Referee Response to Dr. Andy Jacobson.

We thank Andy for his thoughtful comments. The reviewer's comments are in **red bold text**, and our responses follow in black text.

Major Comments:

1. Neither TCCON nor OCO-2 measures carbon dioxide. These instruments measure spectra. It is through a complex retrieval process that the column abundance of CO2 is inferred. This retrieval process is sufficiently complex, uncertain, and subject to systematic bias that it must be referred to as an estimate, model result, or retrieval. These column abundance retrievals cannot be calibrated and it is simply unacceptable to imply a direct relationship between the observed quantity and xCO2. My objection extends to the use of "measurements" in the title of this manuscript.

It is unclear what *would* constitute a measurement by this definition. No instrument (TCCON, OCO-2, Picarro, NDIR, GC) “measures” CO_2 directly. Each provides a “response” which is a number related to the CO_2 amount in an unknown target sample or atmospheric path. An in situ analyser is used as a comparator between the response of an unknown with that of a standard which does have an SI-traceable CO_2 amount. A remote sensing system provides a response which is uncalibrated, but as close to the “true” amount as our knowledge of the physics of the measurement allows. In TCCON and OCO-2 this is better than a few percent. Aircraft or AirCore in situ measurements provide a reference profile against which the remote measurement can be compared to provide a “partial calibration”, albeit with some limitations which are rigorously dealt with in the error budget.

It is our view that the entire process from recording an instrument signal through to a CO_2 amount constitutes a “measurement” in the common definition and usage of the word. We have clarified our usage of the word “measurement” in the text, as requested.

2. TCCON retrievals are not calibrated to the WMO CO2 scale. They have been compared to a modest number of partial-column in situ observations, and made to be (partially) consistent with those measurements. There remain apparent site-to-site biases against the WMO scale, high-frequency stochastic variability with a range of about 2 ppm, and suggestions that diurnal trends in the retrievals exhibit SZA dependencies that exceed plausible variability. Bundling this together into the shorthand phrase "tied to the WMO scale" is acceptable, but only if the shorthand is fully described. Please add text explaining the meaning of this term.

New text has been added (starting at sentence 2 below):

“Placing the OCO-2 data on the World Meteorological Organization's (WMO) trace-gas standard scale is crucial to obtaining accurate flux estimations that are consistent with the state-of-the-art inversions of surface in situ CO2 data, which are carefully calibrated to the WMO scale (Zhao2006). The TCCON data are tied to the WMO scale and serve as the link between the calibrated surface in situ measurements and the OCO-2 measurements.

To tie the TCCON measurements to the WMO scale, over 30 profiles of in situ CO2 have been measured directly overhead of 15 TCCON stations with aircraft carrying carefully calibrated instrumentation {Wofsy2011,Pan2010,Singh2006} or AirCore {Karion2010}. These profiles, the first of which occurred in 2004, vary in altitude range, depending on the vehicle, and thus must be combined with a priori information about the CO2 in the highest altitudes of the atmosphere to generate a full vertical profile. These profiles are then integrated, smoothing with the TCCON averaging kernel and a priori profile to compute the best estimate of the ``true" XCO2 value. Integrated profiles are compared with the retrieved XCO2 from the TCCON spectra and result in a highly linear relationship which defines a multiplicative bias between the TCCON XCO2 and the best

estimate of the “truth”. Removing this bias from the TCCON XCO2 ties it to the WMO scale. The details of this method of tying the TCCON XCO2 to the WMO scale are described in Wunch2010, Messerschmidt2011, and in Wunch2015.”

3. Sections 3 and 4 in general, and with reference to all statistics reported in this manuscript (including the abstract and conclusions). A striking result in this paper is the standing pattern of surface topography- and/or albedo-dependent bias as revealed in the target mode comparisons of figures 9-12. The use of daily medians of target mode retrievals (section 3) and within-box median OCO-2 retrieved xCO2 for the nadir and glint-mode comparisons (section 4) removes much of the effect of this standing pattern of spatial bias. Most, but not all, of the statistics reported here are apparently for these rather abstract daily-median quantities. These statistics are not representative of individual retrievals. One effect of this choice is to reduce systematically the reported magnitude of the differences between TCCON and OCO-2. Given that this paper is likely to be cited during further analysis of the OCO-2 retrievals, I believe that this spatial bias needs to be quantified and included in the reported statistics. That spatial bias is likely to be an important factor to include in analysis of the OCO-2 retrievals, such as surface flux estimation. It appears that this spatial bias is included implicitly in figures 5 & 6, but I’m not entirely clear on that. The maps in figures 9-12 suggest that quantifying this bias is certainly possible for target-mode retrievals. Depending on the density of coverage for the nadir- and glint-mode retrievals, I expect that a similar analysis can be conducted for those operation modes as well.

It was not the intention for individual OCO-2 measurements to be used in isolation (cf. Rayner and Obrien, 2001; Miller et al. 2007), but to be used to constrain regional-scale sources and sinks. The strength of the OCO-2 measurement technique is that it produces millions of measurements, albeit with lower precision, and thus aggregation is the appropriate usage of the OCO-2 data. I would argue the same is true for TCCON data, but to a lesser extent, as the data can be sensibly aggregated somewhat in time, but obviously not spatially. Thus, using medians of coincident OCO-2 and TCCON measurements is appropriate, and we have attempted to quantify the uncertainties on those statistics in this paper. As requested, we have gone through the paper carefully to clarify exactly what the reported statistics mean.

These quantities do serve to reduce the random component of the measurement uncertainty. However, as the reviewer mentioned, they can also mask the fine-scale spatial biases that average out across a target scene that become apparent only when looking closely at the thousands of measurements within a target-mode maneuver. We have, of course, looked for these topography and albedo-dependent biases on a global scale in the science mode measurements, and they do not explain the main sources of bias we see in the data, which have been removed through the standard “bias correction” methodology (i.e., Mandrake et al., 2015).

We have attempted to better quantify the spatial biases we see in the target-mode data in section 3.2.

A similar complaint can be made for the use of daily-median TCCON retrievals in the comparisons, since this removes both the single-sounding noise and potentially unrealistic diurnal variability. I find this use of the daily median more reasonable, since the stochastic component appears like random noise, and the comparison is of course with a sun-synchronous orbiting platform. Some admission of these features of the TCCON retrievals is needed.

A new sentence (the second below) has been added to clarify:

“The median XCO2 within the coincidence box recorded on the same day as the TCCON measurement is compared with the TCCON daily median. We choose to compare OCO-2 nadir and glint mode XCO2 with the TCCON daily median values because the median reduces the random component of the TCCON error budget, it is less sensitive to outlier measurements, and it weights the results to local noon where solar zenith angle

changes are slowest, and the timing is better matched with the overpass time of OCO-2's orbit."

Minor Comments:

L1. OCO-2 does not measure carbon dioxide. See major comment 1.

See response to major comment 1.

LL8-10. If the global relationship between TCCON and OCO-2 can be statistically characterized, then so can the local relationship. Quantify the local-scale residual biases.

Figure A1 shows the OCO-2 comparisons with each TCCON station individually. Due to a low number of coincident measurements at many of the sites, local-scale residual biases are difficult to compute in a rigorous manner. This really illustrates the strength of the target-mode measurements, where one maneuver can generate enough data to "map" XCO2 in a small area with sufficient spatial density to look at local-scale biases.

LL27-28. "unprecedented precision from space" is a grandiose and misleading phrase. The instrument need only be more precise than GOSAT to qualify, and the statement relies heavily on the rather fine distinction between precision and accuracy.

The OCO-2 measurements are much more precise than GOSAT and thus it is possible to see the additional accuracy problems. The word "unprecedented" has been changed to "high".

L29. See major comment 1.

See response to major comment 1.

L31. See major comment 1.

See response to major comment 1.

LL 43-45. It is not clear from this description whether land glint, land nadir, and target mode over land use the same surface reflectance model. A strict reading suggests only that land glint uses the Lambertian surface model. Please clarify.

Clarified: "The glint data are often separated into glint over land (``land glint") and glint over water (``ocean glint"), as the two modes use different surface reflectance models: Lambertian over land (matching the surface model of the nadir observations), and Cox-Munk with a Lambertian component over water."

L48. Boreal forests certainly contribute a significant, perhaps major, fraction of the annual cycle in atmospheric CO2, but they are not THE driver.

Changed "the" to "a" driver of the CO2 seasonal cycle.

L49. Are these 233 orbits the same as the "orbit paths" discussed later (LL86-87)?

Clarified. The new text reads:
"OCO-2 has a geographical "near-repeat" after 16 days. During each 16-day period, the satellite orbits the earth 233 times, each orbit along a distinct "orbital path"."

LL86-87. What is an "orbit path"?

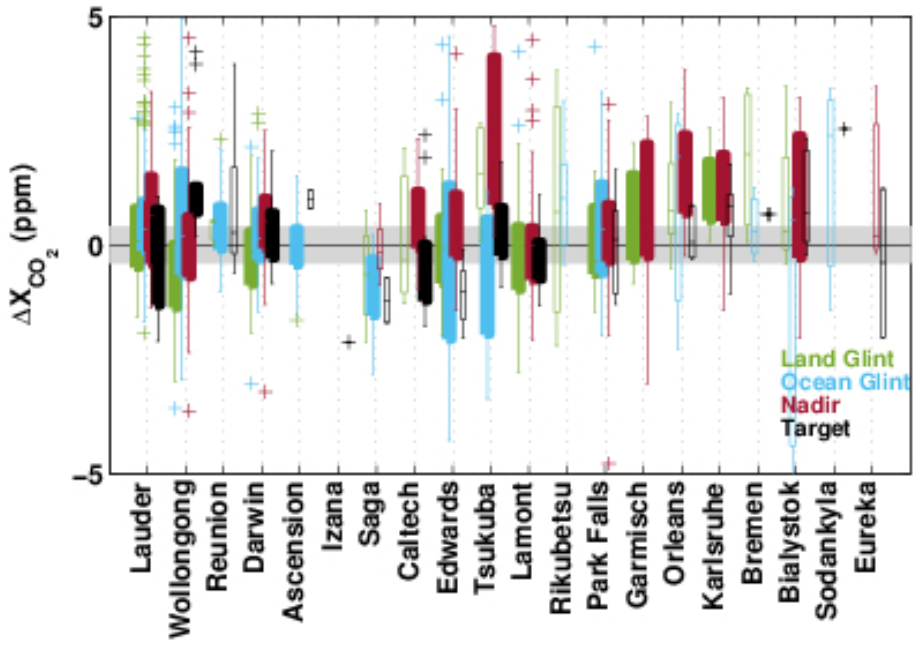
See text in previous response.

L94. Reword. I'd suggest that all TCCON stations are located in areas with spatially varying topography and ground cover. Some of them are characterized by greater variability, some by less. This comes up again later (LL 111-112).

Changed wording to: "There are several TCCON stations that are located in regions with significant spatial variability in topography or ground cover."

L102. Are the apparently X,Y-specific biases made apparent by target model measurements meaningfully indicative of similar biases in nadir or glint mode? Are there sufficient overpasses yet to compile a map of these surface-dependent biases? (Yes, one would need to subttract a time-varying reference like the Mauna Loa CO2 measurements.)

A new figure has been added that shows consistent biases between OCO-2 XCO2 and TCCON XCO2 in the science mode data to those seen in the target mode data:



The surface-dependent biases are much harder to quantify in the data from the science modes.

LL104-105. "Green" and "brown" are not land cover types.

Removed the sentences.

LL120-124. Are there not AirCore measurements at Lamont and Sodankyla?

Yes, there are, but they have not yet been used in OCO-2 comparisons.

L139. When you use the term "selected" referring to target observations, does that mean that these targets were actually observed? Why not use the more accessible term "observed"? Is there some difference between selection and observation?

Changed “selected” to “observed”.

L153. Why is the constant term chosen to be multiplicative, and not an offset?

Added: “The y-intercept is forced through 0 because it is assumed that in the absence of atmospheric CO2, both OCO-2 and TCCON will measure 0 ppm.”

L154. Cite the WMO scale (Zhao and Tans 2006) here.

Done.

L154. I don’t think "constant bias" is a good term to represent a scaling factor.

Changed all instances to “scaling bias”.

L157. Please explain what an OCO-2 footprint is, since the general concept of a satellite footprint may be different.

Added text: “The OCO-2 instrument records 8 spectra simultaneously, each with a slightly different atmospheric path, and hence measures sunlight that has reflected off of a different surface location or “footprint”.”

L160. Please explain why parameter-dependent bias (PDB) needs to be fit sequentially before the constant scaling factor (CSF) is fit. If the residuals are a result of both PDB and CSF, then the natural procedure is to fit both simultaneously. If the PDB is fit first, then some special measure would be needed to allow the CSF to drift freely during that fit. You admit the one-way dependence of CSF on PDB, but do not admit the presumed two-way dependence.

The datasets used to determine the PDB and the CSF are different. The PDBs are determined from the OCO-2 small-areas, the OCO-2 southern hemisphere data, and the multi-model mean-OCO-2 deviations. The CSF is determined from target-mode comparisons of OCO-2 with TCCON. In order to determine the CSF, the PDBs must be applied. This has been clarified in the manuscript.

L175. "Flux inversions" is inaccurate. The term comes from transport inversions for flux estimation–it is transport that is being inverted.

Changed wording to: “Placing the OCO-2 data on the World Meteorological Organization's (WMO) trace-gas standard scale is crucial to obtaining accurate flux estimations that are consistent with the inversions of surface in situ CO2 measurements that are carefully calibrated to the WMO scale (Zhao2006).”

L178. You do not achieve placing the OCO-2 data on the WMO scale by this exercise. You are comparing a set of reference retrievals having their own issues (major comment 2) with another, experimental set. This is a comparison exercise.

Removed the phrase “To achieve this, we compare” and replaced with “We thus compare”.

L180. See major comment 2.

Addressed in response to major comment 2.

L197. The PDB correction may have removed spurious variability in the OCO-2 retrievals, but this is again in comparison with TCCON, which has its own spurious variability (see major comment 2). It is interesting that the PDB exercise improves this comparison of independent estimates, but we cannot truly say that the spurious variability has been removed.

Changed wording to: “The improvement indicates that the parameter-dependent bias correction is effective at removing spurious variability in the OCO-2 data *with respect to TCCON*.”

L200. "Time dependence" is perhaps a bit broad here. Since OCO-2 is in a sunsynchronous orbit, only periods of 1 day and greater are addressed here.

Changed wording to: “The *long term* time-dependence of the difference between the OCO-2 target-mode data and the coincident TCCON data dXCO2, after the scaling bias is removed, is plotted in Fig. 5 [of the revised paper].”

L201. Do not use "constant bias", since this strongly implies an additive offset. The term "constant scaling factor" is more accurate.

Done.

L204. Again, the bias is not constant if you are estimating a multiplicative scaling factor. Reword.

Done.

L204. The best-fit line in panel (b) of Figure 5 is farther away from the 1:1 line than it is in panel (a). Can you explain why the bias correction of OCO-2 estimates apparently increases the global disagreement with TCCON?

Added text: “The parameter dependent corrections can affect the scaling bias if the global mean value of XCO2 at the mean parameter value differs from that of the mean filtered value.”

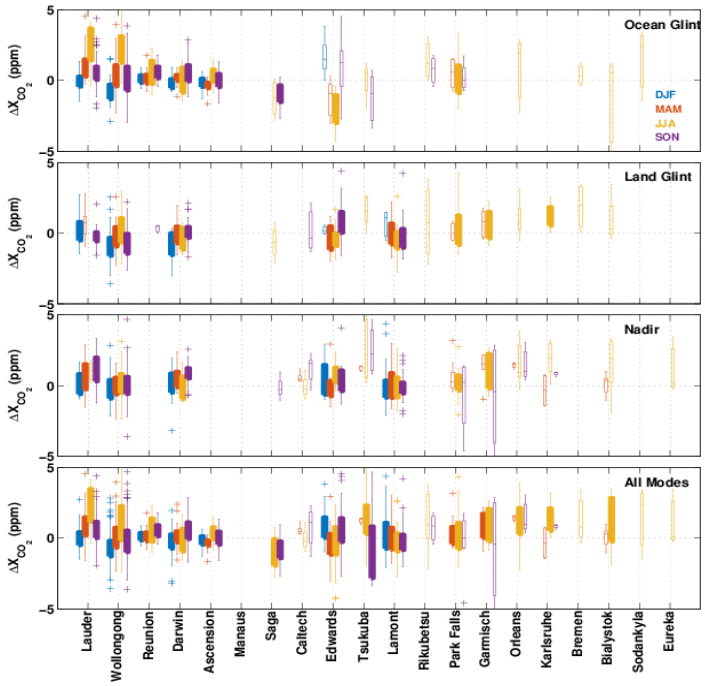
L209. These two broad classes of remaining bias may be obvious, but I think it is too early to say that there aren’t other large biases hiding in the data. For instance: there have been indications that OCO-2 ocean glint retrievals away from the high southern hemisphere fringe are systematically lower than land retrievals at the same latitude. Also: is it not possible that the OCO-2 land glint and land nadir retrievals are subject to the same surface/topography-dependent biases made evident in figures 9-12?

Added “known” and removed “at a target location”: “The *known* residual biases seem to fall into two classes: biases in high southern latitude ocean glint data, and biases caused by the surface properties.”

LL 213-220. This discussion is based on a total of 3 target-mode measurements. That is far too small a number to statistically characterize the potential bias in a robust manner. Due to the small sample size, this evidence is almost anecdotal.

This discussion is based on three target-mode measurements, and confirmed by thousands of nadir and glint-mode measurements. The language has been clarified: “There were three target-mode measurements recorded in the southern hemisphere during that time: two points over Wollongong, and a third point over Reunion recorded during late July/early August 2015 that hint at this feature (revised Fig. 5). The top panel of the figure below clearly illustrates this problem by showing the divergence of the OCO-2 XCO2 measurements in ocean glint

mode over Wollongong and Lauder from the TCCON XCO2 values during June, July, and August."



LL 218-220. How can this bias not affect the relationship between OCO-2 and TCCON? What is this 1:1 line being discussed? Unclear, please fix.

Clarified text: "This bias does not impact the overall scaling bias between OCO-2 and TCCON XCO2 within the uncertainty but does impact the latitudinal gradients (and hence fluxes) inferred by the OCO-2 data."

LL 225-226. Use of "one-to-one plot" is inappropriate. There's no measured-vsretrieved or OCO2-vs-TCCON plot on which a 1:1 line is placed.

The one-to-one plot referred to is Figure 5. Added text to clarify: "Site-dependent differences from the one-to-one plot in Fig 4c, shown in Fig. 6..."

L226. Reword this. TCCON has site-to-site variability, some measure of which is apparently 0.3 ppm. If that estimate is Gaussian, then differences of any magnitude are possible, including some above 0.3 ppm. Furthermore, the number of available comparisons between OCO-2 and TCCON comes in to play here.

Changed text to: "Any differences with magnitudes less than 0.4 ppm could be attributable to TCCON station site-to-site biases (Wunch2010), so we focus on the biases that are significantly larger and thus *most likely* attributable to the OCO-2 data."

L226. My reading of Wunch et al. (2010) suggests an xCO2 error with a mean of 0.4 ppm, and a range of 0.2 to 0.8 ppm. Where does the 0.3 ppm come from? What exactly does it represent, statistically speaking?

The 0.3 ppm should be 0.4 ppm, as listed in the Wunch et al. (2010) paper, and represents an estimate of the

overall calibration uncertainty in TCCON. This is defined as the 1-sigma standard deviation of the TCCON:Aircraft ratios after removing the overall scaling bias. We fixed this throughout the manuscript.

L232. ...in the *OCO-2* xCO2.

Done.

L234. How frequently does this spatially-dependent bias pattern appear? When it does appear is it always similar to the pattern shown in figure 9? Same sign, magnitude, etc.? Are there any hypotheses about why it only appears some of the time? Are the spatially-dependent biases at Wollongong and Lauder similar? Is there evidence for biases like these at other stations? End users of the OCO-2 data need to know more about this.

The feature appears clearly on 10 of 11 clear-sky target-mode measurements over Edwards, with higher values always over the brighter playa. However, the magnitude of the difference across the site differs from target-to-target with RMS ranging from 0.9 ppm to 1.7 ppm across the scene, and the cause of this is currently unknown. Edwards shows the most obvious pattern of all target locations.

L235. ...the *retrieved* xCO2 appears dependent on...

Changed to: “... the *OCO-2* Xco2 appears dependent on...”

LL235-6. These are retrievals of xCO2, not measurements. Two instances, one for OCO-2 and the other for TCCON.

Changed to: “All mean target-mode OCO-2 XCO2 values at Edwards are biased lower than the coincident TCCON XCO2.”

L239. "TCCON measurements" needs to be changed to "estimates" or "retrievals".

Changed to: “The OCO-2 retrievals of XCO2 in target mode are systematically higher than those from the TCCON, and are...”

L243. Change to something like "Even for sites at which OCO-2 retrievals do not have..." Also, while the meaning is clear, the existing text isn’t a sentence (probably meant to drop the "at").

Changed to: “Even for sites at which OCO-2 XCO2 does not appear to have a significant bias with respect to TCCON, the retrievals can show spurious spatially correlated errors.”

L256. The OCO-2 concept of "footprint" might be confusing to the unitiated. Please make reference back to section 3 (L 157) to help.

Added the sentence mentioned above (the reviewer's comment about L157).

L270. See major comment 3.

Changed to: “The median XCO2 within the coincidence box recorded on the same day as the TCCON measurement is compared with the TCCON daily median.”

L277. See major comment 3.

This has been reworded: “The differences between bias-corrected OCO-2 and TCCON XCO2 are all less than 0.4 ppm, and the RMS of the difference is near 1.5 ppm.”

L288. Off-hand reference to "slopes". I infer from context these are the slopes of regressions, but you should be more specific.

Changed to: “...and the slopes of the relationship between OCO-2 and TCCON XCO2 increase...”

L294. Unless you change the methods, this needs to be refined to specify the use of daily (and thus spatial) medians. See major comment 3.

Changed to: “Aggregated OCO-2 XCO2 estimates generally compare well with coincident TCCON data at global scales, with absolute mean biases less than 0.4 ppm and RMS differences less than 1.5 ppm.”

LL 297-298: That the TCCON data are "tied" to the WMO scale is already an abstract concept that needs a careful definition in each document where it is used. This claim that OCO-2 retrievals can be tied to the WMO scale in a doubly-indirect way through TCCON retrievals strains this already delicate claim.

Text was added to an earlier section. See response to Major Comment #2.

L301. The statistical nature of these remaining differences is crucial for use of the xCO2 estimates from OCO-2. Leaving us with "typically less than" is not a satisfying way to quantify the results.

Changed “typically” to “ with RMS differences less than 1.5 ppm”.

Figure 1. Specify that these are OCO-2 measurements.

Done

Figure 2. Can you add to the caption how many individual soundings are represented by this target mode measurement? Also, the inset could be a lot clearer: make the markers smaller, increase the inset size, tighten the lat/lon range...should we see a zig-zag pattern?

There are 3473 soundings in this target-mode measurement. The figure and caption have been updated.

Figure 3. I see orange and red markers, not gold and orange. Either way, the color contrast is pretty low. Recommend changing to colors that are more distinct, e.g. yellow and red.

The colour scale was designed (by MATLAB) to be visually distinct, and I even used different marker styles (stars and circles) to further ensure they were distinct. I'm sure any further attempts I could make would only make this figure worse.

Figure 4. Not an important figure. Most of its content is expressed in words, at line 140. Recommend relegating to appendix.

Removed figure.

FIgure 5. According to the main text, these are statistics of the *daily median* target mode retrievals with the TCCON retrievals. If so, the standing pattern of geographic bias has been filtered out, and would contribute to lowering the RMS difference. Add this detail of data processing to caption. Cf. major comment 3.

I'm not sure what you mean by a *daily median* target mode retrieval. One target mode maneuver only takes a few minutes. It is the median of the XCO2 values within one OCO-2 maneuver (<5 minutes), compared with the TCCON retrievals within an hour of the mean time of that OCO-2 maneuver. These details have been added to the caption.

Figure 5. What do the error bars represent?

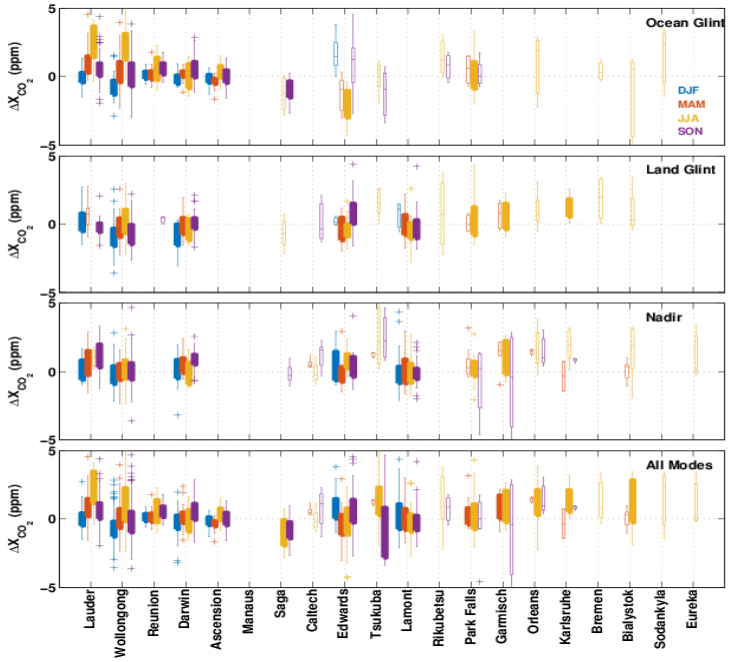
This has been added to the caption: “The error bars represent the standard deviation of the values.”

Figure 6. If I understand correctly, the error quantities displayed here include the "major comment 3" space-time variability of both OCO-2 and TCCON estimates. Is that correct?

These are the sum, in quadrature, of the standard deviation about the OCO-2 median for each target-mode measurement, and the standard deviation about the TCCON median within +/- 30 minutes of each target-mode measurement.

Figure 6. I am concerned with the seasonality of these residuals. I took to drawing lines to connect the estimates for each TCCON station. This analysis is challenging because of gaps in various data records. Could these data be grouped into 3-month climatologies, and displayed by latitude?

The following figure and a short discussion was added to the text to address this concern:



In the above figure, the filled boxes indicate seasons for which there are >10 comparison points between OCO-2 and TCCON; the thin boxes contain at least 3 comparison points. Any site and season for which there were fewer than three comparison points were excluded from the plot. The different colours indicate the different seasons (blue = DJF, orange = MAM, yellow = JJA, purple = SON). The TCCON stations are ordered by latitude, where Lauder is 45S and Eureka is 80N. The equator is between Manaus (3S) and Saga (33N). The high southern

latitude ocean glint bias is clear in the top plot. Otherwise, the main result from this analysis is that the OCO-2 XCO2 appear to have a bias that increases with increasing latitude, most clearly seen in the JJA data north of 45N (Park Falls). A seasonal bias is not clearly apparent, but the data volume is poor, especially in the northern hemisphere north of ~45N. These bias patterns are consistent with those seen in the target mode data (Fig 6 in the revised paper).

Figure 7. My opinion is that this figure should be dropped entirely, especially since it is not discussed meaningfully in the main text. I don't think it really succeeds at comparing meridional gradients between TCCON and OCO-2. This is in part because the OCO-2 quantities displayed here include a full representation of zonal variability, whereas the TCCON estimates are single samples of that zonal variability. This requires the inferential coloring by local fossil fuel emissions.

Figure 7 has been removed.

Figure 7. Given that the hypothesis here is that ocean glint retrievals are biased differently from the land retrievals, I don't think it makes sense to lump all three science modes together here.

Figure 7 has been removed.

Figure 7. Please specify which version of EDGAR is being used here. Those emissions estimates have undergone some significant revisions over time.

Figure 7 has been removed.

Figure 7. If you are going to use medians for the central value of the TCCON distribution, you should probably use 25th & 75th percentiles for the errors bars, not standard deviations.

Figure 7 has been removed.

Figure 8. While it depends on what is represented by this 0.3 ppm number (see comments for L226), this statement that deviations beyond the shaded bar are attributable to OCO-2 issues is statistically vague and probably incorrect. If these are Gaussian distributions, then a deviation of any magnitude is possible. You can make rigorous statements about the likelihood that the two samples are drawn from the same distributions, and this would of course take into account the errors in both TCCON and OCO-2, and the size of the sample.

Changed text to: "Any differences with magnitudes less than 0.4 ppm could be attributable to TCCON station site-to-site biases (Wunch2010), so we focus on the biases that are significantly larger and thus *most likely* attributable to the OCO-2 data."

Figure 8. Why are Sodankyla, Bremen, Izana, and Ascension represented by single markers and not bars? I would guess you need to add some information about the number of samples here (some people use variable bar widths to represent this).

Revised the caption to clarify that these sites only had one target-mode measurement.

Figure 9. Striking figure. The off-hand comment that this spatial bias is not always evident (cf. L234) needs more attention.

Added text main text: "There have been 12 target observations of Edwards, 10 of which had clear skies during the OCO-2 maneuver. On all but one of the clear-sky target maneuver over Edwards, the OCO-2 XCO2 appears

sensitive to surface brightness (revised Fig. 7), with higher XCO2 retrieved over brighter surfaces. However, the magnitude of the sensitivity differs from target to target: the RMS of the target-mode measurements ranges from 0.9 ppm to 1.7 ppm, and the relationship between surface albedo and XCO2 have different slopes (ranging from -2.8 ppm/unit albedo to 10.5 ppm/unit albedo with a mean of 4.5 ppm/unit albedo)."

Figure 10. Is the TCCON station shown with the white star? If so, why are the retrievals all to the northwest (assuming north is up) of the station?

The white star is the location of the TCCON station in Wollongong. The retrievals surround the TCCON station. The elevation map in Figure 10 has a standard digital elevation color scheme which looks similar to the underlying MODIS image.

Figures 10 & 11. It is not clear to me where the coastline is in these figures. It looks like there's ocean on the right-hand side, but what accounts for the near-zero altitude in the upper-left half of the figure? Is that also water?

There is ocean to the east and an escarpment and a dark green surface to the north west. Added the following figure to better orient the reader:

The MODIS true-colour image in this figure is of the Wollongong region of Australia. The solid white line traces the coastline, with the South Pacific to the east. The dashed white box shows the region plotted in the other Wollongong figures (i.e., zoomed in so that the OCO-2 data points are clearer).

Figure 11. Why are the xCO2 differences in a different location than the retrieval altitudes shown in figure 10?

They are not. In Fig 10, all measurements from the various target-mode maneuvers are included; in (revised) Fig

9 individual maneuvers are shown in each subplot. This has been clarified in the caption.

Figure 11. It's hard to distinguish white dots from clouds. Maybe the dots need black borders?

Putting black borders on the white dots make the plot even less readable.

FIgures 13-15. Are these daily medians of the TCCON and colocated OCO-2 retrievals? How many retrievals go into one such point?

Yes, these are TCCON daily medians and co-located OCO-2 retrievals. There is a minimum of 5 TCCON measurements and 5 OCO-2 measurements in the median. There are typically many more TCCON measurements (up to a hundred or two) and OCO-2 measurements (up to a few hundred) included in each daily median.

FIgures 13-15. Given the importance of the annual cycle in atmospheric CO2, and the potential for seasonal issues with sunlight-based estimates like these, I'd like to see the time series of figure A1 not relegated to the appendix and/or the scatter plots of figures 13-15 transformed to show something about the seasonality of the fits. Alternatively, statistics of residuals reported for 3-month averages centered on the annual maximum/minimum and shoulder seasons should be reported.

We added a new figure to address the seasonality:

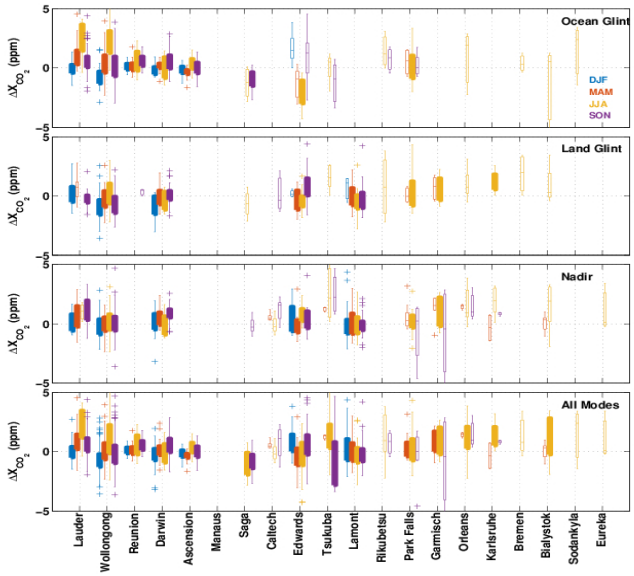


Table 1. Units (specifically, on altitude) need to be specified.

Done.

Table 1. While it might be fun to use musical notation symbols here, the natural and sharp symbols are visually indistinct. Recommend using simple numbers (1,2,3) instead.

Changed.

Table 2. I assume the parameters listed in column 1 are direct variable names either in the lite files or in the full files. Please specify this. It would be meaningful to give a short description of those, especially those that are not clear from the parameter name. For instance, co2_ratio_idp or albedo_weak_co2_fph.

Done.

Table 2. Not clear on the surface type filter. Isn't Cox-Munk for ocean glint retrievals only?

Yes, thanks for noticing. The table entry should say "*Not* Coxmunk"!

Table 3. Many of these entries have N < 10, which brings up the issue of statistics of small sample sizes. Maybe those could be visually noted somehow.

If N>10, it is now marked in bold font.

Table 3. Can you explain why there are ocean glint measurements over a continental site like Park Falls? Are these retrievals over the Great Lakes?

Yes, those are retrievals over the Great Lakes.

Table 3. The "total" line represents what exactly? For instance, is the bias figure weighted by the number N? Does the RMS include this effect?

The "total" line represents all the coincidences aggregated together, each treated independently. This has been added to the caption.

References:

Rayner, P. J., and D. M. O'Brien (2001), The utility of remotely sensed CO 2 concentration data in surface source inversions, Geophys. Res. Lett., 28(1), 175–178, doi:10.1029/2000GL011912.

Miller, C. E. et al. (2007), Precision requirements for space-based XCO2 data, J. Geophys. Res., 112(D10314), 1–19, doi:10.1029/2006JD007659.