Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2016-227-RC1, 2016

Atmospheric Measurement Techniques Discussions

Interactive comment on "Comparisons of the Orbiting Carbon Observatory-2 (OCO-2) X_{CO_2} measurements with TCCON" *by* Debra Wunch et al.

A. Jacobson (Referee)

andy.jacobson@noaa.gov

Received and published: 29 November 2016

This is a well-written paper describing a comparison of xCO2 retrievals from the TCCON network with those from OCO-2. Analyses are reported for three operation modes of the OCO-2 instrument, and comparisons are made for TCCON stations around the globe.

There are some issues with statistics being reported. In one instance (Fig 8, L226) the methods being applied are unclear. With respect to my major comment no. 3, usually but not always a daily median is computed before comparing OCO-2 and TCCON. This biases the analysis, compared to an analysis of sounding-by-sounding differences.

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.

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.

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.

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.

MINOR COMMENTS

L1. OCO-2 does not measure carbon dioxide. See 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.

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.

L29. See major comment 1.

L31. See 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.

L48. Boreal forests certainly contribute a significant, perhaps major, fraction of the

annual cycle in atmospheric CO2, but they are not THE driver.

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

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

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).

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.)

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

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

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?

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

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

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

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

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.

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

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.

L180. See 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.

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

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

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

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?

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?

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.

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.

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

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.

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?

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

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.

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

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

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

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").

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

L270. See major comment 3.

L277. See major comment 3.

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

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.

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.

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.

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

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?

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.

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

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.

Figure 5. What do the error bars represent?

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?

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?

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. Given that the hypothesis here is that ocean glint retrievals are biased dif-

ferently from the land retrievals, I don't think it makes sense to lump all three science modes together here.

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

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 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.

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).

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

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?

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?

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

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

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

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.

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

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.

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.

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

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.

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?

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?

Reviewed by Andy Jacobson, University of Colorado and NOAA Earth System Research Laboratory

---

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-227, 2016.