Response to Reviewer #2
Evaluating the consistency between OCO-2 and OCO-3 XCO_2 estimates derived from the NASA ACOS version 10 retrieval algorithm

T.E. Taylor et al.

Thank you to the reviewer for the suggestions and comments. We appreciate your time and concern. We have addressed each enumerated point below. The original reviewer comment is given in black. Our reply is given in blue. Modifications to the manuscript text are given in red.

1. The use of TCCON as truth proxy for retrieval evaluation does appear to be circular, as indicated by the authors. Have some other data sources (e.g., from airborne campaigns) been considered as truth proxies?

   This is an important field of study that we did not address directly in the submitted version of the manuscript. In this work we did not evaluate OCO XCO_2 against any additional data sources beyond the three truth metrics used in the training of the quality filtering and bias correction. However, in the revised manuscript we have added explicit wording to the discussion in Section 6 as follows:

   This section discusses the evaluation of the OCO v10 good quality-flagged XCO_2 estimates against the truth-proxies used in the quality filtering and bias correction procedure. Although there is some circularity in evaluating the satellite data against the same truth-proxies used for filtering and bias correction, we argue that the multi-parameter parametric bias correction is general enough so as to not over-fit the OCO data. Furthermore, the truth-proxies used for evaluation have been extended in time compared to the data sets used to train the filtering and bias correction. Although it is outside the scope of the current work, OCO-2 data have been validated against a range of other datasets, including in-situ, and NOAA and Aircore vertical observations (Rastogi et al., 2021), aircraft campaigns, e.g. ATom (Kulawik et al., 2019) and ACT-America (Bell et al., 2020), ship-borne and airborne measurements (Müller et al., 2021), and EM27/SUN measurements (Jacobs et al., 2020).

   Section 3.3.1 introduced and discussed the TCCON data as used in the OCO v10 quality filtering and bias correction. Both OCO-2 and OCO-3 were quality filtered and bias corrected against TCCON GGG2014 data, while here we compare to TCCON GGG2020 XCO_2 estimates. Key changes to the retrieval algorithm between GGG2014 and GGG2020 are available on the TCCON wiki page (TCCON). The OCO quality-filtering and bias corrections were trained using data through December 2018 for OCO-2 and December 2020 for OCO-3, whereas the validation data extends through

February 2022 for OCO-2 and August 2022 for OCO-3. This provides some degree of independence in the evaluation.

Regarding the detailed evaluation against TCCON, we note that a detailed and rigorous analysis is underway by members of the OCO project, and a manuscript for submission to peer-review is forthcoming. Unfortunately, we could not cite this work as it has not yet been submitted.

2. The authors discuss the updates from ACOS v8/v9 to v10, and I'm wondering what are the impacts of these updates in terms of retrieval quality? Can the authors compare, for example, the RMSE of ACOS v8 vs. TCCON with that of ACOS v10 vs. TCCON?

   A new plot (new Fig. 1) with descriptive text has been added to the paper to show the RMSE of ACOS v9 and v10 against TCCON, models, and small areas.

   Improvements in successive versions of the ACOS L2FP retrieval are demonstrated in Fig. 1, which compares RMS errors in XCO_2 from v9 and v10 OCO-2 and v10 OCO-3 versus the three truth-proxies for land and ocean-glint observations. There are substantial decreases in the RMSE for OCO-2 from v9 to v10 as compared to both TCCON and to the MMM for both land and ocean-glint. The changes in the OCO-2 RMSE from v9 to v10 for the small area analysis were insignificant between versions, which is to be expected because errors at very small spatial scales are primarily driven by instrument noise, which cannot be further reduced. For all three truth metrics versus land observations, OCO-3 compares worse than OCO-2. We hypothesize that this is driven by both residual pointing errors and L1b calibration errors, both of which are expected to improve in the next data version. The worse agreement of OCO-3 v10 with TCCON as compared to OCO-2 v10 can be explained in part by the limited number of TCCON collocations with OCO-3 that were available at the time of creation of the QTS. Incidentally, the data also demonstrate that for all truth proxies and for both sensors, ocean-glint errors are lower than land errors, indicating higher precision relative to land observations. This result is at odds with previous findings showing unrealistic features in global inversion models which assimilate OCO-2 ocean-glint data (e.g. Peiro et al., 2022; Byrne et al., 2023).

3. Line 9: how do you interpret the differences between OCO-2 and OCO-3 in the context of precision/bias estimates against truth proxies? Do the OCO-2/-3 differences reflect or represent systematic errors or certain components of the systematic errors?

   It is difficult to infer driving mechanisms for differences between OCO-2 and OCO-3 vs TCCON as reported here. The collocations cover somewhat different time ranges, use somewhat different sites, etc. As noted above, the detailed evaluation against TCCON is reserved for a forthcoming manuscript.

4. Line 100: spatial scale for precision and accuracy requirements?

   The discussion has been updated as follows:

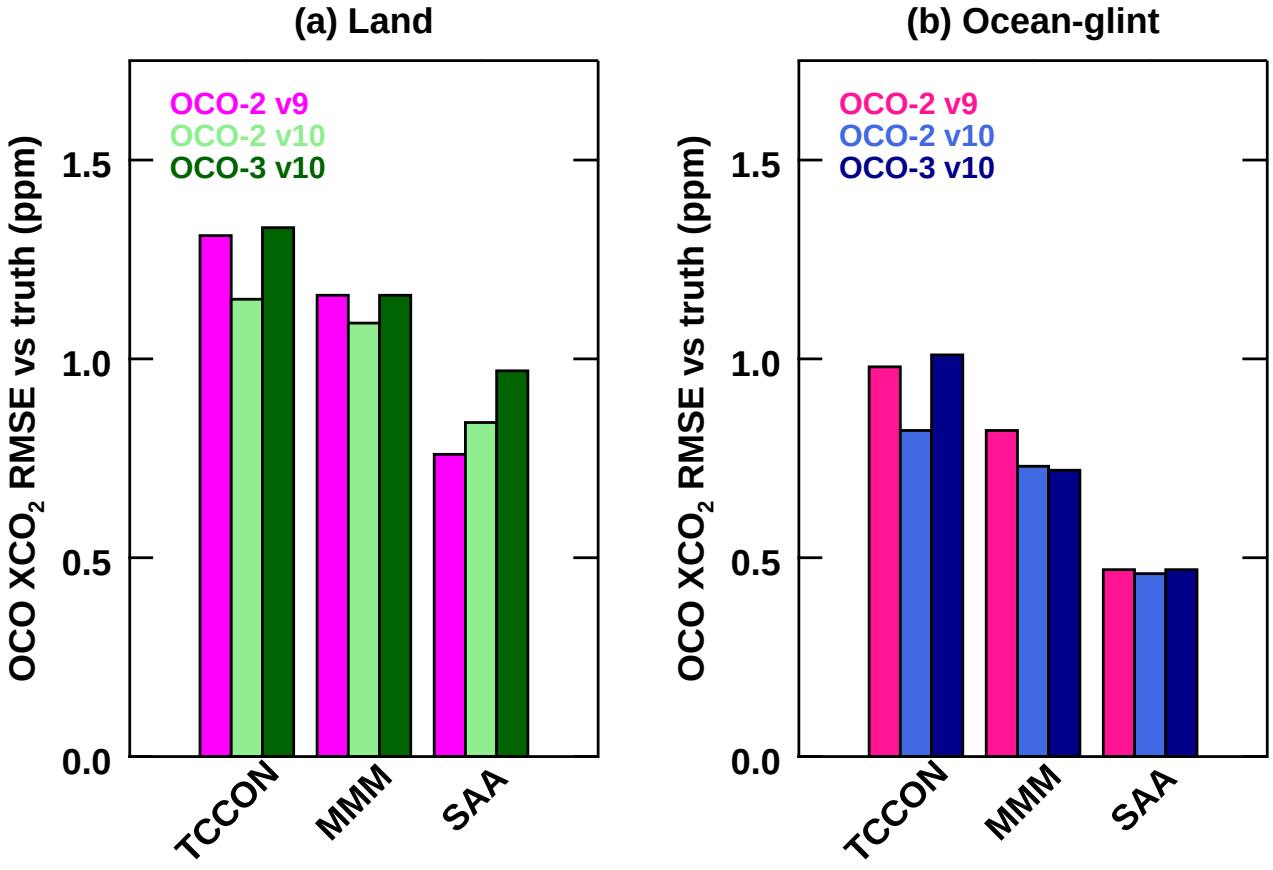

Figure 1. RMS errors for different versions of OCO-2 and OCO-3 XCO_2 versus three truth proxies: TCCON, multi-model-median (MMM), and small area approximations (SAA). Results are derived from single sounding statistics using quality-filtered and bias corrected XCO_2 from the QTS. Results for land observations are shown in panel (a), while panel (b) shows results for ocean-glint observations.

The precision and accuracy requirements for OCO-2 and OCO-3 were originally applied to regional scales, roughly defined as 10° latitude by 10° longitude. Early Observation System Simulation Experiments (OSSEs) indicated that an XCO_2 precision and accuracy better than 1 ppm (less than 0.25%) is needed at this scale to constrain typical natural and anthropogenic sources and sinks of CO_2 (Miller et al., 2007). In practice, the spatial scale for precision and accuracy requirements is determined by the distribution of the validation reference measurements. This is defined by the approximately two dozen TCCON stations and a comparable number of EM27/Sun and Aircore stations distributed over the globe. The system performance on finer scales has also been assessed through comparisons with data collected by aircraft campaigns, e.g. ACT-America (Bell et al., 2020) and ATom (Kulawik et al., 2019), and multi-instrument EM27/SUN Campaigns (Rißmann et al., 2022).

5. Line 125: version 10 here refers to the L1B algorithm? This is a bit confusing. Is this part of level 1 or level 2?

   Here we refer to the v10 L1b calibration. The context has been clarified in the manuscript.

   Keller et al. (2023) describes the current state of the calibration for the L1b OCO-3 v10 data products. OCO-3, unlike OCO-2, cannot view the sun from the ISS, making solar calibration impossible (Rosenberg et al, 2020). Therefore, as compared to OCO-2, more emphasis has been placed on the internal lamp calibration system, which is comprised of three tungsten halogen lamps and a reflective diffuser. The three calibration lamps are illuminated with different cadences and thus age at different rates. For v10 L1b, an algorithm was developed to use information from all three lamps with the goal of mitigating lamp aging while still allowing changes in instrument response to be tracked with the necessary temporal resolution. This is particularly important for OCO-3, as it has exhibited significant, abrupt changes in its overall instrument response. In addition, an update was made to the OCO-3 stray light model used for v10 L1b to account for spatial variability on the detectors.

6. Lines 163-164 - specify the variability of the bias in the two versions?

   For a global test set used during v10 development, the variability in retrieved minus prior surface pressure was reduced from 3.3 hPa to 2.8 hPa.

   This update yielded reduced spatial variability of the bias between the L2FP retrieved surface pressure and the prior value from 3.3 hPa in v9 to 2.8 hPa in v10.

7. Table 2: it appears that Table 2 is not mentioned in text.

   This was an omission. A reference to Table 2 has been added.

   The sounding selection strategy, which determines if a sounding should be run through the L2FP, remains roughly consistent for v10 compared to previous versions. Details are provided in Table 2.

8. Line 252: how is initial median calculated (from the six models in Table 4)? Line 253 mentions “four models”.

As written, this was a bit confusing. There are a number of models employed in both the QF/BC and the actual XCO_2 evaluation. But it varies by sensor. The paragraph was modified as follows.

To augment the sparse spatial coverage of TCCON, we use global atmospheric inversion models in the OCO XCO_2 quality filtering and bias correction process to provide full global coverage (O'Dell, 2018). For ACOS v10, the median XCO_2 was derived from the 4-dimensional (4D) CO_2 fields of models that assimilated only in-situ CO_2 data. To ensure consistency in the models, for each OCO sounding, only the models with XCO_2 that deviated by less than ± 1.5 ppm from the initial median value were retained. Furthermore, soundings were excluded if more than one of the models had been rejected, or if the standard deviation amongst the valid models was >1 ppm. Tables 4, 5 and 6 provide information about the suite of models. An asterisk is used in Table 6 to identify the specific models and data versions used in the quality filtering and bias correction procedure: three for OCO-2 and four for OCO-3 v10. Some of the same models were also used in the XCO_2 evaluation, but using a different model data version and a different evaluation period. A few of the models were used only for the XCO_2 product evaluation.

9. Line 259: how is model/TCCON profile converted to 20 layers of ACOS profiles?

   A simple linear interpolation is performed from the native resolution to the 20 ACOS levels. The paragraph has been modified as:

$$\mathrm{XCO}_{2,ak} = \sum_{i=1}^{20} \mathbf{h}_i \{\mathbf{a}_i \mathbf{u}_{m,i} + (1 - \mathbf{a}_i)\mathbf{u}_{a,i}\}. \quad (1)$$

   Here, $\mathbf{h}_i$ is the pressure weighting function on the $i = 1...20$ ACOS model levels, defined as the pressure intervals assigned to the state vector normalized by the surface pressure and corrected for the presence of atmospheric water vapor. See Appendix A of O'Dell et al., 2012 for details. The vector $\mathbf{a}$ is the CO_2 column averaging kernel, which relates the sensitivity of the retrieved CO_2 to the true atmospheric state of CO_2 at each vertical level, as described in Connor (2008). The vector $\mathbf{u}_m$ is the retrieved TCCON or model profile of CO_2, linearly interpolated from the native vertical resolution to the 20 ACOS levels. The vector $\mathbf{u}_a$ is the ACOS prior profile of CO_2. Generally the averaging kernel corrections are on the order of 0.5 ppm or less.

10. Table 7: what is considered truth xCO2 for SAA?

    The following explanatory sentence was added in Sect. 3.3.1 when the truth proxies are introduced:

    A median value of each small area provides a truth proxy to which each sounding in the small area can be compared.

11. Line 290: course should be coarse?.

    Corrected.

12. Line 290: logarithm is taken for AODfine as well?.

    No, AODfine is in linear space.

13. Line 322: FIn should be In..

    Corrected.

14. Figure 3: Y axis spacing is not linear?.

    This is true. The spacing is scaled by the cosine of latitude. The following explanation was added to the figure caption and to the other similar style figures in the document.

    The ordinate axis is scaled by the cosine of the latitude to elucidate the decreasing fractional surface area of the earth with increasing latitude.

15. Section 4 – I'm wondering if Section 4 can be moved to the Appendix?

    In theory, yes, but we feel that retaining this in the main document is important since the differences in data volumes between the two sensors is likely of interest to data users.

16. Figure 4 – Can you show the correlation between OCO-2 and OCO-3?

    I'm not sure what the reviewer means. The correlation in the gridded values could be shown on a one:to:one scatter plot for grid cells in which both sensors have values, but I'm not sure how informative it would be. Since the bias correction itself is formulated from the delta surface pressure (retrieved minus prior), an aerosol term, and the lapse rate of the CO_2 profile (CO_2 grad del), any correlation would indicate correlations between the sensors in those variables to a certain extent. It is difficult/impossible to directly correlate any variable between OCO-2 and OCO-3 due to the vastly different sampling strategies. It can only be done for the collocated soundings, and then only as a mean. Perhaps the reviewer had a different meaning, but we are unable to interpret.

17. Figure 8 and Section 6.1 – what is the typical variability of individual soundings during each overpass? Do the comparisons change if the size of the domain is changed?.

    The original version of Fig. 8 did not capture this variable. The updated Fig. 8 (now Fig. 9) replaced the not-very-useful "standard deviation of the mean XCO_2 of the overpasses", with the more meaningful "mean of the standard deviation of the XCO_2 of the overpasses". Based on this new statistic, we see that the variability in TCCON is about 0.5 ppm per overpass (in the mean), while the variability for OCO-2 and OCO-3 is significantly higher for Land (0.84 and 0.99 ppm, respectively) and a little lower for ocean (0.49, 0.43 ppm). Presumably, yes, the variability in XCO_2 for each overpass will decrease slightly as the collocation criteria is tightened, although we did not invest the time to carry out the exercise here. In general, the collocation of OCO to TCCON has been well vetted in the literature, and is always a tradeoff between precision and throughput.

    Also note that panels (c) and (f) in Fig. 9 now include all of the OCO footprints, rather than just a single footprint (out of 8 footprints). This provides many more collocations, thereby producing more robust statistics. The comparison statistics changed slightly but did not alter the main conclusion that OCO-2 and OCO-3 tend to agree with one another as well or better than either agrees with TCCON.

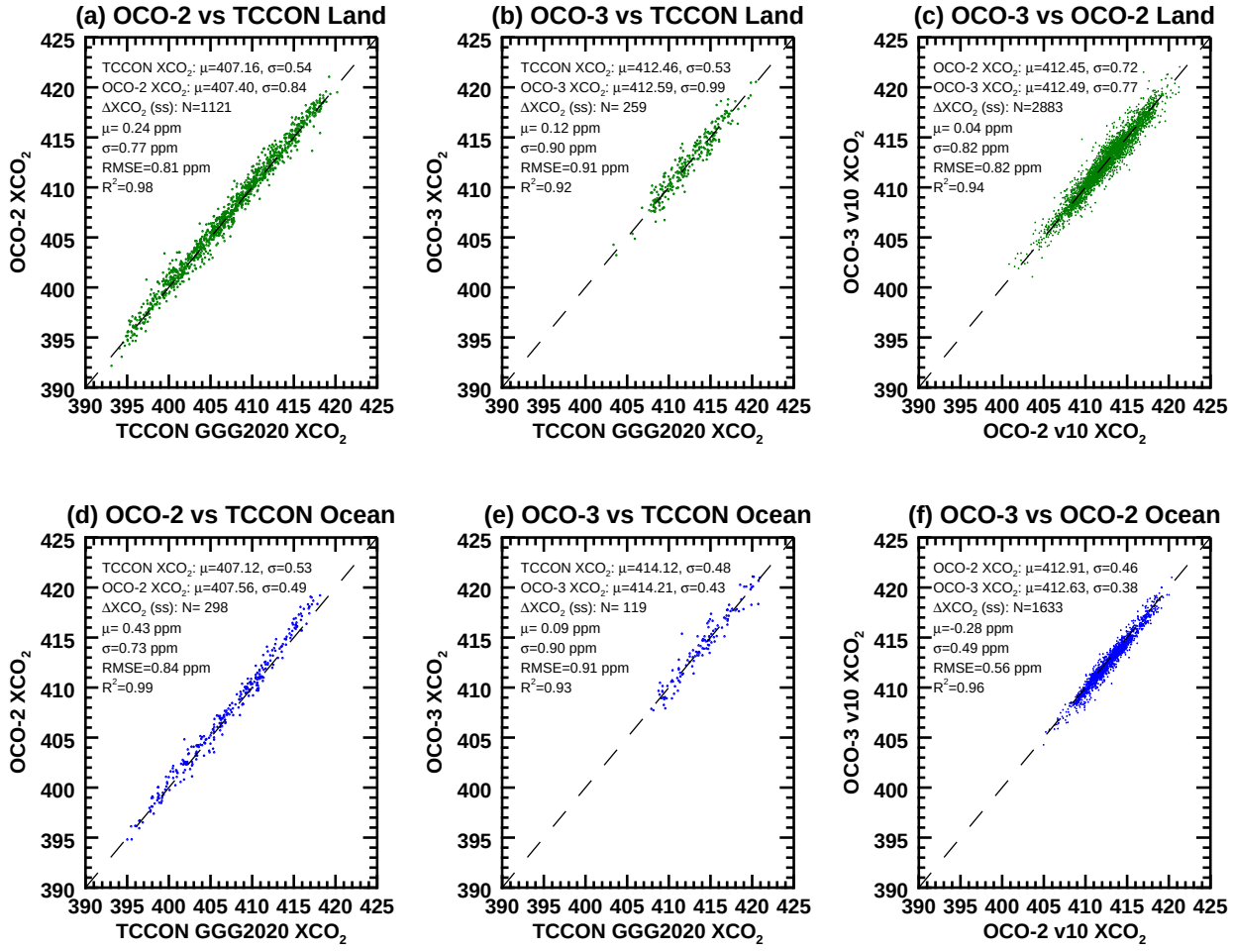

Figure 2. (**Figure 9 in the modified document**) One-to-one XCO_2 correlation plots for land (top row) and ocean (bottom row) observations. Panels (a) and (d) show OCO-2 v10 versus collocated TCCON GGG2020 estimates, while panels (b) and (e) show OCO-3 v10 versus TCCON. Panels (c) and (f) show the correlation in OCO-3 versus OCO-2 XCO_2, respectively, for the set of collocated soundings described in Appendix **??**. In each panel, the top two rows of statistics give the mean XCO_2 for the two sensors being compared for the N collocations, and the mean standard deviation in the XCO_2 (σ). The third through seventh rows of statistics give the number of collocations (N), the mean ΔXCO_2, the standard deviation of the ΔXCO_2, the RMSE ($\sqrt{\mu^2 + \sigma^2}$), and the coefficient of determination (R^2 = the squared Pearson linear correlation coefficient).

18. Line 482: course should be coarse?.

    Corrected.

19. Line 484: Table 4 and Table 5 list six models, not four..

    A total of 6 models were used for both QF/BC and evaluation. Not all of the same models were used for each for the two sensors. I tried to make it as clear as I could in Table 6. I feel like all of the necessary information is there for the reader to interpret, although I agree that it is perhaps slightly confusing. But I cannot think of a different way to pose it. I added a simple reference back to Table 6 in the relevant discussion. And, as given in item 8 above, I tried to make the initial introduction of the models a bit more clear.

    The models chosen for evaluation of the OCO v10 signal, identified in Section 3.3 and Table 6, all fit the following criteria:...

20. Figure 9: the title of right column in the figure is a bit confusing.

    The right column shows the gridded difference of the difference of each sensor from the model. We have replaced the confusing "$\Delta\ \Delta\ XCO_2$" with "Δ signal".

21. Figure 10 and line 514: northward propagation is not that obvious in the plot (especially Figure 10b), perhaps mark it in the figure?.

    The discussion has been slightly modified to read:

    A useful way to investigate the characteristics of the signal is by binning values into zonal (10° latitude) and 10 day bands, as seen in Fig. 10. Here, land observations (includes land-nadir, land-glint, land-TG, and land-SAM) are shown in the left column, and ocean-glint observations on the right, with OCO-2 on the top row and OCO-3 in the middle. The model suite runs only through December of 2020, so the graphs cover a 16 month period, starting in August, 2019. While the zonal mean tends to de-emphasize certain spatial features visible in the global maps, it brings out the temporal variations in the signal. Based on these results, coherent seasonal and latitudinal patterns in the signal are observed for both sensors. For land observations, both sensors tend to have positive signals in the NH and negative signals in the SH, while for ocean-glint observations, the signals tend to be positive poleward of the tropics in both hemispheres, and negative in the tropics. The statistics calculated on the gridded signal data indicate that OCO-3 has higher scatter than OCO-2 (0.62 ppm vs. 0.46 ppm) and a larger bias (-0.30 ppm vs. -0.15 ppm) than OCO-2 for land observations, as seen in panels (a) and (b). The statistics for ocean-glint signals indicate similar scatter between the two sensors of 0.53 ppm and 0.59 ppm for OCO-2 and OCO-3, respectively, with mean biases of 0.24 and -0.23 ppm, as seen in panels (d) and (e).

    The lower two panels of Fig. 10 show the differences in the gridded values between the two sensors for land observations in panel (c) and for ocean-glint observations in panel (f). The gridded mean difference between the two sensors for land observations is -0.08 ppm. The largest differences for land occur in approximately December 2019, immediately

following the OCO-3 PMA calibration that was described in Sec. 2.1.2 of Taylor et al. (2020), and running through approximately January 2020 when the next OCO-3 decontamination cycle occurred. At the time of writing, we have not yet had time to investigate this newly discovered feature in the OCO-3 v10.4 XCO_2 record. The gridded differences for ocean-glint observations, shown in panel (f) indicate a mean low bias of -0.3 ppm for OCO-3 relative to OCO-2. Overall, as was demonstrated with the maps in Fig. 9, these plots suggest that the two sensors tend to agree better with one another than they do with the model suite.

22. Figure 11: describe in the caption what each data point represents in the scatter plots..

The caption Fig. 11 (now Figs. 12) has been updated as follows:

Analysis of XCO_2 uncertainties for land-nadir small areas for OCO-2 (top) and OCO-3 (bottom). Panels (a) and (c) provide the frequency distributions of both the theoretical uncertainties (blue curves) of the retrieved XCO_2 as reported in the L2Lite file product (variable *xco2_uncertainty*) and the actual uncertainties (green curves) calculated from the standard deviation in the XCO_2 for individual small areas. The number of small areas (N) and the mean (μ) and standard deviation (σ) of the theoretical and actual uncertainties are given in the legend. Panels (b) and (d) show the correlation of the actual uncertainties against the theoretical uncertainties, where each black dot is a binned median value. The one:to:one line is shown as a dashed black line.

References

Bell, E., O'Dell, C. W., Davis, K. J., Campbell, J., Browell, E., Scott Denning, A., Dobler, J., Erxleben, W., Fan, T.-F., Kooi, S., Lin, B., Pal, S., and Weir, B.: Evaluation of OCO-2 XCO_2 variability at local and synoptic scales using lidar and in situ observations from the ACT-America campaigns, Journal of Geophysical Research: Atmospheres, 125, https://doi.org/https://doi.org/10.1029/2019JD031400, 2020.

Byrne, B., Baker, D. F., Basu, S., Bertolacci, M., Bowman, K. W., Carroll, D., Chatterjee, A., Chevallier, F., Ciais, P., Cressie, N., Crisp, D., Crowell, S., Deng, F., Deng, Z., Deutscher, N. M., Dubey, M. K., Feng, S., García, O. E., Griffith, D. W. T., Herkommer, B., Hu, L., Jacobson, A. R., Janardanan, R., Jeong, S., Johnson, M. S., Jones, D. B. A., Kivi, R., Liu, J., Liu, Z., Maksyutov, S., Miller, J. B., Miller, S. M., Morino, I., Notholt, J., Oda, T., O'Dell, C. W., Oh, Y.-S., Ohyama, H., Patra, P. K., Peiro, H., Petri, C., Philip, S., Pollard, D. F., Poulter, B., Remaud, M., Schuh, A., Sha, M. K., Shiomi, K., Strong, K., Sweeney, C., Té, Y., Tian, H., Velazco, V. A., Vrekoussis, M., Warneke, T., Worden, J. R., Wunch, D., Yao, Y., Yun, J., Zammit-Mangion, A., and Zeng, N.: National CO_2 budgets (2015–2020) inferred from atmospheric CO_2 observations in support of the global stocktake, Earth System Science Data, 15, 963–1004, https://doi.org/10.5194/essd-15-963-2023, https://essd.copernicus.org/articles/15/963/2023/, 2023.

Jacobs, N., Simpson, W. R., Wunch, D., O'Dell, C. W., Osterman, G. B., Hase, F., Blumenstock, T., Tu, Q., Frey, M., Dubey, M. K., Parker, H. A., Kivi, R., and Heikkinen, P.: Quality controls, bias, and seasonality of CO_2 columns in the boreal forest with Orbiting Carbon Observatory-2, Total Carbon Column Observing Network, and EM27/SUN measurements, Atmospheric Measurement Techniques, 13, 5033–5063, https://doi.org/10.5194/amt-13-5033-2020, https://amt.copernicus.org/articles/13/5033/2020/, 2020.

Kulawik, S. S., O'Dell, C., Nelson, R. R., and Taylor, T. E.: Validation of OCO-2 error analysis using simulated retrievals, Atmospheric Measurement Techniques, 12, 5317–5334, https://doi.org/10.5194/amt-12-5317-2019, 2019.

Miller, C. E., Crisp, D., DeCola, P. L., Olsen, S. C., Randerson, J. T., Michalak, A. M., Alkhaled, A., Rayner, P., Jacob, D. J., Suntharalingam, P., Jones, D. B. A., Denning, A. S., Nicholls, M. E., Doney, S. C., Pawson, S., Boesch, H., Connor, B. J., Fung, I. Y., O'Brien, D., Salawitch, R. J., Sander, S. P., Sen, B., Tans, P., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Yung, Y. L., and Law, R. M.: Precision requirements for space-based X_{CO_2} data, Journal of Geophysical Research: Atmospheres, 112, https://doi.org/10.1029/2006JD007659, 2007.

Müller, A., Tanimoto, H., Sugita, T., Machida, T., Nakaoka, S., Patra, P. K., Laughner, J., and Crisp, D.: New approach to evaluate satellite-derived XCO_2 over oceans by integrating ship and aircraft observations, Atmospheric Chemistry and Physics, 21, 8255–8271, https://doi.org/10.5194/acp-21-8255-2021, https://acp.copernicus.org/articles/21/8255/2021/, 2021.

Peiro, H., Crowell, S., Schuh, A., Baker, D. F., O'Dell, C., Jacobson, A. R., Chevallier, F., Liu, J., Eldering, A., Crisp, D., Deng, F., Weir, B., Basu, S., Johnson, M. S., Philip, S., and Baker, I.: Four years of global carbon cycle observed from the Orbiting Carbon Observatory 2 (OCO-2) version 9 and in situ data and comparison to OCO-2 version 7, Atmospheric Chemistry and Physics, 22, 1097–1130, https://doi.org/10.5194/acp-22-1097-2022, 2022.

Rastogi, B., Miller, J. B., Trudeau, M., Andrews, A. E., Hu, L., Mountain, M., Nehrkorn, T., Baier, B., McKain, K., Mund, J., Guan, K., and Alden, C. B.: Evaluating consistency between total column CO_2 retrievals from OCO-2 and the in situ network over North America: implications for carbon flux estimation, Atmospheric Chemistry and Physics, 21, 14 385–14 401, https://doi.org/10.5194/acp-21-14385-2021, 2021.

Rißmann, M., Chen, J., Osterman, G., Zhao, X., Dietrich, F., Makowski, M., Hase, F., and Kiel, M.: Comparison of OCO-2 target observations to MUCCnet – is it possible to capture urban X_{CO_2} gradients from space?, Atmospheric Measurement Techniques, 15, 6605–6623, https://doi.org/10.5194/amt-15-6605-2022, 2022.

TCCON: wiki page, Tech. rep., https://tccon-wiki.caltech.edu/Main/DataDescriptionGGG2020, "last accessed May 8, 2023".