Response to Reviewer #1
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 feedback. 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. Abstract line 15: "show encouraging results": Encouraging for what? Please modify this sentence if possible, using a more quantitative statement.

   The sentence has been replaced with one that is more in-line with the overall message that the two sensors have similar behaviors.

   Evaluation of uncertainties in XCO_2 over small areas, as well as XCO_2 biases across land-ocean crossings, also indicate similar behavior in the error characteristics of both sensors.

2. Relevant for several figures: Explain at one place or in the corresponding figure caption what the "mu", "sigma", and "N" values stand for (and other parameters, e.g., "G").

   Updates have been made to many of the figure captions to explain the displayed statistics.

   Figure 2: Version 10 data volumes from a single detector footprint (4 of 8, 1-based) for the year 2020 gridded at 2.5° latitude by 5° longitude resolution for OCO-2 (top row) and OCO-3 (bottom row). The total number of measured soundings (N) for each sensor are shown in panels (a) and (c). Panels (b) and (d) show the number of soundings (N) that were assigned a good quality flag in the L2 Lite XCO_2 product. The percent of the total number of measurements is given as G. Grid cells containing less than 10 soundings are colored gray.

   Figure 3: Bar plots of the monthly number of good quality-flagged soundings for OCO-2 (a) and OCO-3 (b) by observation mode (colors) for the time period August 2019 through February 2022. The fractional percent for each observation mode is listed in the legend, along with the total number of good quality-flagged soundings (N).

Figure 5: Monthly maps of the bias correction for OCO-2 (top row) and OCO-3 (bottom row) for April 2020 (left) and August 2020 (right) gridded in 2.5° latitude by 5° longitude bins. The number of single soundings (SS) are given by N, while the mean (μ) and standard deviation (σ) of the gridded (bin) data are reported. Grid cells with less than 10 soundings are colored gray.

Figure 6: Monthly maps of XCO_2 for OCO-2 (top) and OCO-3 (bottom) for April 2020 (left) and August 2020 (right) at 2.5° latitude by 5.0° longitude resolution. The number of single soundings (SS) are given by N, while the mean (μ) and standard deviation (σ) of the gridded (bin) data are reported. Grid cells with less than 10 soundings are colored gray.

Figure 7: Meridional XCO_2 gradients at 5° latitude resolution by sensor and observation mode for April 2020 (a), and August 2020 (d). Only latitude bins containing at least 1000 soundings are shown. The total number of soundings (N) for each sensor and observation mode is given in the legend. Panels (b) and (e) show the differences in the monthly binned values (OCO-3 − OCO-2) for both land (pink diamonds) and ocean (purple circles) observations. Panels (c) and (f) show the differences in the monthly binned values (land − ocean) to demonstrate the land/ocean bias. The mean (μ) and standard deviation (σ) of the binned differences are given in the legend. Here, land observations include land-nadir, land-glint, land-TG and land-SAM, while ocean includes ocean-glint observations.

Figure 8: Good quality-flagged and bias corrected XCO_2 gridded at 10° latitude by 10 days for the overlapping time period August 2019 through February 2022 for OCO-2 land (a) and OCO-3 land (b). The gridded differences for land observations (includes land-nadir, land-glint, land-TG and land-SAM) are shown in panel (c). Panels (d), (e), (f) are the same, except for ocean-glint observations. The ordinate axis is scaled by the cosine of the latitude to elucidate the decreasing fractional surface area of the earth with increasing latitude. Data cells with less than 10 soundings are colored gray. In panels (c) and (f) the number of valid grid cells (N) is given, along with the mean (μ), standard deviation (σ), and maximum (max) and minimum (min) differences in the gridded values.

Figure 9: 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 B3. In each panel, the top two rows of statistics give the mean (μ) of the XCO_2 from all of the collocations, and the mean standard deviation in the XCO_2 (σ) from all of the collocations. 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)

Figure 10: Maps of XCO_2 signal ($XCO_2^{OCO} - XCO_2^{MMM}$) at 2.5° latitude by 5° longitude resolution for April 2020 (top row) and August 2020 (bottom row), for OCO-2 (left) and OCO-3 (middle). The number of single soundings (N SS) for all observation modes combined (all), combined land-nadir, land-glint, land-TG and land-SAM (land), and water-glint (water) are given. The mean (μ) and standard deviation (σ) of the binned values (bin) are also given for each observation mode. Grid cells containing less than 5 soundings are colored gray. Panels (c) and (f) show the differences in the binned values (OCO-3 − OCO-2) for grid cells in which both sensors have valid data. Here, the statistics are given only for all observation modes combined.

Figure 11: The XCO_2 signal ($XCO_2^{OCO} - XCO_2^{MMM}$) gridded at 10° latitude by 10 days for the time period August 2019 through December 2020 for OCO-2 land (a) and OCO-3 land (b). The gridded differences for land observations (includes land-nadir, land-glint, land-TG and land-SAM) are shown in panel (c). Panels (d), (e), (f) are the same, except for ocean-glint observations. The ordinate axis is scaled by the cosine of the latitude to elucidate the decreasing fractional surface area of the earth with increasing latitude. Data cells with less than 10 soundings are colored gray. In panels (a), (b), (d), and (e) the number of single soundings (N SS) is given, along with the mean (μ), standard deviation (σ), and maximum (max) and minimum (min) values of the gridded (bin) values. In panels (c) and (f) the number of valid grid cells (N) is given, along with the mean (μ), standard deviation (σ), and maximum (max) and minimum (min) differences in the gridded values.

Figure 12: 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, using binned median values (black filled circles) to highlight deviations from the one-to-one line (dashed black line). A least-squares linear fit to the binned data is shown (dotted red line), along with the correlation coefficient (R), the slope of the fit (m) and the fit offset (y).

Figure 14: Analysis of the coastal crossings data set for OCO-2 v10 (top row) and OCO-3 v10 (bottom row). Panels (a) and (c) show ΔXCO_2 (land-glint − ocean-glint) in 5° latitude bins with one standard deviation error bars as thin vertical lines. The number of latitude bins (N), and the mean (μ) and standard deviation (σ) of the binned values are given in the legend. Panels (b) and (d) show the frequency distributions of ΔXCO_2 for the individual coastal crossings. The number of coastal crossings (N), and the mean (μ) and standard deviation (σ) of the XCO_2 values for the individual crossings are given in the legend.

Figure A1: Bar plots of the monthly number of good quality-flagged soundings for a single footprint (4 of 8) for the full OCO-2 v10 data record, spanning 6-Sep-2014 through 28-Feb-2022. Observation modes are distinguished by colors. The fractional percent for each observation mode is listed in the legend, along with the total number of soundings (N).

Figure A2: OCO-2 v10 XCO_2 gridded at 10° latitude by 10 days for the time period September 2014 through February 2022. Panel (a) includes land-nadir, land-glint, land-TG and land-SAM (Land) soundings, while panel (b) is for ocean-glint soundings. The ordinate axis is scaled by the cosine of the latitude to elucidate the decreasing fractional surface area of the earth with increasing latitude. Grid cells containing less than 10 soundings are colored gray.

Figure A3: CO_2 concentrations and calculated atmospheric growth rates (AGR) from the OCO-2 v10 data record (orange), with a comparison to ACOS GOSAT v9 (gray) and NOAA GML marine surface values (blue), similar to Fig. 2 in Buchwitz (2018). Panel (a) shows the monthly CO_2 concentrations for each product. Panel (b) shows the calculated monthly values of the AGR with vertical error bars (see text for description of how the error is calculated). The mean (μ) monthly AGR is indicated with corresponding standard deviation. Panel (c) shows the calculated annual AGRs. The linear Pearson correlation coefficient (R) of the satellite versus the NOAA GML values is given, along with the mean difference in the annual AGR ($\mu\Delta$AGR) with corresponding standard deviation.

Figure A4: Maps of XCO_2 signal ($XCO_2^{OCO} - XCO_2^{MMM}$) at 2.5° latitude by 5° longitude resolution for April 2019 (top row) and August 2019 (bottom row), for OCO-2 v9 (left) and OCO-2 v10 (middle). The number of single soundings (N SS) for all observation modes combined (all), combined land-nadir, land-glint, land-TG and land-SAM (land), and water-glint (water) are given. The mean (μ) and standard deviation (σ) of the binned values (bin) are also given for each observation mode. Grid cells containing less than 5 soundings are colored gray. Panels (c) and (f) show the differences in the binned values (OCO-2 v10 − OCO-2 v9) for grid cells in which both sensors have valid data. Here, the statistics are given only for all observation modes combined.

Figure B1: Comparison of OCO-3 pointing optimization for vEarly (pink) and v10 (blue) spanning an approximately one year time period. Panel (a) shows the frequency distribution of the optimization distance. The orbit ranges and number of swaths are indicated in the legend, along with the mean, median, and maximum optimization distances (dx) in kilometers. Panel (b) shows the cumulative frequency distributions. The percent of the swaths (Frac N) with optimization distances greater than 1.25 km, 2.5 km, and 5.0 km are indicated in the legend.

Figure B3: Ad hoc bias correction of OCO-3 v10 XCO_2 for footprint 4 (FP-4) land measurements. Panel (a) shows the ΔXCO_2 (OCO-3 − OCO-2) for a set of collocated clusters of soundings versus the OCO-3 orbit number. The number of collocations (N), and the mean (μ) and standard deviation (σ) of the ΔXCO_2 are given in the legend. Panel (b) shows

the correlation between the ΔXCO_2 and the OCO-3 WCO_2 ZLO used to determine the correction. Panel (c) shows the OCO-3 WCO_2 ZLO (left) and magnitude in ppm of the ad hoc XCO_2 bias correction (right) versus OCO-3 orbit number. Panel (d) is similar to panel (a), except with the OCO-3 XCO_2 ad hoc correction applied. In all panels, the small grey dots indicate individual collocations, while the large black dots are binned median values. The vertical shaded regions in panels (a), (c), and (d) indicate the time period during which the OCO-3 instrument was powered down for a decontamination cycle.

Figure B5: The XCO_2 signal (OCO-3 − MMM) gridded at 15° latitude by 10 days for the time period August 2019 through December 2020. Panels (a) and (c) show results before and after the ad hoc XCO_2 bias correction for land-nadir observations, while panels (b) and (d) are for ocean-glint observations. The ordinate axis is scaled by the cosine of the latitude to elucidate the decreasing fractional surface area of the earth with increasing latitude. Data cells with less than 10 soundings are colored gray.

Figure B6: Analysis of OCO-3 XCO_2 from a set of paired intersecting orbits, i.e., self-crossings, over land. Panel (a) shows the scatter in XCO_2 between the early and late orbits. The number of collocations (N) and the mean (μ), standard deviation (σ) and maximum ΔXCO_2 (Δmax) are given in the legend. Also shown in the legend are the percent of the collocations having ΔXCO_2 greater than 0.25, 0.5, 1.0, 2.0, and 3.0 ppm. Panel (b) shows the ΔXCO_2 versus the time difference between self-crossings. Panel (c) shows the ΔXCO_2 versus the L2FP retrieved total aerosol optical depth (combined for the ascending and descending nodes).

3. Page 2, line 42: I would have expected that random error is mainly due to instrument noise but this most important factor is missing here.

   This was an omission. The phrase has been added to the following sentence:

   These biases and random errors are associated with multiple factors, such as instrument measurement noise, uncertainties in instrument calibration, error in CO_2 and O_2 gas absorption cross sections, complications in accurately representing aerosols and surface characteristics in the retrieval, and lack of accurate knowledge of the prior estimates of the atmospheric state that are used in the retrieval algorithm

4. Page 3 line 50: Please add ":" at the end of the first sentence.

   Corrected.

5. Page 3, line 51: "December 2019": In the Abstract (line 6) August 2019 is listed as first month. Please check and correct, if necessary.

   Corrected.

...the OCO-2 and OCO-3 v10 XCO_2 data volumes are analyzed for the overlapping time period August 2019 through February 2022...

6. Page 4, line 96: move “e.g.” into brackets.

Corrected.

7. Page 6, line 178: Reference to Crisp paper should be in brackets after ATBD. There are also several other places with format related reference issues, please check and correct.

Corrected.

8. Page 10, line 264: How is the "variability" / “real variability” defined in this case? This is important for Figures 11 and 12, too. Please provide more details in the Figure 11 and 12 figure captions.

The paragraph was modified to include definitions for “theoretical” and “actual” uncertainties related to the discussion of Figs 11 and 12:

Small areas, as introduced in Sec. 3.3.1, were used as XCO_2 truth-proxies in the development of the v10 quality filtering and bias correction. Small areas can also be used to derive realistic estimates of XCO_2 uncertainties for assimilation into inversion systems (Baker et al, 2022, Peiro et al. 2022). For each small area, the “theoretical” uncertainty is calculated as the median value of the XCO_2 uncertainties, which are described in Appendix B of O’Dell et al., 2012 and recorded in the L2Lite files. The “actual” uncertainty is calculated as the standard deviation of the retrieved XCO_2 in each small area. A minimum of 40 OCO soundings are required for each small area. Figures 11 and 12 show the results of an analysis on the small area XCO_2 uncertainties for land-nadir and ocean-glint observations, respectively.

As stated in item 2, the Fig 11. caption was updated. We also updated the text in several places throughout the manuscript to consistently use the term “uncertainty” (quantified by the standard deviation, σ) rather than “error” or “variability” (σ^2).

9. Page 10, line 268: Please add the definition of the pressure weighting function (it is the pressure difference of / over a layer divided by surface pressure?).

Correct, but it is also corrected for the presence of water vapor to allow for “dry-air” calculations. The equation and following paragraph was modified to:

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

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.

Note that, in the equation above, the vector quantities are now shown in **bold**, following journal guidelines.

10. Page 10, line 268: I assume that the CO2 averaging kernel is (essentially) the change of the retrieved XCO2 divided by the change of the true XCO2 (effectively obtained by perturbing the CO2 profile). Not clear what is meant here with "normalized"?

The response to the item above addresses the averaging kernel.

11. Page 12, line 291: Have the bias correction coefficients obtained using only the good (i.e., quality flagged data with QF=0) data or using all data?

Yes, only good quality-flagged soundings are used in determining the bias correction coefficients. The following clarification has been added to the preceding paragraph:

Bias correction coefficients are derived using only soundings that have been assigned a good quality flag.

12. Table 7: I recommend to always take the same number of digits after the decimal point.

The table was cleaned up to use consistent number of digits.

13. Table 7: Unit of N: maybe better write "(10^6 soundings)"?

This change was made. Furthermore, the units were changed to 10^3 to obviate the need for fractions.

14. Table 7: Unit of Fraction Good QF: either in the header of write it for each number individually.

The "(%)" was removed from the header in the table.

15. Table 7: Why is the Fraction Good QF of the models as truth proxy always smaller than for the others (e.g. 44% compared to 67% and 69% of the other proxies)?

The statistics reported in Table 7 are based on our Quick Test Set (QTS) data. The QTS is comprised of a subset of the full OCO data record. Soundings are "hand selected" to allow for algorithm development, e.g., derivation of the Empirical Orthogonal Functions used in the L2FP retrieval, quality filtering and bias correction, and a variety of error checking of the initial XCO_2 results. By nature, the selected TCCON and small area analysis tend to be only the soundings of the highest quality, whereas the selection of model data is more forgiving. For this reason, the fraction of good quality model data tends to be lower compared to TCCON and SAA.

16. Table 7: What is the reason to mention the average of all three truth proxies? Aren't they different by default and not really comparable?

Using an average was necessary in order to make more general statements in the discussion. E.g., “Overall, the fraction of good-quality soundings remains the same at approximately 60% for both sensors for land and ocean-glint.” But, yes, as stated in the comment above, the selection of the data itself yields imbalance in the good quality-flagged throughput.

17. Page 15, line 314 following: "The same is not true": Maybe correct this to "This is different".

    The sentence was modified to:

    For ocean-glint observations, the variance explained by the bias correction is similar for OCO-2 and OCO-3 at 58% and 55%, respectively.

18. Page 16, line 320: Section B -> Appendix B.

    Corrected.

19. Page 16, line 321: FIn -> In.

    Corrected.

20. Page 16, line 330: Please replace “M” by million (if this is what is meant here).

    Corrected.

21. Page 17, line 362 following: "while OCO-3 has a sinusoidal-like pattern of higher/lower densities over mid-latitude land": Why is this the case?

    It’s related to the precessing orbit of the ISS. We attempted a brief explanation here:

    The time-latitudinal coverage of OCO-2 is much smoother than OCO-3 due to the repeating sun-synchronous polar orbit. In contrast, OCO-3 has a sinusoidal-like pattern of alternating high and low densities over mid-latitudes, with the maximum value alternating in time between the northern and southern hemispheres. This is due to the precessing orbit of OCO-3 aboard the ISS, which introduces periodic variations in the portion of the earth that is viewed during daylight hours.

22. Figure 1: The country and continent boarders are too thick. It is difficult to see the data over land. Please use thinner lines. This is also true for Figs. 4, 5, 9 and A4.

    Figure 2 (1), Fig. 5 (4), Fig. 6 (5), Fig. 10 (9), and Fig. A4 have been updated with thinner coastlines. The new (old) figure numbers are provided for convenience.

23. Figure 1: What is the difference between white and grey? Please add this info to the figure caption.

    Figure 1 caption was updated. See response to item 2 above.

24. Figure 2: The meaning of N is not explained. I assume it is the total number of "good quality" measurements?

    The value N represents the total number of measured soundings, not the number of GQF. The Figure 2 caption has been updated. See response to item 2 above.

25. Figure 2: I would have assumed that Land Glint has less "Good Quality" pixels than Land Nadir. Why is this not the case? In addition, the percentage number is relative to the total number of "Good Quality" flagged data. Shouldn't this be relative to all measurements of the respective category, i.e. how many Ocean Glint are "good" of all Ocean Glint measurements? I think this would be better comparable between panel a and b?

    Did the reviewer mean to imply that there should be less GQF **ocean-glint** compared to **land-nadir**? There is no information in this figure concerning the distinction between **land-glint** and **land-nadir**! The percentage clear (G) is relative to the total number of satellite observations.

26. Page 17, line 351: Using "Figure" instead of "Fig." at the beginning of a sentence.

    Corrected.

27. Page 18, line 366: Explain JEM-EF.

    The JEM-EF was already defined in Section 2, third paragraph, second sentence. For clarity, we have now separated the single sentence into two sentences for better clarity.

    The OCO-3 instrument is mounted as an external payload on the Japanese Experimental Module Exposed-Facility (JEM-EF) aboard the ISS. The ISS flies in a precessing orbit with a varying time-of-day local overpass across a 63-day illumination cycle.

28. Page 20, line 393 following: "thereby forcing the models to rely more on the prior spatial variations": What does this mean? What are "prior spatial variations" and why do models for OCO-2 rely on the high latitudes? Do you mean deriving global flux estimates using both instruments?

    We agree that the phrase "prior spatial variations" is ambiguous. What we meant was simply "the model prior CO_2 fields". The discussion has been updated as follows:

    This has meaningful consequences for the interpretation of flux estimates derived from inverse modeling of the OCO-2 and OCO-3 XCO_2 concentrations. For example, OCO-3 cannot directly capture the strong summer draw down of CO_2 in the northern boreal forests. For this time and location, an inversion of OCO-3 CO_2 fluxes must rely more on the model prior values since there is no information provided by satellite measurements, whereas an assimilation of OCO-2 XCO_2 for this same time and location would provide information to the models since the satellite observed this location at this time.

29. Page 20, line 406 following: I think you describe the mean values (mu values?) in Fig. 6, but are the given standard deviations (sigma?) a measure for uncertainty? If yes, the mean values accounting for the uncertainties are much closer to each other.

Correct, the reported σ give the uncertainty in the mean. I think I understand your point that the mean land and ocean differences are not statistically significantly different from one another. The figure caption was updated (see response to item 2) and the discussion has been slightly modified:

The difference plots (panels b and d) indicate that OCO-3 ocean-glint is generally biased low relative to OCO-2 by about 0.3 to 0.4 ppm with uncertainty, σ, of approximately 0.2 ppm. For land observations, the differences vary significantly with latitude, making inferences difficult.

30. Figure 4: Do positive values mean underestimation of uncorrected data? Please add explanation.

The subtraction of an annual, global median XCO_2 bias correction value was meant to center the color scale on zero. This was a legacy from previous research on ACOS GOSAT v9 data, where the size of the bias correction was very different between satellite viewing modes. For OCO-2 and OCO-3, this is not the case, i.e., both land and ocean-glint viewing have bias corrections of comparable magnitude. We have therefore taken the liberty to regenerate the maps without the annual, global median XCO_2 bias correction removed. The color scale now ranges from 0 (blues) to 4 (reds) ppm, and shows the actual magnitude of the bias correction. We feel that this is more intuitive for interpretation. We hope that the reviewers agree. The new figure is given below, along with updates to the discussion. This is now Fig. 5 in the updated manuscript.

Figure 5 shows maps of the magnitude of the bias correction (ppm) for both sensors for April and August, 2020. The patterns look qualitatively similar, with bias corrections ranging from zero to ≈ 2 ppm in the midlatitudes and polar regions, and bias corrections of up to ≈ 4 ppm over the Sahara and dust outflow regions, as well as the tropical oceans. The mean global bias corrections are slightly larger for OCO-2 compared to OCO-3 for both months, but the uncertainties are slightly smaller for OCO-3. The 2020 annual median bias correction was 1.81 ppm for OCO-2 and 2.11 ppm for OCO-3. Note that, the OCO-3 v10.4 time-dependent ad hoc bias correction discussed in Appendix B3 has been removed.

31. Figure 6 caption: Are "Ocean" and "Land" observation modes? Do you mean glint over ocean and nadir over land?

See item 2 above. This is now Fig. 7 in the updated manuscript. The figure caption has been updated to state that:

..."land includes land-nadir, land-glint, land-TG and land-SAM observations, while ocean includes ocean-glint observations".

32. Figure 6 (a): Is there a reason for the nearly 2 ppm land ocean difference at northern latitudes?

This is now Fig. 7 in the updated manuscript. This appears to be a latitudinally-dependent land/ocean bias in the XCO_2 data. We added an additional panel to the figure to demonstrate this, along with brief discussion. At the moment we have not been able to identify the underlying cause, and additional investigation will be required. The new figure and updated discussion are shown here:

To further demonstrate the agreement between the two sensors, panels (a) and (c) of Fig. 7 show the meridional behavior of XCO_2 for both sensors and observation modes for April and August 2020, respectively. Here, the resolution is 5°

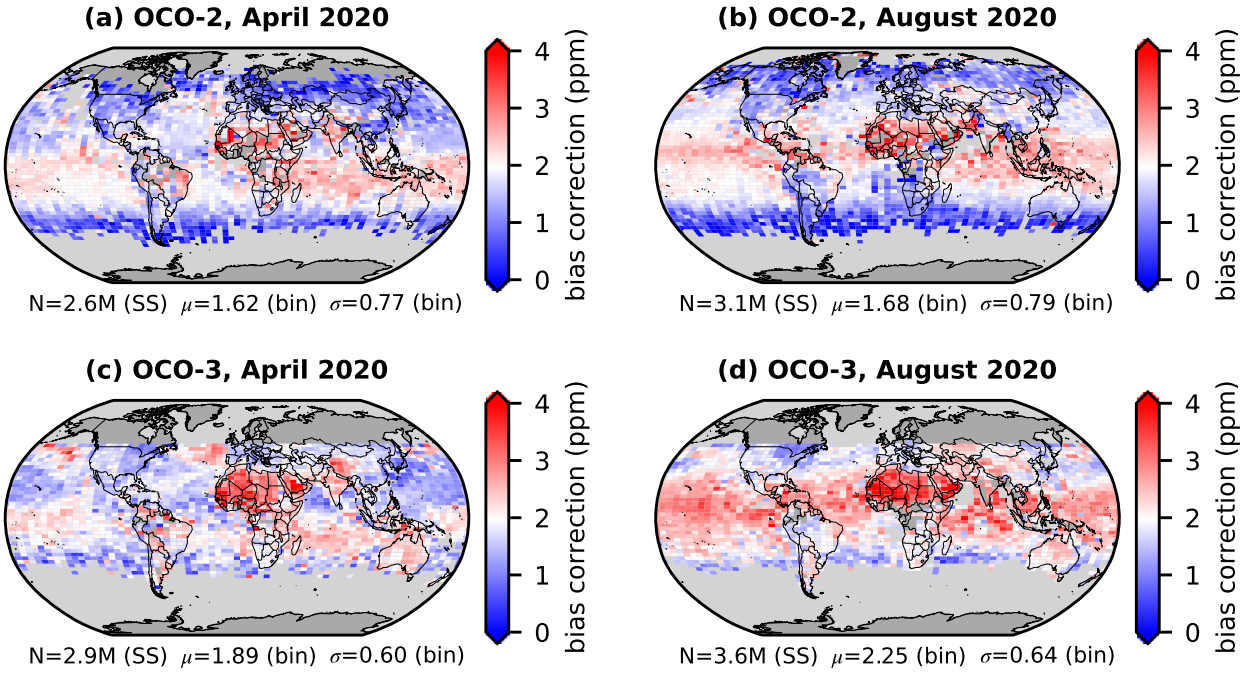

Figure 1. **(Figure 5 in the updated manuscript)** Monthly maps of the bias correction for OCO-2 (top row) and OCO-3 (bottom row) for April 2020 (left) and August 2020 (right) gridded in 2.5° latitude by 5° longitude bins. The number of single soundings (SS) are given by N, while the mean (μ) and standard deviation (σ) of the gridded (bin) data are reported. Grid cells with less than 10 soundings are colored gray.

latitude bins, and the monthly median OCO-2 XCO_2 has been subtracted. In April, when XCO_2 concentrations are near their annual maximums in the extra-tropical northern hemisphere, the meridional gradients are strong over both land and ocean. In August, when the northern hemisphere biosphere is fully active, XCO_2 is within ≈ 1 ppm of the global median for latitudes below approximately 40° N, but much lower than the global average at higher northern latitudes. The difference plots (OCO-3 − OCO-2), shown in panels (b) and (e), indicate that OCO-3 ocean-glint is generally biased low relative to OCO-2 by about 0.3 to 0.4 ppm with uncertainty, σ, of approximately 0.2 ppm. For land observations, the differences vary significantly with latitude, making inferences difficult. Panels (c) and (f) show the zonally averaged differences between land and ocean observations, which are expected to be close to zero for both sensors. Based on the results for these particular months, OCO-2 and OCO-3 are in approximate agreement, with land/ocean biases ranging $\approx \pm 2$ ppm, with significant variation by latitude. These same behaviors were observed for most months in 2020. This latitudinally-dependent land/ocean bias is an unexpected feature of the data set that requires further investigation. The analysis for April and August 2021 (not shown) was qualitatively very similar.

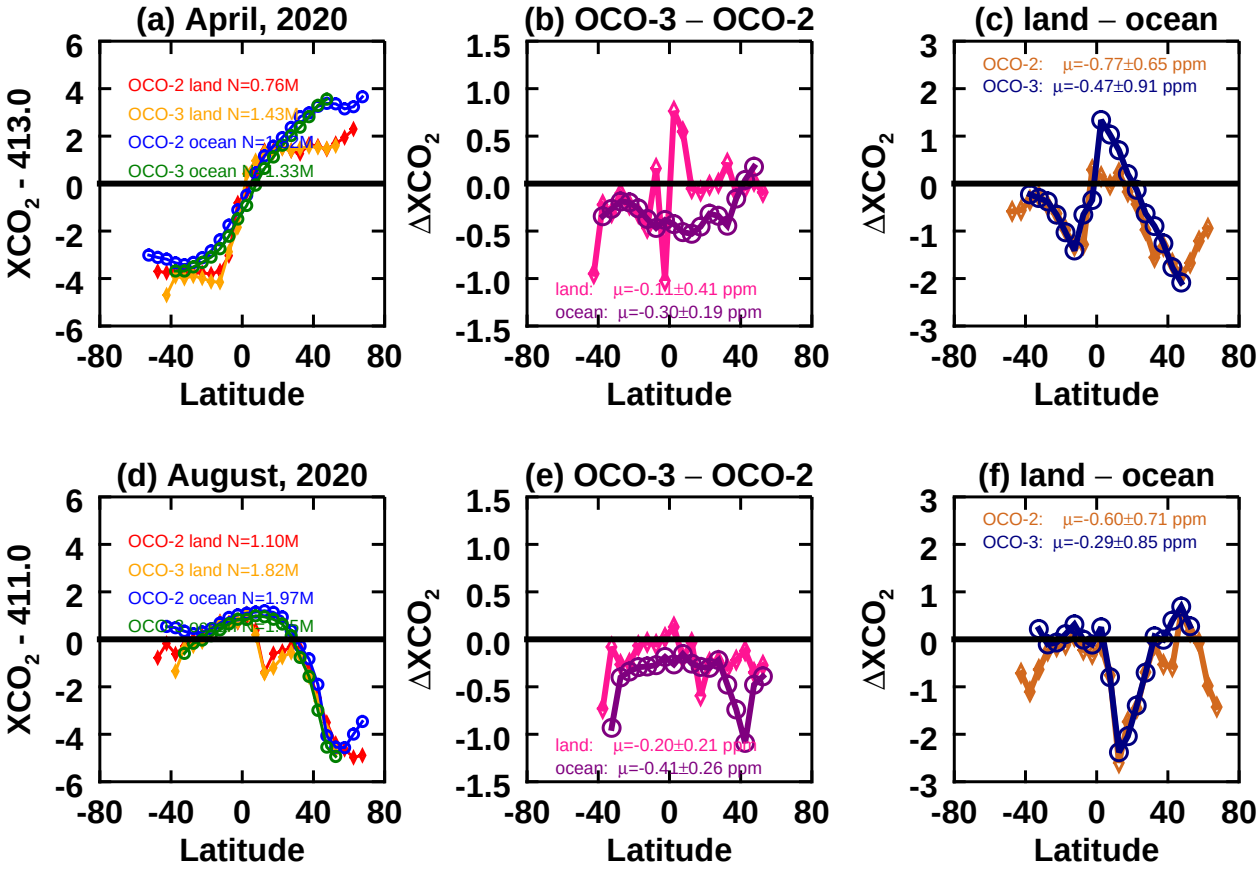

Figure 2. (Figure 7 in the updated manuscript) Meridional XCO_2 gradients at 5° latitude resolution by sensor and observation mode for April 2020 (a), and August 2020 (d). Only latitude bins containing at least 1000 soundings are shown. The total number of soundings (N) for each sensor and observation mode is given in the legend. Panels (b) and (e) show the differences in the monthly binned values (OCO-3 − OCO-2) for both land (pink diamonds) and ocean (purple circles) observations. Panels (c) and (f) show the differences in the monthly binned values (land − ocean) to demonstrate the land/ocean bias. The mean (μ) and standard deviation (σ) of the binned differences are given in the legend. Here, land observations include land-nadir, land-glint, land-TG and land-SAM, while ocean includes ocean-glint observations.

33. Page 23, line 436 following: "Only overpasses with at least one hundred good quality-flagged OCO soundings were retained.": Does this mean, at least one hundred good quality soundings for one overpass within the 2.5x5.0 box?

Correct. The sentence was modified to:

Only overpasses with at least one hundred good quality-flagged OCO soundings within the 2.5° latitude by 5.0° longitude grid box were retained.

34. Page 24, line 440: "(includes land-nadir, land-glint, and land-target)": This explanation should be added to the description of the previous figures, too, where "land" and "ocean" observation are mentioned.

This correction has been made.

35. Page 26, line 491: The bias statements are only true over the oceans.

The statement has been slightly modified to:

Both sensors exhibit spatially coherent biases against models on the order of half of a ppm. Over oceans, the satellite estimates of XCO_2 are generally biased low relative to the models in the SH and biased high in the NH.

36. Page 27: line 512 following: I don't understand this sentence. Do you mean that the signal changes sign at the equator starting from February 2020?

Both reviewers had trouble interpreting this discussion. The discussion has been slightly modified to:

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 December 2019, immediately following the OCO-3 PMA calibration that was described in Sec. 2.1.2 of Taylor et al. (2020), and continuing through January 2020 when

the next OCO-3 decontamination cycle occurred. The origin of this feature in the OCO-3 v10.4 XCO_2 record is not currently understood. 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.

37. Page 27, line 513 following: Where does this northward propagation come from?

    The mention of the northward propagation has been removed in the discussion, as given in the previous item.

38. Page 28, lines 524-526: But isn't this the same metric as in Fig. 9 where you wrote that this is not the difference between the OCO instruments, but a measure how far from the MMM they are?

    Yes, both Fig 9 and Fig 10 (OCO vs MMM) (Figs. 10 and 11 in the updated manuscript) indicate that the sensors tend to agree better with one another than either agrees with the MMM. Because the two instruments are being compared to the same MMM, these results illustrate a difference between the instruments.

39. Page 28, line 537: Is the "theoretical error" the one that is reported as uncertainty in the XCO2 Level 2 product files?

    Yes, that is the correct interpretation. That section of the text has been slightly updated to:

    Small areas, as introduced in Sec. 3.3.1, were used as XCO_2 truth-proxies in the development of the v10 quality filtering and bias correction. Small areas can also be used to derive realistic estimates of XCO_2 uncertainties for assimilation into inversion systems (Baker et al., 2022). For each small area, the "theoretical" uncertainty is calculated as the median value of the XCO_2 uncertainties, which are described in Appendix B of O'Dell et al., (2012) and recorded in the L2Lite files. The "actual" uncertainty is calculated as the standard deviation of the retrieved XCO_2 in each small area. A minimum of 40 OCO soundings are required for each small area. Figures 11 and 12 show the results of an analysis on the small area XCO_2 uncertainties for land-nadir and ocean-glint observations, respectively.

40. Page 28, line 533: "land": maybe write land-nadir?

    Done.

41. Page 28, line 537: "it is highly correlated": What is the correlation coefficient?

    The plots and discussion have been updated as follows:

    The frequency distributions of the XCO_2 uncertainties over land-nadir small areas, as shown in panels (a) and (c) of Fig. 11 for OCO-2 and OCO-3, respectively, indicate that the actual uncertainties (green curves) are slightly larger with a wider distribution of values compared to the L2FP noise-driven theoretical uncertainties (blue curves). Although the actual uncertainties tend to be biased high, they are highly correlated with the theoretical uncertainties, having R values of 0.95 and 0.98 for OCO-2 and OCO-3, respectively, as seen in panel (b) and panel (d) for OCO-2 and OCO-3, respectively. Here, the median binned values are shown, for clarity, rather than the uncertainties for individual small areas. These results imply that there are additional spatially-correlated systematic uncertainties in the ACOS retrieval

385 over small areas, and these additional uncertainties are similar for both sensors.

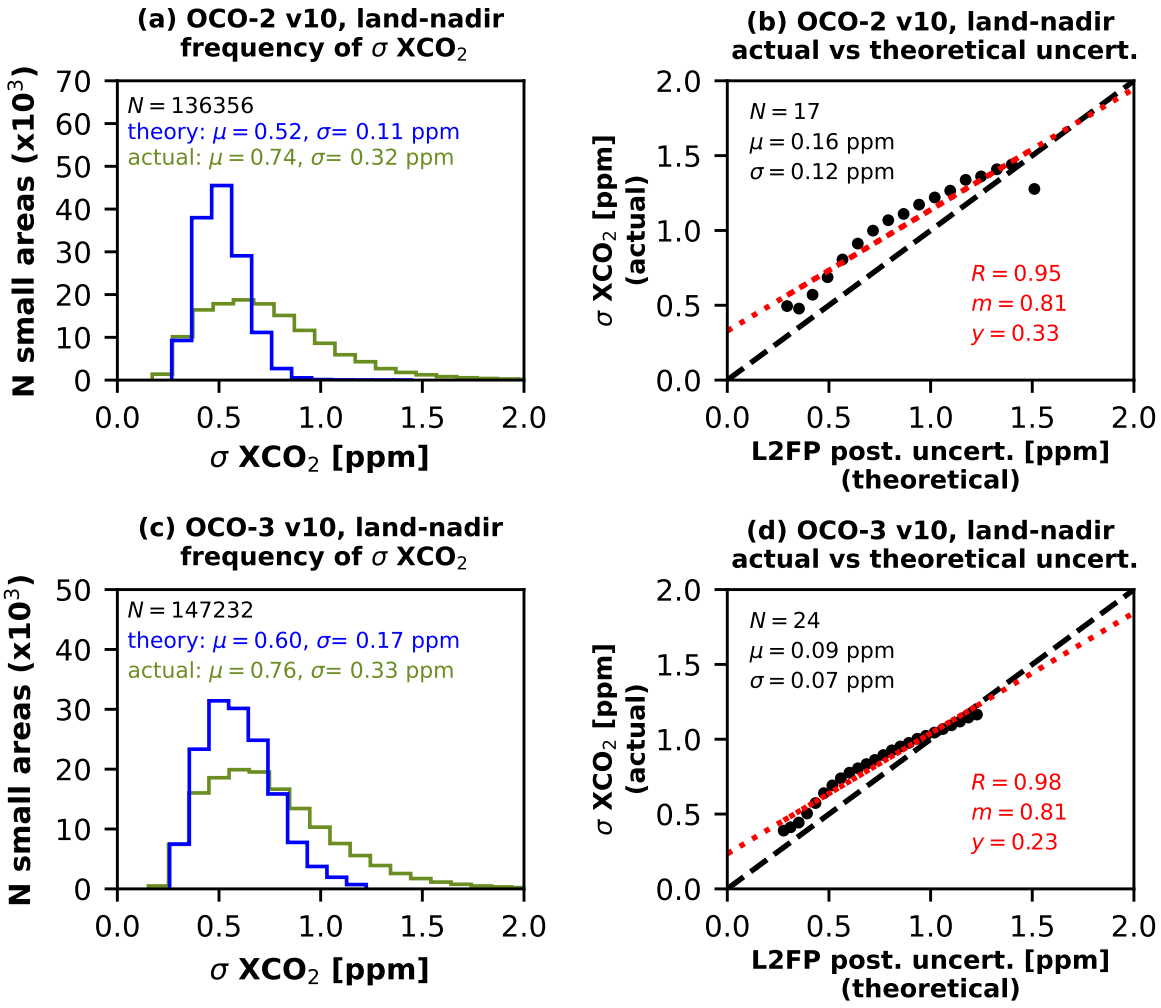

Figure 3. (Fig. 12 in the updated manuscript) Analysis of XCO_2 uncertainties for land-nadir small areas for OCO-2 (top) and OCO-3 (bottom). Panels (a) and (c) provide the frequency distributions of both the theoretical uncertainties (blue curves) of the retrieved XCO_2 as reported in the L2Lite file 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.

For ocean-glint observations, shown in Fig. 13, XCO_2 uncertainties in small areas have different characteristics compared to land observations. The frequency distributions of the XCO_2 uncertainties, shown in panels (a) and (c) for OCO-2 and OCO-3, respectively, indicate that the actual uncertainties (green curves) are often lower than the L2FP noise-driven theoretical uncertainties (blue curves), especially for OCO-3. Furthermore, even though the actual uncertainty correlates reasonably well with the theoretical uncertainty (R=0.53 for OCO-2 and R=0.89 for OCO-3), the line-of-best-fit falls well off of the one:to:one line, with a slope of 0.18 for OCO-2 and 0.31 for OCO-3. For the lowest theoretical uncertainties, the actual uncertainty is near or somewhat higher than anticipated, but when the theoretical uncertainties are large, the actual uncertainties are significantly lower than anticipated. This is unexpected for a well-characterized retrieval, and indicates that there is some non-linearity or other systematic behavior in the v10 ocean-glint retrieval. Early efforts to develop the OCO-2 v11 product suggest this is due to the parameterization of the ocean surface reflectance model. Additional investigation is underway.

42. Page 28, line 537: "values around 43%": Isn't a discrepancy of 43% a large difference?

    The discussion has been modified to be more clear based on the statistics and line fit presented in the updated figures. See previous comment.

43. Figure 10: Why is there a double "Delta" letter in the captions of (c) and (f)? Maybe you can adapt these titles to match with Fig. 9 right column?

    The plot titles have been updated to use the new "Δ signal" nomenclature rather than $\Delta\Delta XCO_2$.

44. Figures 11 and 12: using sigma for "variability" and the standard deviation in the figure is pretty confusing. Maybe correct the sigma in the captions by "var."?

    Here, and elsewhere, in the text, terminology has been made consistent to use σ to represent "uncertainty". The use of the term "variability" has been removed.

45. Figures 11 and 12: Is the "uncertainty" in the title of panels (b) and (d) the variability?

    Here, and elsewhere, in the text, terminology has been made consistent to use σ to represent "uncertainty". The use of the term "variability" has been removed.

46. Figures 11 and 12: I don't understand how panel b is generated from the data shown in panel a. I would have expected many more dots (about 136356).

    Panel (b) plots binned median values, rather than the individual points for each small area. Plotting the later produces a super-noisy scatter plot that is difficult to interpret - it would need to be a heat map. But we feel that plotting the binned median is the most elucidating way to display this particular data because it clearly demonstrates deviations from the one:to:one line. The discussion has been modified and the figure caption has been updated to include the statement:

    Panels (b) and (d) show the correlation of the actual uncertainties against the theoretical uncertainties, using binned median values (black filled circles) to highlight deviations from the one:to:one line (dashed black line).

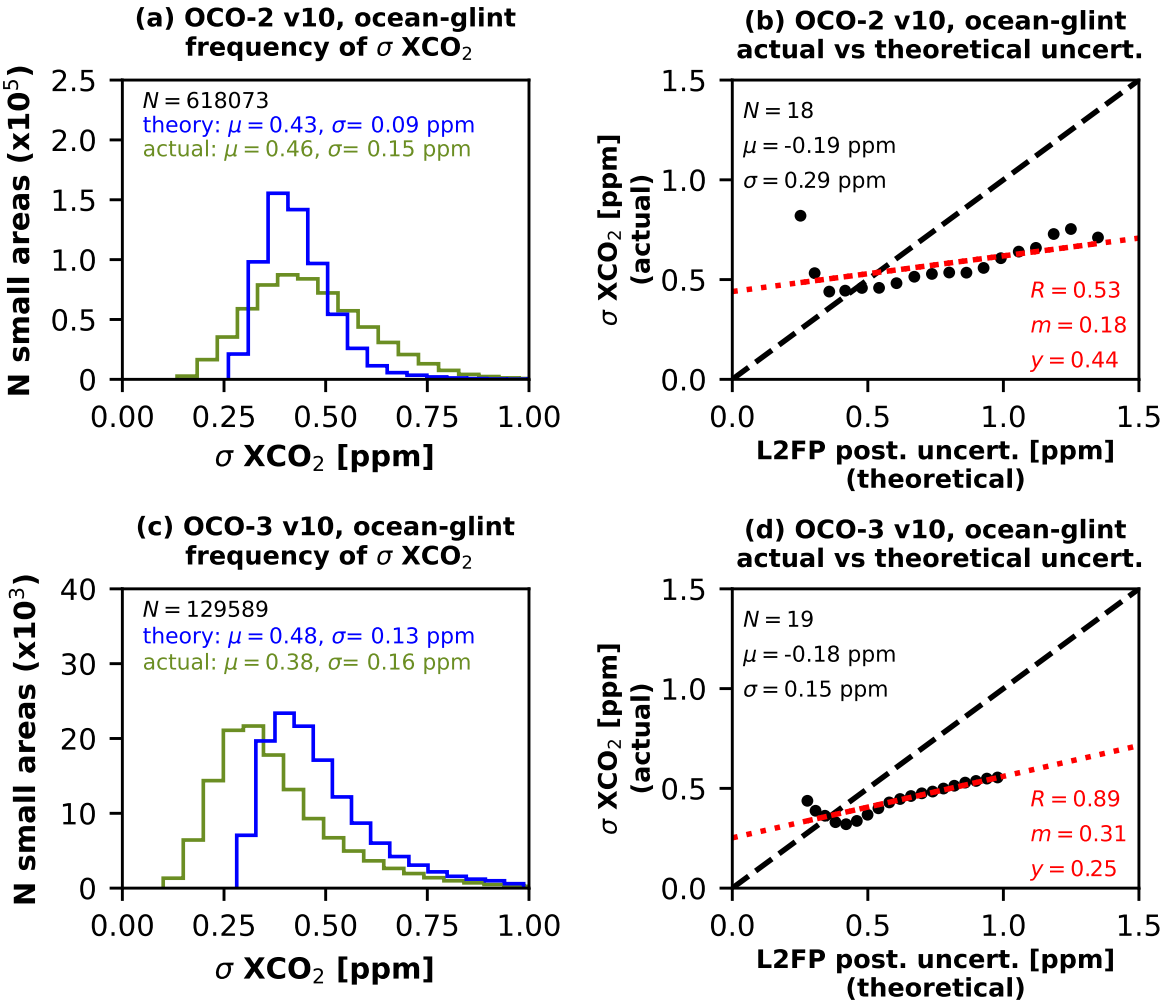

Figure 4. (Fig. 13 in the updated manuscript) Same as Fig. 3, but for ocean-glint small areas.

47. Figures 11 and 12: What is the unit of "sigma XCO2" in panels a and c of Fig. 11 and 12?

The units are parts per million (ppm). This was added to the x-axis labels.

48. Page 30, line 540: Where does the value of 1 sigma come from?

The statement has been removed.

These results imply that there are additional spatially-correlated systematic uncertainties in the ACOS retrieval over small areas, and these additional uncertainties are similar for both sensors.

49. Page 30, line 546: Would you expect the "uncertainty" to be on the one-one-line?

Correct. The discussion has been rewritten to be more clear.

Small areas, as introduced in Sec.3.3.1, were used as XCO_2 truth-proxies in the development of the v10 quality filtering and bias correction. Small areas can also be used to derive realistic estimates of XCO_2 uncertainties for assimilation into inversion systems (Baker et al., 2022, Peiro et al., 2022). For each small area, the "theoretical" uncertainty is calculated as the median value of the XCO_2 uncertainties, which are described in Appendix B of O'Dell et al., (2012) and recorded in the L2Lite files. The "actual" uncertainty is calculated as the standard deviation of the retrieved XCO_2 in each small area. A minimum of 40 OCO soundings are required for each small area. Ideally, the actual uncertainties are highly correlated to the theoretical uncertainties, with the relationship having a one-to-one dependence (slope, m=1), and the y-intercept falling at zero (y=0). Figures 11 and 12 show the results of an analysis on the small area XCO_2 uncertainties for land-nadir and ocean-glint observations, respectively.

50. Page 31, line 553: "land and ocean": Are they all in glint or all in nadir mode?

Previously, all land observations were mixed in with ocean-glint observations. Now only land-glint data are used in the analysis to compare to the ocean-glint observations. This did not affect the OCO-2 data much since complete orbits are made in glint viewing. But for OCO-3, which has more dynamic pointing, only about half of the coastal crossings were in land-glint, while the other half were in either land-nadir or land-TG/SAM. The statistics changed slightly, but overall the summary is the same. The text in that paragraph was modified to:

Figure 13 shows analysis of the coastal crossings data set for the v10 OCO-2 (top row) and OCO-3 (bottom row), containing $\approx 20\times10^3$ and 0.5×10^3 crossings, respectively. Panels (a) and (c) show ΔXCO_2 in 5° latitude bins with one standard deviation error bars as thin vertical lines. The mean land-ocean difference tends to be positive (negative) in the southern (northern) extra-tropical latitudes for both sensors. Panels (b) and (d) show the frequency distributions of ΔXCO_2 for the individual coastal crossings. The means are biased -0.10 $\pm$ 0.83 and -0.11 $\pm$ 0.88 ppm for OCO-2 and OCO-3, respectively. The uncertainties are due presumably to local geometry, aerosol, and surface effects. In the future, we plan to assess whether these biases can be explained by any retrieved parameters or other independent information such as population centers, which may help to shed light on the cause of these ubiquitous land-ocean XCO_2 differences.

51. Page 31, line 562: Please replace "k" by kilo (if this is what is meant here).

The "k" notation was replaced with $\times 10^3$.

52. Page 31, line 574: Please explain TSIS-SIM.

    The Total Solar Irradiance Sensor (TSIS) Spectral Irradiance Monitor (SIM) was defined in Sec. 3.

53. Page 34, line 640: Please replace "y" by years (if this is what is meant here).

    Done.

54. Figure A1: Caption: "4 of 8": Numbering starts with 1, or?

    Modified caption:

    (4 of 8, 1-based)

55. Page 42, line 749: Please replace "m" by minutes (if this is what is meant here).

    Done.

56. Page 42, line 766: "Researchers are urged to use these files . . . ". Please add this important recommendation to Section "Data availability" to minimize the risk to overlook it.

    The following statement was added to the Data Availability section:

    For OCO-3, researchers are urged to use the v10.4 Lite files, and avoid use of the XCO_2 reported in the v10 L2 Standard product which do not have the ad hoc bias correction applied.

57. Figure B3: Definition of delta_XCO2: Is it OCO-3 minus OCO-2? Please add this information.

    Yes, (OCO-3 − OCO-2). This information is already included in both the figure caption and on the ordinate label.

58. Figure B5: Caption: Please add that a difference is shown and not absolute XCO2 values.

    This was already included on the ordinate label. The figure caption was updated to include:

    The XCO_2 signal (OCO-3 − MMM) gridded at 15° latitude by 10 days for the time period August 2019 through December 2020.

59. Page 46, line 790: Add after Panel (a): "of Fig. B6".

    Done.

60. Page 47, line 818: What is the "light-weight per-day format"? Is this the format of the "Lite files"? If yes then please use the term Lite files.

    Correct. Sentence changed to:

    The Lite files, which include the quality flag and bias corrected estimates of XCO_2, can be obtained at...

61. Page 13, line 303: Variance: I think sigma_1 and sigma_2 need to be squared.

The calculation was performed correctly, but the wording was not accurate. "Variance" has been replaced with "uncertainty" and "standard deviation". The text has been updated as follows:

The two right-most columns of the table show the percent of the variance explained by the quality filtering (compared to the raw XCO_2) and the bias correction (compared to the quality filtered XCO_2), respectively. The variance explained is calculated as $1-(\sigma_2/\sigma_1)^2$, where σ_1 is the original uncertainty (standard deviation) in the ΔXCO_2 and σ_2 is the remaining uncertainty. For each sensor and for both land and ocean-glint observations, the mean values from the three truth proxies are provided to help summarize the statistics.

62. Page 16, line 321: Fin -> In.

Corrected.

63. Page 21, line 409: "10 day" -> "10 days"

Corrected.

64. Figure 10, caption: Add space before "gridded".

Corrected.

65. Page 26, line 481: course –> coarse?

Corrected (two instances).

66. Page 36, line 676: /gt

Corrected.

67. Page 42, line 741: add comma after B3.

I'm sorry - I could not find this typo in the original manuscript!