

Interactive comment on “Characterization of OCO-2 and ACOS-GOSAT biases and errors for CO₂ flux estimates” by Susan S. Kulawik et al.

Anonymous Referee #1

Received and published: 22 January 2020

General Comments:

The manuscript describes an evaluation of biases in remotely sensed CO₂ concentration from GOSAT and OCO-2 using the ACOS retrieval algorithm. It is extremely critical to understand such biases in order to make full use of these remote sensing datasets within flux inversion frameworks. In practice it is very difficult to obtain this information because comparison datasets are much sparser, sample different atmospheric volumes, and have their own biases relative to the unknown truth. As such, I think the focus of the paper is very timely and useful to the research community. As far as I am aware, this is a novel study because it brings together multiple comparison datasets (TCCON, aircraft, and surface flask data) and also attempts to characterize errors and biases in some of the comparison methods themselves, such as sampling

C1

mismatches.

However, the paper itself has numerous problems in presentation and some potential methodological problems. The major specific problems and my suggestions are listed below, grouped into respective sections.

Methodology:

In general, I question whether all the quoted uncertainties should have been rounded to 0.1 ppm. Since most of the uncertainties are < 1.0 ppm, this means most numbers are rounded to a single significant figure. This causes problems in interpretation of many of the results - for example, just the first set of numbers in Table 1, we have the total co-location error for Geometric (CT2017) listed as 0.4 (0.3, 0.2). I think the value of 0.4 is the quadrature sum of 0.3 and 0.2; if we had two significant figures this quadrature sum could actually range from $\sqrt{0.34^2 + 0.24^2} = 0.42$ to $\sqrt{0.26^2 + 0.16^2} = 0.31$. This problem is worse later when two uncertainties are subtracted (Table 2). The authors should report an additional figure, or explain the justification for rounding to one significant figure.

Line 315 - "The LMT reaches the maxima and minima at least one month before XCO₂" I do not think this statement is not supported by figure 5. The error bars on the time series are very large, so the claimed time offset does not appear to be statistically significant. More sophisticated analysis needs to be performed to support this claim. In addition, I would argue this aspect of the data is unrelated to the focus of the paper, so should be omitted.

Section 4.2. Line 482 - To assess the relative variability of LMT versus XCO₂, the authors are "Looking at a few random days" - this is not sufficient analysis, particularly since the data underlying the manuscript is much more extensive. Can you just compare the overall daily standard deviations, averaged over the multiple years of available data? Otherwise the reader is left the impression these 3 days are selected for some omitted reason.

C2

Section 5.3.1. - These simplistic calculations are too simple in my opinion. There is some dimensional mismatch that is not explained: why is a concentration anomaly - the regional bias - directly converted to a flux? The Baker 2006 numbers are in PgC/year, while the anomalies are ppm. I do not see any need for this section at all, since the following section (5.3.2) is a much more realistic representation of how the concentration anomalies would actually impact flux estimates. I would also add detail to section 5.3.2, there are some lacking details. How is the bias correction term assimilated, exactly? How do you explain 'no overall bias resulting from the bias [correction term] assimilation', when the mean of the bias correction term is nonzero?

Presentation:

I would recommend the paper be reviewed by a co-author or colleague, focusing on improving the clarity of the language. In particular, I feel there is a lack of consistency of what is precisely meant with terms like "bias", "error", "correlation", which makes it very difficult for the reader to understand the analysis.

Somewhere in Section 1 or 3 (I am not sure the best place), there need to be equations spelling out how the different error terms are applied to either the satellite data, the validation data, or the differences. There must be some assumed underlying statistical model here; explaining that clearly would make the manuscript much more understandable. This would also make the different terms less ambiguous.

Section 3.4 needs to be reorganized. There is a bulleted list of different uncertainty aspects of the validation process, but each bulleted item has a long description. The last bullet however, just says "this is described in more detail below", which is then just the rest of section 3.4. This could be split into subsections, or just make one short abbreviated list followed by separate paragraphs describing each aspect. The section describing the averaging kernel error is very unclear.

Section 5.2, there is no specific definition of what quantity is being computed, I am not sure what "the square of the XCO₂ error versus time" means - I think this is similar to

C3

the Torres et al 2019 paper, but in that paper they clearly define the statistic they are using first (the semivariance, equation 8 in the paper), and then the analytic model they use (equation 9). What statistic is being used here?

Later in Section 5.2.1, the section describing correlations of bias correction terms versus the XCO₂ is not clear and needs more explanation. In addition, I do not find this part of the manuscript is well aligned to the main focus of the manuscript (characterization of biases relative to validation data), and I would recommend removing it.

Section 5.2.3, this section also needs an unambiguous description of the statistic being used. "Time correlation of the error (OCO-2 minus TCCON) is used to calculate correlation of error versus time." - this is not meaningful.

Another common issue in the presentation, is there are many unneeded parenthetical statements, or unrelated comments. Some examples: Line 155: "It would also make more sense to use a more relaxed co-location ... but we found sufficient co-locations within +/- 1 hour" This entire sentence could be omitted, there is no reason to discuss some processing aspect that was not needed. Line 295: this statement about other co-location techniques is unrelated to what was actually used, it should be removed or moved to the discussion section at the end. Line 392: "... described below (although it is possible that different models or schemes have biases in the same direction)." This is unclear and seems unrelated to the rest of the paragraph.

Figures: these are all very low quality, difficult to read, and not publication quality. Particularly figure 1, where many of the labels are unreadable. If possible, all figures should be recreated at a higher resolution and not compressed (specifically, use PNG not JPG).

Figure 2, and other similar plots: in the lower panels that show difference plots, there is appears to be a dashed line along $y=0$ that looks the same as the CarbonTracker derived difference. Remove the $y=0$ lines (otherwise I cannot tell which is which) or replot the $y=0$ as a dotted or thinner dashed. What do the vertical marks at 2015 and

C4

2017 on the $y=0$ line signify?

Figure 6 - I cannot tell what this represents, are these boxplots, or just scatter plots where all the points are piled on top of each other? Please use a more visually intuitive plot here.

Figure 8 - the linear fits on the left side are not very meaningful. It appears that the fit for the small bin numbers is a linear fit to 3 data points. This is not a robust result. If the authors want to claim a robust difference in slope between the two scale ranges, this needs a more sophisticated analysis, with an assessment of statistical significance between the two slopes.

Figures 9, 10 - Legends need to be plotted with better spacing. Why does the bias term fit in figure 10 have a -NaN value? What are the chi² values listed in the legend? If these are the typical reduced chi² values, why are they all well below 1?

Minor issues:

Line 57 "systematic errors by a factor of 2 for land observations and improves errors by ~ 0.2 ppm for ocean." Describe the magnitude of the errors either multiplicatively or additively for both datasets, instead of mixing the types. The reader cannot assess which data has higher systematic errors with this description.

Line 70: The description of OCO-2 observation modes is not accurate. There are three modes: Nadir, Glint, Target; the data in the "land glint" and "ocean glint" datasets is collected in the same mode. Later in Line 74 the phrase "standard modes" is referenced but not described.

Line 81: Need a reference for the CO₂ variations. Is this XCO₂ or CO₂ concentrations anywhere in the column?

Line 95: the bulleted list is not clear: What "effects" of bias correction are being characterized? this is too vague. The last bullet does not agree with the lead in sentence fragment - a clearer statement might be "the magnitude of false surface fluxes induced

C5

by regional biases assimilated in an inversion model"

Line 110: there needs to be a DOI, or at least a URL, for where this data file was obtained.

Line 142: here "The TCCON averaging kernel is applied" but then in line 145 "the OCO-2 averaging kernel applied" - which is it?

Line 160 - 165, what is 'Lauder125' and 'Tsukuba125' ? are the "125" suffixes meaningful or are these typos?

Line 218 - Please define the "predicted total error from the v8 lite product"? Is this the xco₂ uncertainty from the level 2 algorithm?

Line 220 - what uncertainty is assigned to the average of the 1-second averages?

Line 221 - now this is "85 observations per bin" - but this is 85 observations per 10-second average, correct? This section would be clearer without using the word "bin" for each stage in the averaging process, or perhaps call them "10-second bins" or "1-second bins" as appropriate

Line 244 "The field "xco₂_std+uncertainty" is the predicted error for the averaged product." Is this referring to a specific field inside the v8 Liu dataset? Is this relevant? There is no publicly available dataset in this case, so I am not sure what use this information is to the reader.

Line 288 - Keppel-Aleks 2012 used the potential temperature at 700 hPa, is that what was used here?

Line 478 - "However, recently aircraft measurements were found to have a low bias on the order of -0.6 ppm for LMT ..." reference needed.

Line 670 "The LMT product is most useful when the 2 partial 670 columns behave differently" - This sentence is unclear and vague, what is meant by "behave differently"?

C6

