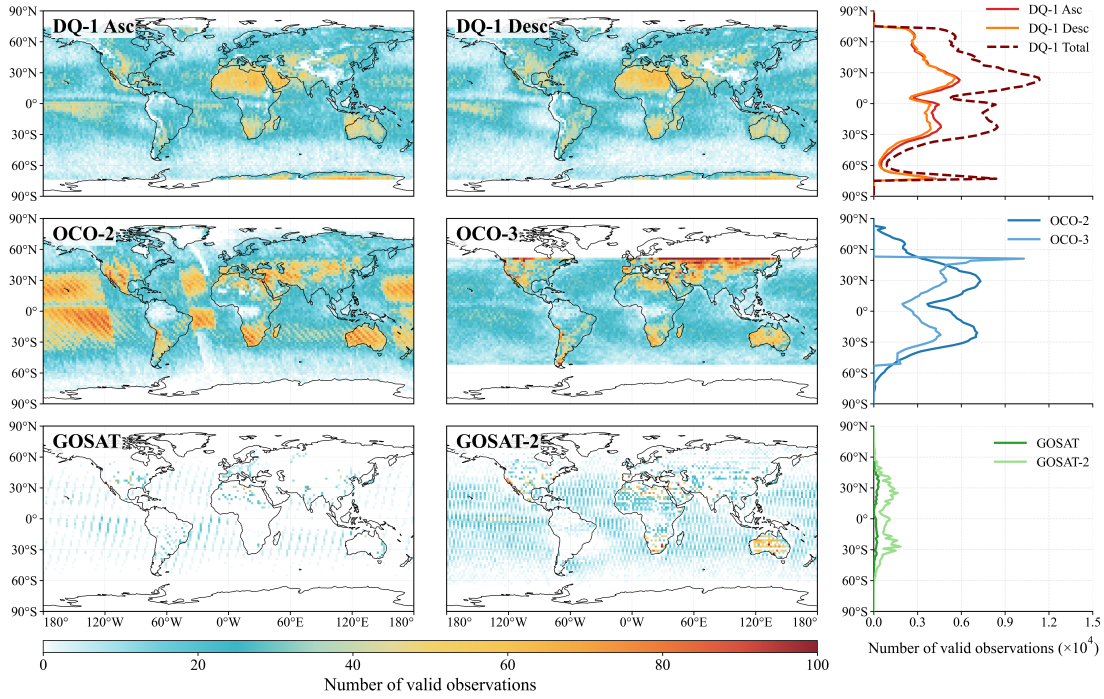


Figure A1. Spatial distributions and zonal totals of cumulative number of valid observations for the satellite XCO_2 products at 2° × 2° resolution.

Comment Minor 8

Figure 2 and the associated text: The term "valid-observation cell-days" does not accurately describe the reported quantity and should be replaced with "number of valid observations". Please define the counting unit explicitly at its first occurrence and revise the terminology consistently throughout the relevant methods, results, figure captions, and axis labels.

The manuscript should also state explicitly that DQ-1 Asc+Desc is calculated as the direct sum of the ascending and descending observation counts.

ANSWER: Revised. We use "number of valid observations" and define the counting unit in Sect. 2.2.1: one sensor, one day, and one 2° grid cell count as one valid observation when the accumulated along-track sampling length reaches at least half of the local grid-cell diagonal. It is therefore a gridded availability count, not the number of raw soundings. DQ-1 Asc and Desc are counted separately, and Total (Asc+Desc) is

their direct sum. The terminology has been updated in the sampling figures and related text.

Changes in the manuscript :

Sect. 2.2.1, p. 9, lines 240–250.

Revised text: "To ensure stable multi-satellite observational coverage in the candidate 2° grid cells, we construct a satellite sampling availability score. For each satellite, daily XCO_2 observations after quality control are first assigned to the same 2° grid according to the latitude and longitude of the observation center. Following the grid-based spatial-coverage method of Han et al. (2024) and the path-length-based definition of valid gridded observations described by Han et al. (2025), we do not use the raw number of observations as the only criterion for determining valid gridded observations. Instead, for each sensor, day, and grid cell, we calculate the along-track sampling length by accumulating the great-circle distances between adjacent valid observation centers along the same orbit track. For each sensor and day, only grid cells with an accumulated sampling length greater than or equal to half of the diagonal length of the local 2° grid cell are counted as one valid observation. DQ-1 Asc and Desc are counted separately, and DQ-1 Total (Asc+Desc) is calculated as the direct sum of the two counts. The number of valid observations is then accumulated over the study period and normalized to the range 0–1 using the 95th percentile among grid cells with positive values, yielding the satellite sampling availability score."

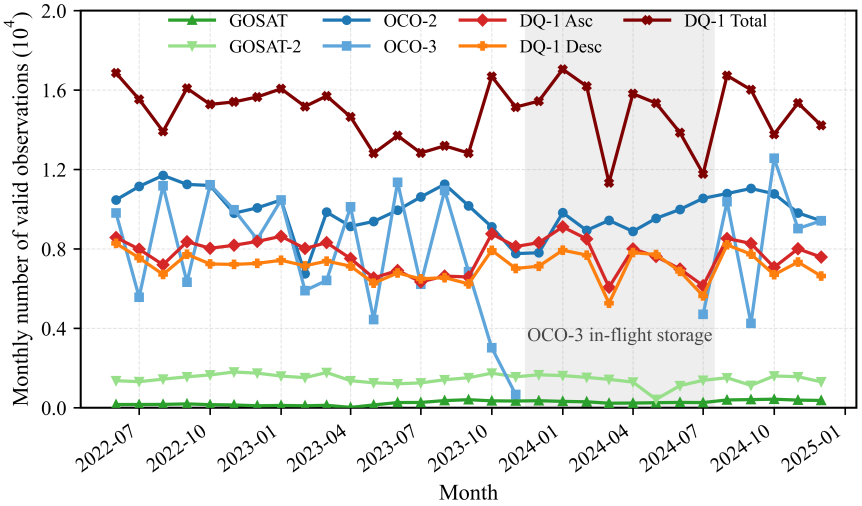


Figure A2. Monthly variations in the number of valid observations for the satellite XCO_2 products. The gray-shaded region indicates the OCO-3 data gap associated with its temporary storage on the ISS.

Comment Minor 9

P14, L366: “Final score” appears within a sentence and should be changed to lowercase “final score” for consistency with sentence-style capitalization.

ANSWER: Revised. “Final score” within the sentence has been changed to “final score”.

Comment Minor 10

Figure 4 and Table A2: Please clearly specify the sign convention used for the mean bias or mean difference, for example, “the product shown on the y axis minus the product shown on the x axis”. The regression method used to derive the slope and intercept should also be specified in the figure caption or table note.

ANSWER: Revised. Bias is explicitly defined as the y-axis product minus the x-axis product in Sect. 2.2.2, the caption of Fig. 4, and Table S3. The revised Fig. 4 and Table S3 report R, N, bias, and RMSE rather than regression slopes and intercepts, so the former inter-satellite regression-parameter table is no longer retained. The fits displayed in the TCCON comparison, now Fig. 3, are identified in its caption as linear least-squares fits.

Comment Minor 11

Table A1: The caption refers to the “valid observation periods” of the TCCON sites, but the table does not appear to include a corresponding observation-period column. Please add this information or revise the caption accordingly. The column headings could also be standardized as “Latitude (°N)”, “Longitude (°E)”, and “Altitude (m)”.

ANSWER: Revised. The site-information table is now Table A2 and includes a “Valid observation period” column. Its note defines the period as the overlap of site data availability with 1 June 2022–31 December 2024. The coordinate and altitude headings are “Latitude (°N)”, “Longitude (°E)”, and “Altitude (m)”.

Table A2. Abbreviated names, geographic information, and valid observation periods of the TCCON sites used for validation. The valid observation period denotes the overlap between data availability at each site and the study period (1 June 2022–31 December 2024).

Abbreviated name	TCCON site	Latitude (°N)	Longitude (°E)	Altitude (m)	Valid observation period
an	Anmyeondo [anmeyondo01]	36.54	126.33	20	2023-02-09–2024-03-03
br	Bremen [bremen01]	53.10	8.85	30	2022-06-01–2024-12-31
bu	Burgos [burgos01]	18.53	120.65	40	2022-06-01–2024-12-31
ci	Caltech [pasadena01]	34.14	−118.13	240	2022-06-01–2024-12-31
db	Darwin [darwin01]	-12.46	130.93	40	2022-06-01–2024-12-31
df	Edwards [edwards01]	34.96	−117.88	700	2022-06-01–2024-12-31
et	East Trout Lake [easttroutlake01]	54.35	−104.99	502	2022-06-01–2024-12-31
gm	Garmisch [garmisch01]	47.48	11.06	743	2022-06-01–2024-01-16
hf	Hefei [hefei01]	31.91	117.17	40	2022-06-01–2024-12-31
hw	Harwell [harwell01]	51.57	−1.32	142	2022-06-01–2024-12-31
iz	Izana [izana01]	28.31	−16.50	2370	2022-06-01–2023-08-30
js	Saga [saga01]	33.24	130.29	9	2022-06-01–2023-12-13
ka	Karlsruhe [karlsruhe01]	49.10	8.44	110	2022-06-01–2024-12-27
lr	Lauder [lauder03]	-45.04	169.68	370	2022-06-01–2024-12-31
ni	Nicosia [nicosia01]	35.14	33.38	185	2022-06-01–2024-12-31
ny	Ny-Ålesund [nyalesund01]	78.92	11.92	20	2022-06-01–2024-12-31
oc	Lamont [lamont01]	36.60	−97.49	320	2022-06-01–2024-12-31
or	Orléans [orleans01]	47.96	2.11	130	2022-06-01–2024-12-31
pa	Park Falls [parkfalls01]	45.94	−90.27	476	2022-06-01–2024-12-31
pr	Paris [paris01]	48.85	2.36	60	2022-06-01–2024-12-31
rj	Rikubetsu [rikubetsu01]	43.46	143.77	380	2022-06-01–2024-12-31
so	Sodankylä [sodankyla01]	67.37	26.63	188	2022-06-01–2024-12-31
wg	Wollongong [wollongong01]	-34.41	150.88	30	2022-06-01–2024-12-31

Abbreviated name	TCCON site	Latitude (°N)	Longitude (°E)	Altitude (m)	Valid observation period
xh	Xianghe [xianghe01]	39.80	116.96	40	2022-06-01–2024-12-31

Comment Minor 12

The manuscript uses TCCON data, but the corresponding data providers are not acknowledged. Please add an appropriate acknowledgement and include the relevant dataset citations or DOIs for the TCCON sites listed in Table A1.

ANSWER: Revised. We have acknowledged the TCCON investigators and data providers, added a site-to-reference mapping beneath Table A2, and included the corresponding GGG2020 dataset citations and DOIs in the reference list.

Comment Minor 13

Author contributions: "Author contributions." appears twice at the beginning of this section. Please remove the duplicated text.

ANSWER: Revised. The duplicated "Author contributions" text has been removed.

Comment Minor 14

References: The bibliographic information for Han et al. (2017, 2018) appears incomplete. Please provide the journal name, volume, page range or article number, and DOI, following the same format used for the other references.

ANSWER: Revised. The bibliographic entries for Han et al. (2017, 2018) have been completed and standardized to include the journal information, volume, page/article information, and DOI in the same reference style as the rest of the manuscript.