

## Review of Carbon dioxide retrieval from OCO-2 satellite observations using the RemoTeC algorithm and validation with TCCON measurements by Lianghai Wu.

This paper describes the XCO<sub>2</sub> retrieval from OCO-2 spectra with the RemoTeC algorithm for two years of data, focused on TCCON collocations for validation purposes. OCO-2 currently delivers the most accurate and the largest dataset of NIR / SWIR radiance measurements for XCO<sub>2</sub> estimation. RemoTeC is widely acknowledged as a state-of-the-art retrieval algorithm, already successfully applied to GOSAT. The retrieval of OCO-2 with RemoTeC is therefore widely expected and this work largely deserves a dedicated publication.

The paper is well written and concise. The main properties of OCO-2 are clearly reminded. The main assumptions of RemoTeC are recalled, but I understand that a detailed description of the algorithm requires to read references, which may be a weakness for the self consistency of the paper. Maybe the paper should be more precise about the modification of the algorithm for OCO-2. The methodology is based on the systematic comparison with several TCCON stations. This is a classical, rigorous and probably the most accurate strategy for XCO<sub>2</sub> missions, as the column sensitivities are similar and strong efforts have been made to trace the TCCON network to the WMO XCO<sub>2</sub> standard. Such validation work requires estimation of random error, global and regional biases, which can only be obtained at a reasonable cost with the large data set of the TCCON network. The choice of a period larger than 1 year is essential to remove the seasonal effects.

The main result is that the residuals biases of OCO-2 / RemoTeC with the TCCON global network is lower than 0.1 ppm in absolute value, and up to 1 ppm when looking at individual stations. These low values, of the same order as the OCO-2 L2, prove the quality of RemoTeC and its application to OCO-2. The remaining station to station biases are still high for the needs of the flux community, meaning that research must continue to improve the retrieval scheme and the understanding of the instrument (beyond the scope of this paper). The bias correction shows its efficiency to empirically reduce the biases, but the magnitude of the correction is still too high to give a solid confidence in the final value.

I have some questions and remarks that I would like the authors to address before publication. These will probably not require new calculations, but only precisions and additional materials. I will try to focus my questions on the application of RemoTeC to OCO-2 and not to the RemoTeC algorithm itself which was the subject of previous papers properly quoted. One of the drawbacks of this paper is that literature on this topic is large, and the reasons of some assumptions have now become implicit (and could sometimes deserve to be questioned once again). I also noticed that several results are given only in the text whereas they should be given dedicated figures or tables (see comments). This has to be corrected before publication. Finally, I was sometimes lost in the different statistics indicator (target, land, ocean; all footprints or daily averages; global bias, station to station bias, standard deviation). A clearer presentation and interpretation of them would be welcome before publication.

### MAJOR COMMENTS

- Page 3 line 13: The objective of the study requires further justification than « to enhance the reliability and confidence of the data product ». What does this study aim at? To challenge the official OCO-2 Level 2 (L2)? To improve RemoTeC through its application to the new OCO-2 dataset, more accurate than GOSAT? Will a new OCO-2 / RemoTeC be proposed in the future?
- As already mentioned, there is a lack of description of the algorithm, largely given by references. This is however very important to understand the differences with the OCO-2 L2.
- p4 l12: 5° around a TCCON station is very large (~500km). In such an area the CO<sub>2</sub> may not be considered as uniform. What is your justification? Did you make any error budget, any sensitivity study?
- p4 l16: Why do you restrict VZA to <30° and not to a larger value? Is there also a restriction on SZA? Table 1 gives some information but in contradiction for VZA, maybe because it only applies to land and glint? Please precise.
- p5 & 6: I think the paper deserves a table describing exactly the content of the state vector.
- p6 l13: the description of the cloud screening is too light, I understand it is a copy of what is done by OCO-2. Do you use the information from the OCO-2 pre-processing, or did you develop your own algorithm? Did you make any performance study, and associated XCO<sub>2</sub> sensitivity study? 30% is higher than the performance reached by OCO-2 (for land and ocean).
- p6 l28: please explain the reason why you separate land and ocean evaluation. Is it based only the aerosol argument (p7 l19)? Was it decided from the OCO-2 feedback ? OCO-2 does also but with another separation between land nadir and land glint.
- p6 l33: the assumption that TCCON station to station variability is zero is very strong and may not be excluded when interpreting the results.

- p7 l8: you talk about retrieval uncertainties; these uncertainties may be instrument dependent. Compared to Butz et al 2011, Gueret et al 2013b, did you reconsider your filters for OCO-2?
- p7 l25: I don't understand why you say « we look for possible correlations of errors with instrumental, geophysical, meteorological and retrieved parameters ». Actually, here you do not look for such correlations (as would the OCO-2 Bias Correction do), you only calculate a regression with  $\chi^2$ , which is different. This is an original bias correction and, as far as I know, it is the first time it is applied. What made you adopt such methodology? To my mind its drawbacks are that you lose interesting spectral information about the residuals. An error in retrieved albedo may lead to a large  $\chi^2$  whereas it has very limited impact on XCO<sub>2</sub>. An error in line-mixing may lead to a small  $\chi^2$  residual but have a strong impact on XCO<sub>2</sub>. I clearly do not say the approach is wrong, but I think that it is new and should really deserve deep study. The shape of the  $\chi^2$  spectra would deserve attempts of interpretation. Why would you have only to regress with  $\chi^2$  in SWIR-1 and not in the other bands? Why would this bias correction be required only for land, not for oceans? The spectroscopy and the instrument are the same. You say in section 4.1 the aerosol contribution is weak in ocean glint measurement, that could be an explanation but aerosols are not the only source of bias.
- p22: figure 4 should exhibit a fit, be given for lands and oceans, and for the 3 spectral bands.
- p8 l9: you say some correlation with parameters of table 1 are reduced, you clearly have to present these correlations by a figure or table before and after bias correction. Otherwise we cannot accept such affirmation.
- p8 l19: please show fig 2 before and after bias correction. Giving a rough value in the text ( $\sim 0.1$  ppm) is not enough.
- p8 l23: please define your averaging. I understand that in fig 2 (no averaging), you plot every OCO-2 single footprint minus the TCCON of the area at the same time. I understand that in fig 5, 6, 7 you average every OCO-2 single footprint – TCCON at the same in a window  $5^\circ \times 5^\circ \times 2h$ , is that true? Is it  $\text{mean}(\text{OCO-2}) - \text{mean}(\text{TCCON})$ ? You mention a « daily averaging » p8 l32, but this term is confusing because you may encounter several collocations with several TCCON during the same day. In such a case, are the data of different TCCON in the same average? If not, you should maybe talk about « overpass averaging »?
- p8 l23: Please explain why you make an averaging in 4.2 whereas you do not in 4.1. I guess that in 4.1 you need to keep the individual parameters for your bias correction derivation, but this clearly needs to be explained. For this reason, fig 2 and fig 5,6,7 cannot be directly compared, and that is why the effect of bias correction is difficult to assess.
- p9 l1 and fig 5,6,7: Please also give the figures before bias correction for fig 5,6,7, and give the associated standard deviation as for p8 l33.
- p9 l10: please illustrate the effect of bias correction by giving figure 8 also for before bias correction.
- p9: I know you made the assumption the TCCON stations are consistent (p6 l34), but I am disappointed that you do not try to interpret the station to station variability in terms of residual bias of OCO-2 / RemoTeC or actual differences between TCCON stations. The fact that you do not see the same station to station bias in land and oceans modes could suggest there are still biases in OCO-2 / RemoTeC.
- p10 l6: Did you try also the same retrieval as fig 6 without the fit of the offset in the O<sub>2</sub> band? As you mention later, there could be a link with the stronger internal reflections in this band.
- p10 l12: I am not convinced by the explanation of the lack of SIF fitting, since the behaviour of the SIF and the offset is very different (SIF exhibits atmospheric absorptions).
- p10 section 5: I think the comparison with the previous OCO-2 / ACOS – TCCON and with the previous GOSAT / RemoTeC – TCCON are very important. But here the discussion is poor, mentioning only the common results in terms of standard deviation. This section really deserves to compare biases (global, station to station, etc.), as it was the case for section 4. This could help understand the origin of biases (from TCCON, from the instrument, the retrieval code). This should be done before and after bias correction.

#### MINOR COMMENTS

- p2 l9: for clarity I would make a new paragraph.
- p4 l3: please precise if you use the tabulated instrumental functions given in the OCO-2 products.
- p4 l12: you use a requirement in degrees rather than in km, why? This makes a distance criterion in km variable according to the station, which is not suitable.
- p4 l18: please precise if you make your own calculation of surface pressure from ECMWF and MNT (this information could also be read from the OCO-2 L2 data). How do you interpolate ECMWF and how to select SRTM grid points? This question makes sense since you do not retrieve surface pressure in your state vector.
- p4 l23: Your initial guess for CO<sub>2</sub> and CH<sub>4</sub> comes from different years, therefore it is subject to inter-annual variability. What is the sensitivity of your retrieval to this first guess?

- p6 l24: figure 1 is given as an illustration but is not interpreted. It is very quickly mentioned on p9 l22, but not comparable. Please emphasize the scientific value of the figure, or discard it.
- p6 l6: by « second order spectral dependence of the Lambertian surface albedo », do you mean you retrieve 3 coefficients of a polynomial describing the albedo? Please precise.
- p6 l31: the « SRA » clearly deserves a mathematical definition, and not just a reference to Dils et al.
- p7 l1: « however » is inappropriate (before you discussed station to station bias, after you discuss year to year variability and global uncertainty).
- p7 l3: the interest of a validation with TCCON deserves a stronger argumentation (see my introduction).
- p7 l24: For clarity I would cut here and create a new paragraph: one for « Filters », one for « Bias correction », since these are 2 different topics.
- p7 l15: how do you define exactly  $\Delta_i$  ?
- p8 l7: is  $\chi^2$  the  $\chi^2$  of the SWIR-1 band, as suggested by the above text? Please precise in the formula.
- p22: legend should precise it is land.
- p8 l18: avoid the terms « precisions » and « accuracy » unless you unambiguously define them. Prefer bias and random error.
- p9 l4 and figure 8,9,10: please give the number of collocation per station.
- p21: fig 2 has 3 panels, fig 5,6,7 could also be the 3 panels of a single figure. Please harmonize.
- p10 l9: please explain why you compare the slope of retrieved offset with that of noise. An offset may be related to the signal (for example straylight) or not (for example wrong dark current estimation).
- p10 l11: why do you say the offset in the O2 band shows a less strong dependence on the signal? The slope is the same as in the other bands (there is only a difference in the noise slope).
- fig8,9,10: giving the actual value of bias and standard deviation for each station would be of great value.
- p11 l14: « a posteriori » is a bit misleading, since at least SZA, VZA and sev are probably pre-filtered.
- p11 l22: you compare target and lands without reminding the values. I guess you talk about station to station variability given by figures 8 and 9 (0.35 for target and 0.41 for land), and not the global biases given by figures 5 and 6 (-0.07 for target and 0.00 for land). This should be recalled.