Reply-to-reviewer 1

Manuscript submitted to Atmospheric Measurement Techniques

Title “Using ocean-glint scattered sunlight as a diagnostic tool for satellite remote sensing of greenhouse gases”

Authors: A. Butz, S. Guerlet, O. Hasekamp, A. Kuze, H. Suto

We thank the reviewer for the positive view on our work and thoughtful comments. Find our reply below. Line and page numbers refer to the AMTD version of the manuscript. Reviewer comments are reproduced in bold face italic letters.

<<This manuscript presents an analysis of roughly 3 years of GOSAT ocean-glint observations. A method is presented to identify very clear soundings that are free from light-path modifications in the O_2A band. This uses the so-called “upper-edge” method, which calculates a ratio between a retrieved O_2 column under a non-scattering assumption, to that of the predicted O_2 column based on meteorological reanalysis from ECMWF. The authors demonstrate that using the method allows them to see some changes in the GOSAT instrument over time (at least in the O_2 A band), as well as evaluate the inter-consistency of CO_2 retrievals from different spectral windows.

The paper is well-written and should be published in AMT after addressing the following list of minor questions and issues.>>

<<Specific Comments & Questions>>

<<Line 155. How does selection of the R-branch of the O2 A band enhance sensitivity? Is this due to instrument problems that are worse in the P-branch? It seems that if you only use the R-branch, you make your results more temperature-dependent (because the temperature jacobians are opposite in the P vs. R branches). Also, please state if you fit anything related to temperature, or if you simply fix the T-profile to the prior meteorology.>>

The R-branch of the O_2A-band contains optically thicker absorption lines than the P-branch. Tentatively, the optically thicker the absorption lines the more single-scattering (light-path shortening) effects dominate over multiple-scattering (mostly light-path enhancing) effects. So, the argument is that using the R-branch alone renders the method less susceptible to contamination by double-layer (multiple-scattering) effects discussed further in the manuscript. In order to make this point clearer we change the wording at

l.25, p.4375: “scattering effects” -> “single-scattering effects.”

We did not test sensitivity to errors due to erroneous temperature input for this particular study. However, we tested the effect of fitting/not fitting an offset to the temperature profile for our standard RemoTeC retrievals and concluded that there is no benefit in fitting temperature. Therefore, we fix the temperature profile to the one provided by the ECMWF ERA interim analysis (interpolated to the place and time of the GOSAT sounding). We add this information at

l.26, p.4376: “Wind speed is not retrieved but interpolated from the meteorological fields.” -> “Surface wind speed, temperature and pressure profiles are not retrieved but interpolated from the meteorological fields.”

<< Line 195: How sensitive are your results (in terms of selecting the upper edge ensemble) to the fitting or not of the O_2A band offset?>>

The `upper edge' is sensitive to fitting/not fitting the radiance offset in the $O_2$A band. If we do not fit the offset and, thereby, leave a known instrument deficiency unaccounted for, the `upper edge' is less compact and shows seasonal variability.

If the offset is not fitted, the soundings identified as `upper edge' correspond to cases where the offset is found negative by the standard fit. This is in line with our understanding of the retrieval: if there is a negative offset, O_2 absorption lines are (relatively) deeper than expected. Thus, a retrieval that does not fit the offset attributes deeper absorption to higher O_2 concentration. The highest O_2 concentrations are then selected as `upper edge'.

We would consider fitting the radiance offset the standard approach to cope with non-linearity of the TANSO-FTS band-1 detector electronics. If an offset is not fitted, other measures have to be taken to remedy the effect of non-linearity on the spectra.

<<Line 190-215: It seems that the retrieval you describe is completely independent among the different windows. Ie., there are no parameters that have non-zero jacobians in more than one fit window. If this is the case, please state it in the manuscript, just to add clarity to what you've done.>>

This is correct. We add the information at

l.2, p.4377: "This implies that the derivatives of the forward model with respect to the retrieval parameters are uncoupled among the retrieval windows ie. each retrieval parameter corresponds to non-zero Jacobians in exactly one fit window."

<<Line 311-352: This is extremely interesting! For GOSAT data, can you state what additional amount or fraction of data are screened when you apply the 0.05 albedo criterion? It would be interesting to know how many of these double-layer cases there appear to be.>>

A similar question has been raised by the other reviewer. The double-layer filter rejects 4.3% of the upper-edge data (when applied after all the other filters). We add this information at

l.11, p.4381: "Applying the double-layer filter developed in section 3 as a final screening step, it rejects 4.3% of the 'upper edge' data."

<<Line 365/Fig 4: Figure 4 implies that the distribution of the O_2 ratio only depends on time, but not on space. Did you examine if there is any kind of spatial dependence (if you aggregate over reasonably large regions)? Please explicitly state in the manuscript if/that you assume this, and if you've seen any evidence of any kind of spatial/latitudinal dependence.>>

Indeed, we assume that for the aspects discussed in the manuscript time dependent effects drive the pattern. In our opinion, this is justified for the rather short-term effects such as the version/operation switches in August 2010 and April 2011 (Fig. 5).

However, since GOSAT's ocean-glint observation pattern varies seasonally with denser coverage in the Northern hemisphere for boreal summer and denser coverage in the Southern hemisphere for boreal winter, temporal and spatial dependencies are per se entangled, at least on seasonal timescales. When restricting the selection of the `upper edge' to narrower latitude bands and examining correlations with viewing geometry, we observe some seasonal dependencies. These findings are, however, preliminary and inconclusive, so far. We would prefer not to discuss these aspects in the manuscript but to defer it to forthcoming studies.

We add this information at

l.27, p.4382: "So far, the discussion assumes that the `upper edge' is only affected by time dependent effects. However, GOSAT's ocean-glint observation pattern varies seasonally with denser coverage in the Northern hemisphere during boreal summer and denser coverage in the Southern hemisphere during boreal winter. Therefore, temporal and spatial variability are entangled. Seasonal variability of the `upper edge', for example, could be caused by a truly seasonal pattern or by a latitudinal pattern that appears time dependent due to the seasonal coverage. We observe some seasonal variation when restricting the selection of the `upper edge' ensemble to narrower latitude bands, though care must be taken to avoid persistently cloud covered regions. We will investigate dependencies on latitudes and other parameters such as viewing geometry in forthcoming studies. Short-term effects such as changes in level 1 version and ocean-glint operation pattern should not be affected by spatial variability."

<<Line 380: If I understand the method correctly, you take what is between the 95^{th} and 99^{th} percentile of the O2 ratio you derive, for each 10-12 day period. This implies that you ALWAYS select EXACTLY 4% of the soundings in one time period. Have I got this correct? It implies that if one period is much cloudier than another, it doesn't matter; you will always select exactly 4% of the ocean glint soundings (at least with the upper-edge criterion alone; I realize the h2o water line screen and the lambertian albedo screen will further remove some soundings). If so, it would be helpful to state this in the paper, and please compare it to the throughput rate of some of the more traditional approaches over ocean. It seems that 4% is rather strict, but it is hard to say.>>

It is correct, that we always take exactly 4% (95%-to-99% percentile) of the prescreened soundings within each 10-12 day period. Relating the number of 'upper edge' data to the total number of unscreened (except for instrument anomalies) soundings yields a fraction of roughly 1.4%.

Thereby, we assume that the 95% percentile is always on the safe side of what can be considered clean for the ensemble of the prescreened GOSAT soundings aggregated globally over 10-12 days. Prescreening considers the criteria described in the manuscript: instrument anomalies, non-convergence of the retrieval, quality of the fit, cirrus filter using highly absorbing water band, double-layer filter based on the Lambertian albedo correction term. Except for the instrument related criterion, these filters preferentially remove scattering-contaminated scenes.

Comparing the overall fractional yield of 1.4% for the `upper edge' to independent estimates for clean scenes over ocean is difficult since GOSAT's onboard cloud-and-aerosol imager (CAI) cannot be used for ocean-glint scenes. The cloud-flagging algorithm has not been designed for (bright) ocean-glint cases and developing our own CAI algorithm is beyond the scope of this work. Breon et al., 2005, find significantly higher fractions of "totally clear" (aerosol and cloud free) sky over most parts of the oceans. However, our goal is to be on the safe side in order to investigate instrument and retrieval issues.

We add the following caveat at

l.24, p.4381: "Further, our method can only work if the ensemble from which the `upper edge' is selected contains a sufficient number of clean scenes. When using the 95%-to-99% percentile, here, we assume that at least these 4% of the (prescreened) soundings can be considered clean within each temporal bin of 10-12 days over GOSAT's geographic sampling range. Since our prescreening preferentially removes scattering-contaminated soundings this seems a safe assumption. If the method is to be applied to spatially small regions and short periods of time, however, care must be taken to allow for a sufficient number of clean cases."

<< Section 5 general comment: What are the typical mean squared fit residuals of the fit in each window, expressed as a %? It would be interesting for the reader to see this; large variations in this between the bands might be further suggestive of spectroscopy shortcomings in particular windows. In fact, it might be information to see a plot of mean fractional fit residuals in each fit window. If you find such a plot informative, please consider adding it to the paper.>>

We follow the reviewer's recommendation. For each sounding of the `upper edge' ensemble, we calculate the difference between the measured and the fitted spectrum and divide the difference by the continuum radiance in each window. Averaging these relative differences over all soundings gives the average fitting residual relative to the continuum radiance from which we calculate the average root-mean-square difference (RMS). New figure 8 shows the fitting residuals and the derived RMS for the year 2010. Fitting residuals are smaller for the

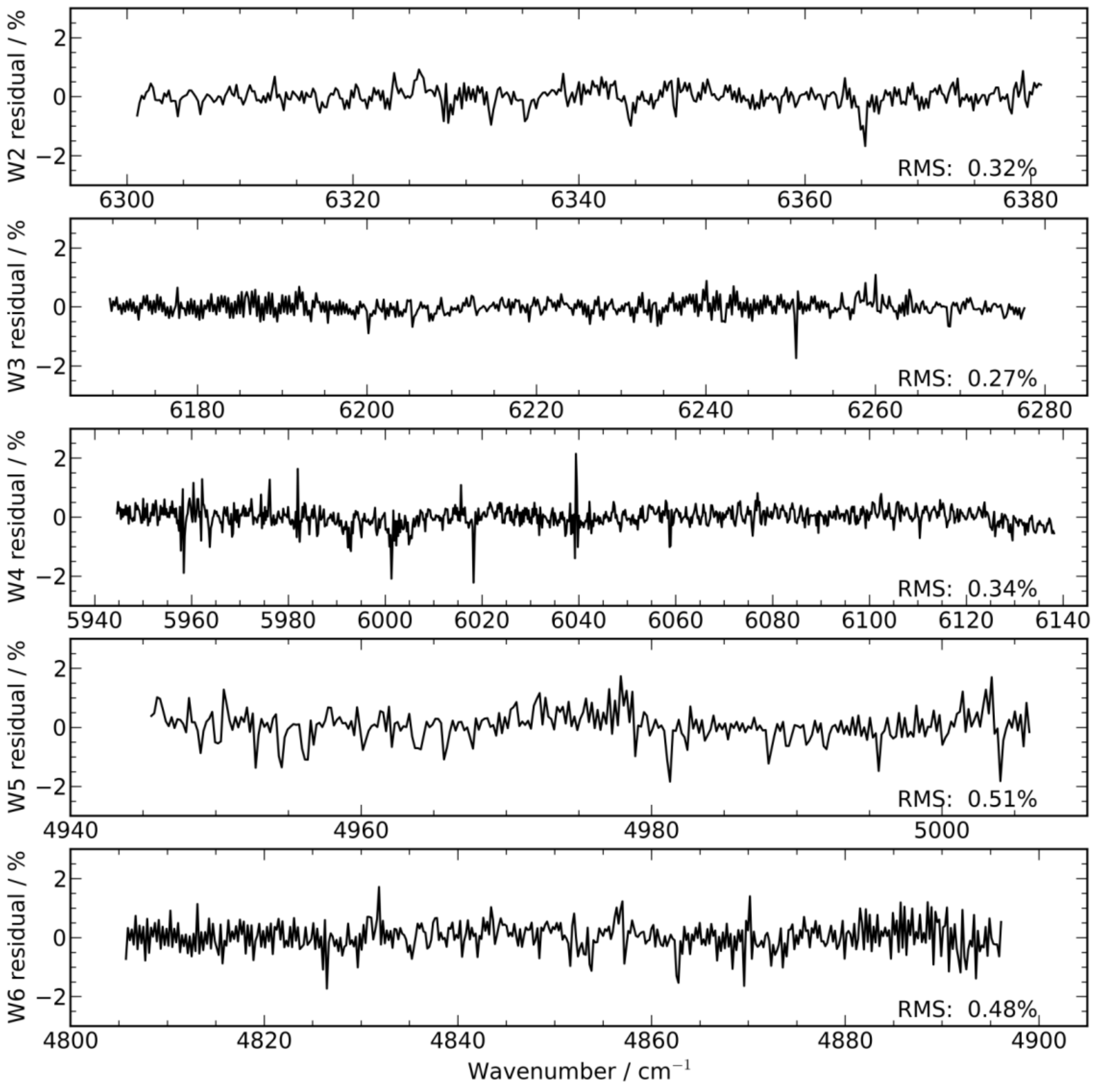

Fig. 8. Fitting residuals for windows W2 (upper panel), W3 (second panel), W4 (third panel), W5 (fourth panel), W6 (lower panel) relative to the continuum radiance. We average all 4518 `upper edge' soundings in the year 2010 such that the residuals are dominated by systematic errors. The residual root-mean-square (RMS) is quoted in the lower right of each panel.

windows W2, W3, W4 (around 1.6 micron) than for the windows W5 and W6 (around 2 micron) which confirms that uncertainty with respect to spectroscopic parameters and line shape models is most critical in W5, W6.

We add the following discussion and Fig. 8 to the manuscript at

l.19,p.4385: “Going beyond the assessment of line strengths, line shape errors can be investigated by evaluating the quality of the spectral fit. Figure 8 shows the relative fitting residuals among the `upper edge' soundings for the year 2010. The residuals are calculated by dividing the difference between measured and fitted spectrum by the continuum radiance in each window. Following our rationale that the `upper edge' ensemble is not affected by unaccounted scattering effects, the residuals in Figure 8 are dominated by systematic errors of the spectroscopic parameters or by deficiencies of the employed line shape models. Clearly, line shape calculation seems most challenging in windows W5 and W6 which cover strong CO_2 (and H_2O) absorption lines. Here, we employ the line-mixing model of Lamouroux et al., 2010, for calculating the CO_2 absorption line shape. Figure 8 indicates that even more subtle line shape effects might need to be considered for the strongly absorbing CO_2 bands [Thompson et al., 2012].”

Further, we add a remark in the conclusion at

l.29,p.4387: “Examining the spectral fitting residuals shows that line shape errors are most pronounced for the strongly absorbing CO_2 bands (W5, W6 around 4970, 4850 cm^{-1}).”

<<Section 5 comment 2: I notice that the predicted (posterior) error of (W6/W2)*100, at 0.26%, is significantly larger than the observed scatter in that quantity (0.62%). This is the only window for which this is true. I further notice that W6/W2 experiences large outliers in 2009-mid2010, which seem to go away. Please comment. Are these outliers primarily driven by W6 alone? This result implies that W6 retrievals appear dependent on something that is time-dependent, to which W2 and W3 are not dependent. >>

The manuscript already mentions this observation at l. 22, p.4383. Indeed, the observed scatter for the windows W5 and W6 (both windows!) is more than a factor 2 greater than the scatter expected from propagation of the noise error. Beside the discussion in the manuscript, one could argue that W5 and W6 are sub-ranges of GOSAT’s band 3 detector while W2, W3, W4 are sub-ranges of GOSAT’s band 2 detector. One might speculate that the band 2 noise estimate is better than the one for band 3. Though, this seems unlikely since we use the out-of-band signal for determining the noise in each band.

In general, we find that our standard RemoTeC retrievals show a factor ~2 greater data scatter than expected from noise error propagation [eg. Butz et al, 2011, paragraph 17]. We attribute this to erroneous input parameters of our retrieval forward model such as instrumental, meteorological or spectroscopic parameters. These sources of error are not necessarily of random nature but nevertheless might show up as enhanced scatter for a geophysical ensemble covering various ambient conditions. W5 and W6 might be more sensitive to such pseudo-noise than W2, W3, and W4.

Tentatively, we agree that the number of outliers seems to be larger for W5 and W6 (both windows!) in the early phase of the mission than in the later phase of the mission. Though, the number of outliers is small in general, remembering that more than 14,000 soundings are shown in Figure 6. It seems as if the number of outliers drops after August 2010 when image motion compensation was switched on. Since the observation is rather vague and we do not have a good physics explanation for it, we prefer to not speculate and not to add any discussion to the manuscript.

<<Technical/Grammatical Comments>>
Line 158: furtheron ! further on
Line 215: constraint ! constrained
Line 334: is ! are
Line 484: is ! are
Line 512: detected difference ! detected mean difference
Line 525: later ! latter >>

All corrected.

References

Bréon, F. M., D. M. O’Brien, and J. D. Spinhirne: Scattering layer statistics from space borne GLAS observations, Geophys. Res. Lett., 32, L22802, doi:10.1029/2005GL023825, 2005