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."

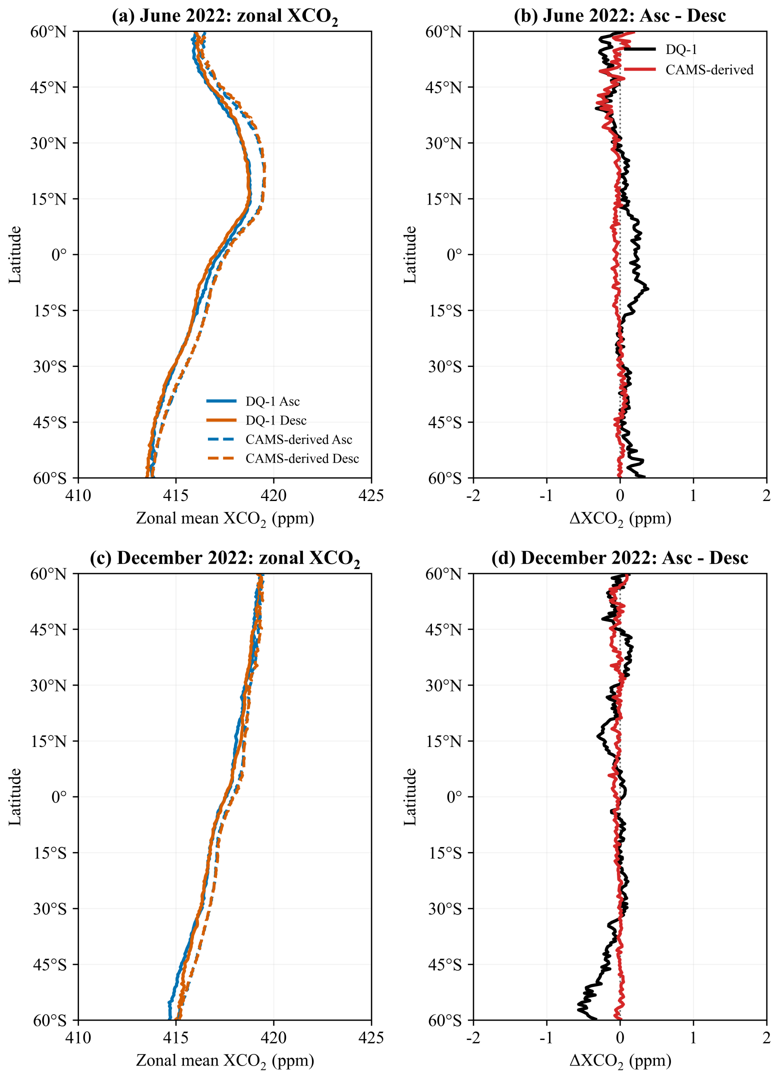

Figure S8. Zonal-mean comparison of ascending (daytime) and descending (nighttime) DQ-1 XCO2 and the corresponding CAMS-derived XCO2 under the DQ-1 column definition for (a, b) June 2022 and (c, d) December 2022. Panels (a) and (c) show the zonal-mean XCO2 distributions for the Asc and Desc observations,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

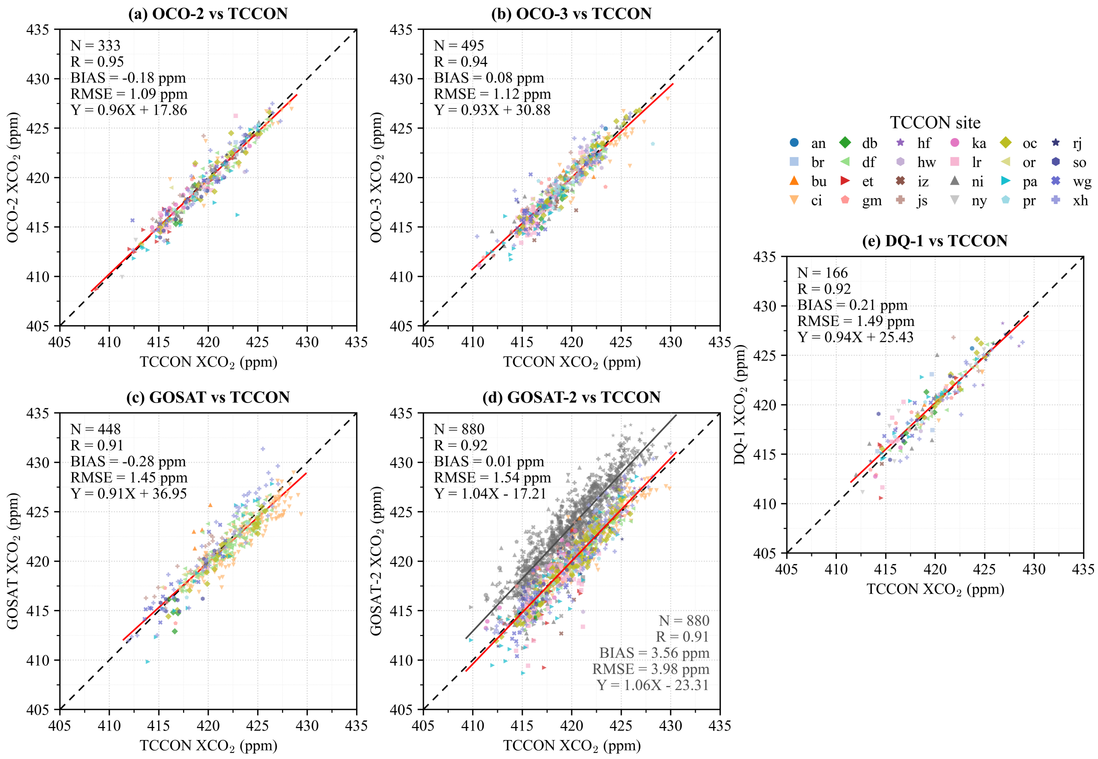

Figure 3. Comparison of satellite XCO_2 products with TCCON observations. (a–e) OCO-2, OCO-3, GOSAT, GOSAT-2, and DQ-1 XCO_2 compared with collocated TCCON XCO_2, respectively. Different colors and symbols denote individual TCCON sites. In panel (d), colored and gray symbols denote empirically bias-corrected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

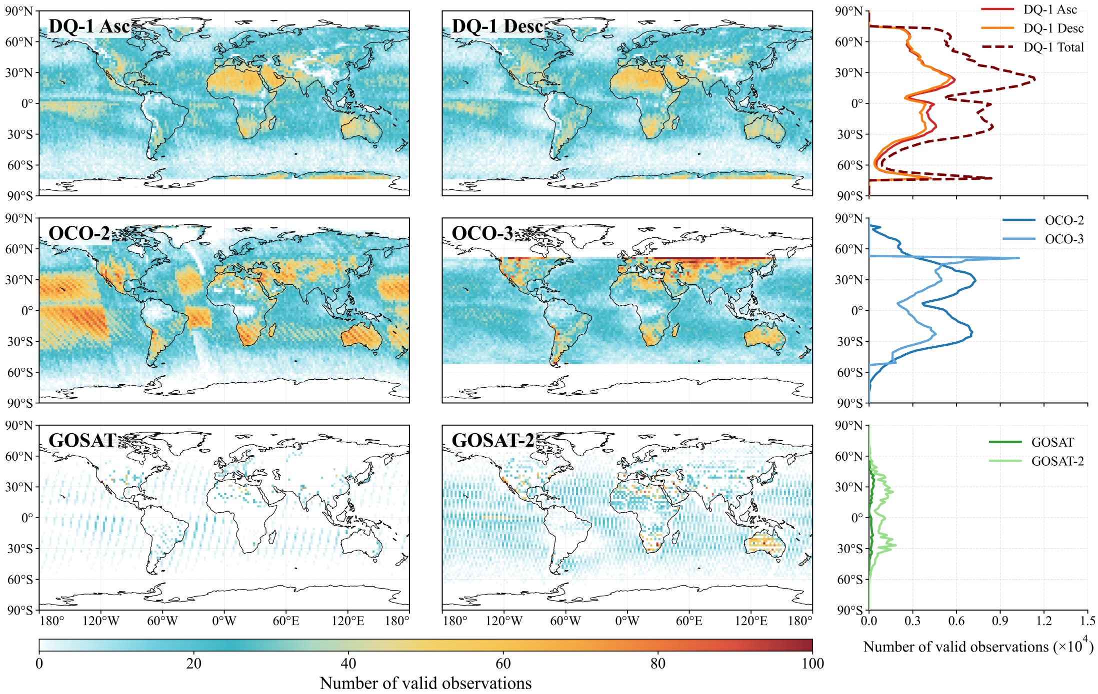

Figure A1. Spatial distributions and zonal totals of cumulative number of valid observations for the satellite XCO_2 products at 2° × 2° resolution.

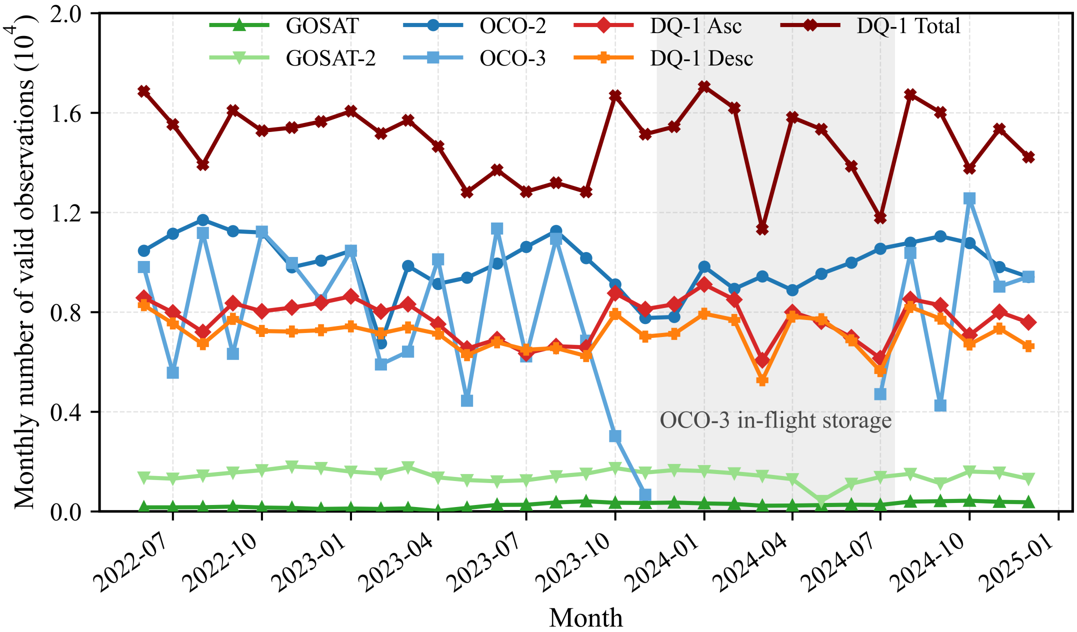

Figure A2. Monthly variations in the number of valid observations for the satellite XCO_2 products. The gray-shaded region indicates the OCO-3 data gap associated with its temporary storage on the ISS.

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."

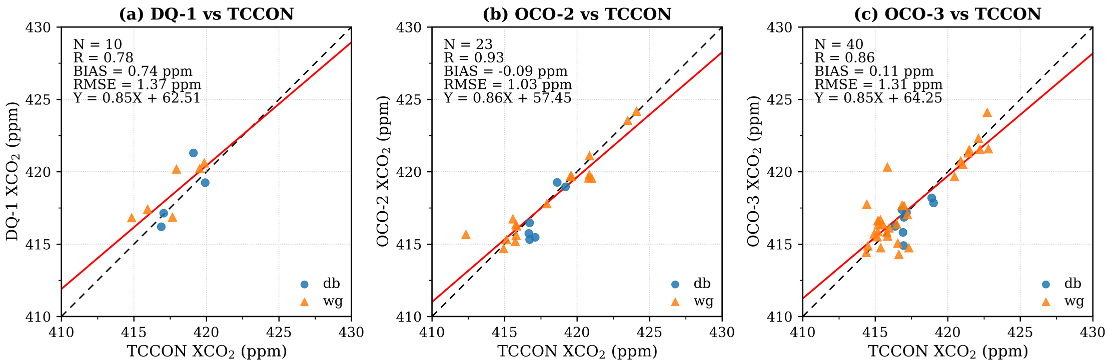

Figure S5. Scatterplot comparisons of DQ-1, OCO-2, and OCO-3 XCO_2 with TCCON observations at the Darwin (db) and Wollongong (wg) sites in Australia. Panels (a–c) show the comparisons for DQ-1, OCO-2, and OCO-3, respectively. The dashed black line denotes the 1:1 line, and the red line shows the linear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

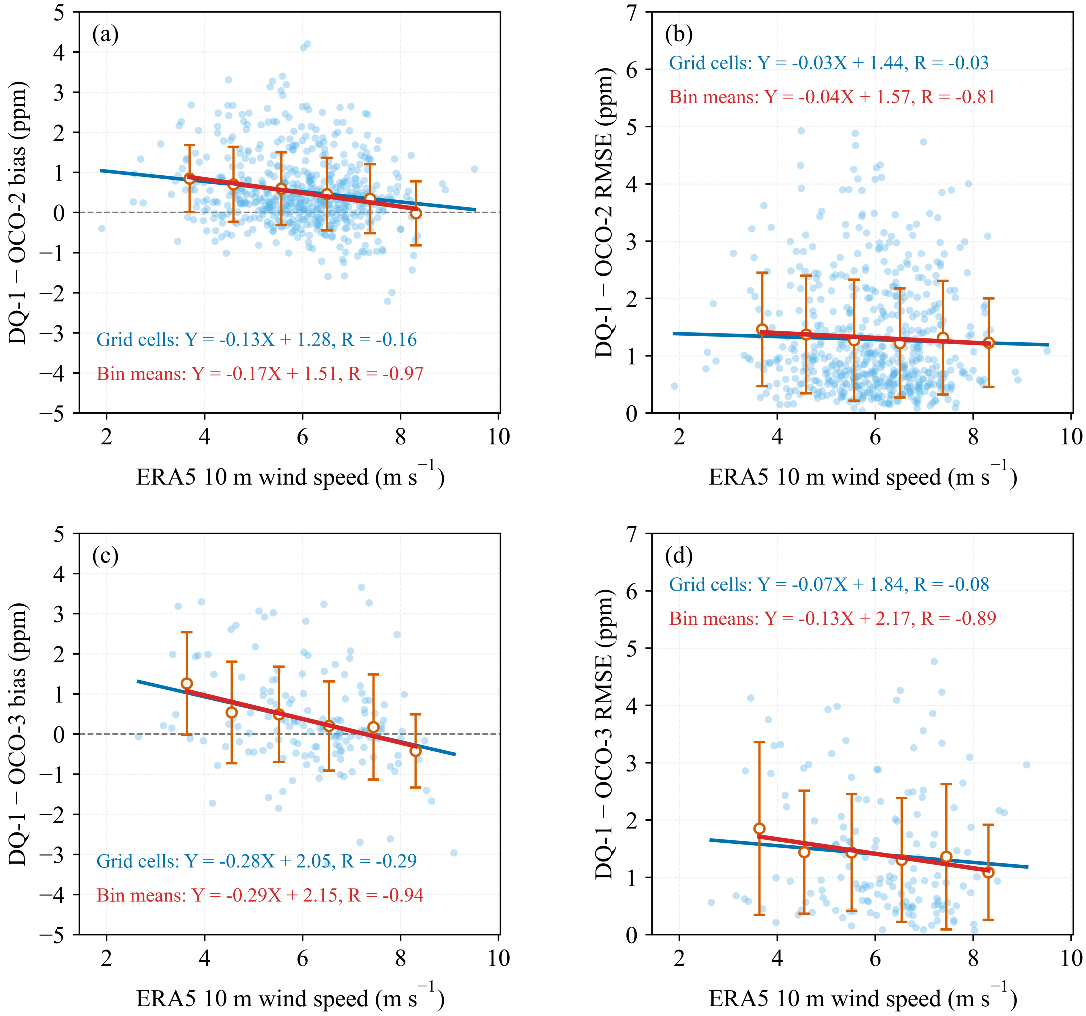

Figure S6. Dependence of oceanic DQ-1–OCO-2 and DQ-1–OCO-3 bias and RMSE on ERA5 10 m wind speed. Panels (a, b) show DQ-1–OCO-2 and panels (c, d) show DQ-1–OCO-3. Light blue dots denote individual oceanic 2° grid cells, while orange circles and error bars show 1 m s−1 wind-speed-bin means ±1 standard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