Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2020-453-AC2, 2021

Atmospheric Measurement Techniques Discussions

Interactive comment on "XCO_2 retrieval for GOSAT and GOSAT-2 based on the FOCAL algorithm" *by* Stefan Noël et al.

Stefan Noël et al.

stefan.noel@iup.physik.uni-bremen.de

Received and published: 9 March 2021

Reply to referee 1

We thank the referee for the review and the constructive comments. They will be considered in the revised version of the paper. In the following, the original reviewer comments are given in *italics*, our answer in normal font and the proposed updated text for the revised version of the manuscript in **bold** font.

Answers to General comments:

1. *Overall, I found this paper useful and interesting, and will serve as an important reference. The subject matter is important, the layout of the paper is logical, the reasoning sound, and the results are generally laid out well. However, there are a number of problems that need to be addressed. While details of the retrieval, filtering, and bias correction were presented in a straightforward way, it was quite dry with little learned. Especially in the part about the random forest filter, which was used for both filtering and bias correction, but with little attempt on the part of the authors to explain the relevance of the features identified. The same goes for the prefilters, where it appeared that thresholds were drawn somewhat out of thin air for some of the parameters. It would have been useful if the authors had shown even a couple example plots of some of the prefilters and how thresholds were determined.*

   Regarding random forest filtering, the relevance of the different features is determined by the random forest method and shown in Figs. 7 and 8. A physical explanation of these relevances is not possible – this is a general problem of this method. Therefore we think that plotting individual maps of parameters is not helpful here.

   There is indeed no well defined procedure to define the prefilters, this is essentially based on "scientific knowledge", e.g. trying to set (in a physical sense) reasonable limits for some parameters. Therefore – as written in the paper – the choice of the XCO2 error limit or the maximum optical depths of the scattering layer is somewhat arbitrary and essentially based on trying out different values and looking at the resulting scatter of the data after filtering. It is therefore not possible to show e.g. a single example map of one of these parameters from which the limits can be derived. or justified.

   We are aware of this problem, and for future product versions we will try to improve the filtering method.

2. *There were 25 figures in this paper, and in my opinion, many more than are useful, especially some of the earlier plots. I suggest the authors try to remove some panels in some plots, or some plots altogether, to show representative plots. For instance, all the noise model coefficients are given in Tables 6 & 7. Therefore, the authors can reduce Figs 3-6 to probably a single 2 or 4 panel plot (e.g., Fit Windows 2 & 3 for both GOSAT and GOSAT-2, P-polarization only). The same goes for Figs 9-12 (a single one would do) and Figs 13-16 (again, a single one would do, and not all bands are necessary). Plots are in the paper to explain findings, not to exhaustively present ever detail of the study, especially if some plots or features of plots are never discussed in the main body of the paper.*

   We agree and will combine / remove plots to reduce the number of plots accordingly. For the noise model we will show windows 2P (O2 band) and 6P (strong CO2 band) for both GOSAT and GOSAT-2 in one figure. Spectra and residuals will be combined into one plot, and we will only show P polarisation, so there will be one plot for GOSAT and one for GOSAT-2.

3. *Finally, it appeared that many important previous works by other authors are never referenced, or included in the reference section but never cited in the main body. In general, referencing needs to be much improved in this work.*

   We will include additional references as suggested by the referee and also check existing references.

4. *Therefore, I recommend publication of this manuscript after a major revision to fix the issues with the burdensome # of plots and problems with referencing, as well as addressing all the specific concerns raised below.*

   See below for our answers to the specific comments.

Answers to Specific comments:

1. *Section 2.3: This is a unique approach to a truth database to my knowledge – it needs more information (plots, etc) on how big this contiguous regions are / how much the TCCON data are expanded through this approach. A map of a month or a season of data density would fulfill this, and I think be very interesting for readers. Otherwise, it's not clear how much this really expands over just using TCCON directly.*

   We will add an example map of the true database.

2. *Section 2.3: Secondly, you say the requirement for contiguous regions, but you never say how close the ak-corrected CT value at the TCCON location & time has to agree with TCCON itself. Is that also 0.75 ppm? You imply this but never say – please correct this.*

   The contents of the "true" database is not selected based on a 0.75 ppm maximum difference to TCCON data but to a subset of CT data at TCCON locations. This subset has been selected as having a maximum difference of 0.75 ppm between the ak-corrected CT value at the location of the station and the TCCON value.

   The "true" database therefore contains only CarbonTracker data, which are confirmed by TCCON measurements but may differ by up to 1.5 ppm from the TCCON value. This is explicitly stated in the manuscript at the end of this section:

   "Please note that the "true" database does not contain any TCCON data - it only contains CT data which were confirmed by TCCON, but individual values may differ by up to 1.5 ppm."

   We will clarify this in the text.

   We also decided to rename the "true database" to "reference database" to indicate that the content is not necessarily the (essentially unknown) "true" XCO2 at a certain time and place but only an estimate which should on average reproduce large scale features correctly.

3. *Section 3.1: Your terms "cloud albedo" and "water vapor path" are neither. These terms already have definitions in use by the community, and they are not how you define them. I suggest you rename "cloud albedo" to "effective albedo" or "effective scene albedo". Note you will screen out some bright desert scenes with your albedo filter, though probably not many. It looks like your 1.98 μm filter is doing most of the work. Regarding "water vapour path", it's nothing of the sort. It's more like an SNRwv(wv="water vapour"), or SNR 1.93 (since this band is roughly at 1.93 μm). Low SNRwv= clear, high SNRwv= cirrus present. So please rename it to something else.*

   We agree that the nomenclature we used here is misleading. We will therefore replace "cloud albedo" by "effective albedo" and will rename "water vapour path" to "water vapour filter" to clarify this. The albedo filter may indeed remove scenes with very high surface albedo, but this seems to be uncritical in our case as we still have data over deserts after cloud filtering.

4. *Section 3.2 – Please MOTIVATE why you use both polarisations separately. Do you believe you obtain more information than if you averaged them together, or do you believe you cannot accurately average because certain instrument properties (such as ILS) are different for the two polarisations, and they themselves cannot be averaged together?*

   We use both polarisation corrections mainly for the following reasons:

   - In principle, information is lost when averaging S and P spectra.
   - In general, the sensitivity of the instruments and therefore the calibration of the measured spectra is different for S and P. For example, as mentioned by the referee, the measured ILS is given independently for S and P.

   - S and P include different information on scattering, which can also be used for filtering and/or bias correction.

   We will mention this in the paper.

5. *Section 3.2.1 Near line 263, you talk about the "NIR", but early in the paper you refer to ALL the bands you use as "SWIR". I realize most scientists label the O2A band as NIR and everything past 1 micron as SWIR. Can you please go through the paper and ensure consistency between NIR and SWIR labels throughout?*

   We will harmonise the use of "NIR" and "SWIR" in the revised version of the paper.

6. *Section 3.2.1 – Way too many plots, as I said in the general comments. As a rule of thumb, try not to overwhelm readers with a bunch of plots that all look essentially the same. Each panel of each plot should contribute to the story you are telling.*

   We agree and will remove / combine figures as suggested.

7. *Section 3.3.1 – In general, your "basic filter" through the RSR filters (I'm looking at your figures 1-2 for this information) really does seem basic for GOSAT, as it filters out only 8 percent of the data (35.0%–>27.2%), and most of that comes from convergence. However, for GOSAT-2 not only do twice as many soundings fail to converge as for GOSAT, but the window 5 RSR also accounts for many failed soundings (5% for GOSAT-2, versus 0.3% for GOSAT, if I am counting right). Can you please comment on why this may be happening for GOSAT-2? Window 5 is the methane band I think. You may wish to split things out separately as land versus ocean – you may find very different behaviors for the two categories. In any event, please devote a few words in this section as to why this is happening. And please do say how differently the filters act on land vs. ocean. Actually, looking at this further, I think it is the "broadband oscillation" in the fit residuals*

you mention for GOSAT-2 that may be causing the problem. Are those oscillations really correlated with retrieved XCO2 quality? If not, you may wish to loosen that constraint for GOSAT-2, to save more soundings.

We use the term “basic filter” for all post-processing filters applied before the random forest filter. Indeed, the number of data removed by the RSR filter is much larger for GOSAT-2 than for GOSAT. We assume that this is mainly related to the GOSAT-2 calibration which e.g. currently does not consider remaining polarisation sensitivities to the U (45°) component of incoming light, which especially affects band 2 and may be the reasons for the broadband features in window 5.

Furthermore, as explained in the paper, there seem to be some problems with the noise of GOSAT-2 data over water, which is why we base our GOSAT-2 noise model on land data only. This is also why we cannot provide separate RSR filters for land and ocean for GOSAT-2. For GOSAT different RSR filters are not necessary as there is no discrepancy in the land/ocean noise data. This is why we apply the same convergence / RSR filters for land and water (but later split the filtering and bias correction).

We rely on the calibrated spectra given in the data product and thus cannot quantify the real impact of the deficiencies in the calibration on the retrieval results. We therefore prefer not to add speculations about possible reasons to the paper.

However, the different performance of the filters indicates that the filtering for GOSAT-2 needs some further optimisation, which is planned for the next version of the products.

We will mention the latter in the paper.

8. *Secton 3.3.2 Near line 295, please also reference Mandrake et al (2013, AMT, “Semi-autonomous sounding selection for OCO-2”), who did something similar for OCO-2.*

   We will add this reference.

9. *Section 3.3.2, near line 310. I’m nearly certain that for water, SAA, VAA, SZA, VZA will be correlated with latitude. Because the orbit is sun-synchronous and you’re looking to the glint spot over water, I’m willing to bet that any machine learning algorithm or even a simple correlation analysis can probably figure out where you are based on those quantities (or even only one or two of them). I suggest you be exceedingly careful in including those quantities. Please include a comment to this affect in the paper.*

   We agree that for water / glint there are specific relations between viewing geometry and geolocation and will mention this in the paper.

10. *Section 3.3.2 – can you state how many training soundings total there were for GOSAT and GOSAT-2, for each of land and water? I wonder if your training set is general enough to avoid over-fitting. Also, please define “Relevance” as you use it in Figure 7 & 8.*

    The number of soundings used for the random forest filter are:

    - GOSAT land: 54317
    - GOSAT water: 109414
    - GOSAT-2 land: 10625
    - GOSAT-2 water: 40459

    From these, we use 90% for training and 10% for testing, so this should be sufficient for 10 parameters. We think that over-fitting is not a problem here, because as mentioned in the paper we only get accurate filter results in two thirds of the cases (for both training and test data sets). This would not be the case for over-fitting where the training set would have a much better performance than the test data set.

“Relevance” is a quantity coming out of the random forest method which describes the relative importance of each feature for the filtering. Relevances are normalised such that the sum of all relevances is 1.

We will include this information in the paper.

11. *Section 3.3.3 – This community did XCO2 bias correction long before OCO-2. Can you please reference earlier works on the subject? (The earliest I know of is Wunch et al., 2011, ACP “A method for evaluating bias...”; I believe there are similar references for GOSAT for the UoL retrieval, the NIES retrieval, and the RemoTeC retrieval). Are you really using 10 parameters in your bias correction? This is way more than most groups usually use (which is typically 1-4; as I remember, Reuter et al.(2017) didn’t use any in their OCO-2/FOCAL paper). Be careful – there could almost certainly be overfitting here. So my comment is 10 parameters simply doesn’t seem to be justified based on past experience and the published methods of nearly all other retrievals for the last 10 years. Therefore, your using 10 parameters requires more justification than simply “this is what came out of the random forest algorithm”.*

   We will add additional references to other bias correction methods.

   The choice of filter parameters is mainly based on the output of the random forest method, which associates to all possible parameters a relevance value (shown e.g. in the top of Fig. 7 for GOSAT). As explained above, this relevance value describes the relative importance of a suggested feature for the filtering or regression. As can be seen from this figure, the relevance drops off quite rapidly after a few variables. The number of variables to be used is then a trade-off between many variables (explaining all relations with a risk for over-fitting and high computational effort) and few variables (no over-fitting, but maybe missing some relations). We decided to use 10 variables as a good compromise, and since we do not seem to have problems with over-fitting (see answer to previous comment) this seems justified. Note that even if we would include too many non-relevant pa-

rameters, this would be no major problem as most filtering is done by the relevant parameters.

   We will include this explanation in the paper.

12. *Section 4, nearly line 378. Just a comment. The higher XCO2 variability over land has long been seen. I highly doubt this is due solely to surface variability. I think it is also caused by different scattering pathways that are not present over water. In particular, photons scattered downward by the atmosphere can be reflected off the surface back into the beam accepted by the sensor; this mechanism doesn’t happen over water, so there are more ways for atmospheric scattering to degrade a retrieval. But that’s mainly just a hypothesis.*

   Thank you for the information. We agree that surface variability may not be the only reason for the higher XCO2 variability over land. We will adapt the text accordingly and take your comment into account.

13. *Section 5 – page 13. Please include appropriate references for each of these algorithms here. Also you say for the validation of the GOSAT and GOSAT-2 FOCAL products, but really these comparisons to the other products are just for GOSAT only. You may wish to state upfront here that the vast majority of the presented validation is only for GOSAT. Only subsection 5.3 mentions GOSAT-2, and it only appears in a single validation figure (25). In fact, this paper is really begging for some basic comparison plots of GOSAT and GOSAT-2 to TCCON, to see how well your algorithm works on GOSAT-2 as compared to GOSAT. Can you please add something to that effect?*

   Detailed references for all products are already given in Section 2.4. Indeed, most comparisons are done for GOSAT, except for the time series (section 5.3).

   We did not include a comparison of GOSAT-2 with TCCON because of the limited amount of the GOSAT-2 data. We will add this in the revised version but note also that the results are probably not representative.

14. *Line 438 – I do not understand this statement about a "bias anomaly". Please be more clear about what you did here. Did you subtract some kind of mean bias with respect to TCCON from each algorithm? Please don't! Or if you did, you have to state somewhere what number you subtracted off each algorithm. If ACOS is high by 1 ppm relative to TCCON and you simply subtracted that off before making plots, it's critical to state that somewhere. It would be much better simply to NOT subtract off that bias, unless you can throughly justify why you did.*

    We subtracted for each algorithm the mean of the bias for all stations. This is because for most applications this mean bias is not relevant since most information is contained in gradients. Subtracting a mean bias also facilitates a comparison of different bias patterns between the algorithms.

    The subtracted mean station bias is actually small (0.17 – 0.64 ppm). We will mention this in the paper.

15. *Section 5.2 end, L445 – Even if you don't have "sufficient data" for full seasonal cycle fits for all GOSAT-2 data vs. TCCON, you've got enough to make some basic plots. Please do so – the community is really interested in them. If not, there isn't a lot of point in including GOSAT-2 in this paper at all.*

    As mentioned above, we will include some plots for GOSAT-2 vs. TCCON.

16. *References: It looks like you have way more references in the Reference section of the paper, than you actually reference in the main body of the paper. A rule of papers: you MUST cite each reference in your references section somewhere in the main body of the paper. Please make sure this is the case.*

    We will check the references, but actually all entries in the References section are cited in the paper. However, the references for the used TCCON data are only listed in Table 1, maybe this is why the referee misses them in the main text.

Answers to Technical/Grammatical comments:

1. *47: Tansat, GOSAT, and OCO-2/3 instruments The Tansat, GOSAT, and OCO-2/3 instruments*

    Will be corrected.

2. *L280: XCO2 error is ambiguous. Suggest you change this to "XCO2 posterior uncertainty" or something more clear that it is the posterior error estimate from the OE itself, and not some error as compared to TCCON or something.*

    Agreed, will be changed.

3. *L292: Remove the word "exemplary". This isn't really an example, I assume this is a full indication of what is happening.*

    Indeed, "exemplary" may not be the correct term here. However, the figures only show an example for one month of data (as written in the caption). We will rephrase this sentence:

    Figs. 1 and 2 show how many data points are typically filtered out in this step.

    Note: The numbering of figures above refers to the current paper, it will change in the revised version.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2020-453, 2020.