

Atmos. Meas. Tech. Discuss., referee comment RC2
<https://doi.org/10.5194/amt-2022-122-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.



Comment on amt-2022-122

Anonymous Referee #4

Referee comment on "Intercomparison of Sentinel-5P TROPOMI cloud products for tropospheric trace gas retrievals" by Miriam Latsch et al., Atmos. Meas. Tech. Discuss., <https://doi.org/10.5194/amt-2022-122-RC2>, 2022

"Intercomparison of Sentinel-5P TROPOMI cloud products for tropospheric trace gas retrievals", Latsch et al.

This manuscript shows a large variety of descriptive statistical comparisons between multiple cloud products from the TROPOMI instrument on S5P and VIIRS on SNPP. The information is thorough in scope (though limited in time, as only 4 days of data are used), and is valuable to the remote sensing and applications communities that use S5P data. The thoroughness of the comparisons means there is a great number of figures, and in some places the organization is tough to follow, and I have made suggestions below. There are also a number of places where the discussion and analysis is unclear or misleading. I do think the manuscript is publishable but in the present state I would recommend a large number of changes.

Overall major comments:

It is unclear if the VIIRS cloud mask should even be included in the manuscript. Whenever the VIIRS product appears in a figure, the correlation with other products is poor, and the reader is reminded "the geometric VIIRS cloud fraction is expected to have the largest differences...". Since it is a fundamentally different product, why is it included in this comparison? Are there cases where the VIIRS cloud mask should be considered from a user's perspective?

It is unclear why the region over China was selected, particularly if the "values for Europe ... and China .. behave very similar" (Line 344). Could the China regional subset simply be excluded from the paper entirely? (Which would substantially reduce the number of figures). I would recommend moving the plots related to the China regional subset to the

Supplement only, rather than switching regions in the main manuscript, with no justification (e.g., Europe and Africa are used for all Cloud fraction comparisons, but then the presentation switches to China for Cloud height. Why was this done? It would be clearer to do the comparisons for only the Europe subset in the main paper.)

On the slopes and intercepts (appearing first in Figure 2 and many figures thereafter): for many of the comparisons, there are subpopulations in different regions, which makes the single linear fit questionable. For example, figure 7F is one of the worst examples: for the main group of points, the intercept is nearly zero and the slope is slightly larger than 1, whereas the single linear fit shows a large intercept and slope less than 1. Therefore, many of the differences between algorithms for the slopes & intercepts are of similar size to biases in the slopes and intercepts due to the data distributions themselves. I would recommend removing the fits entirely, and removing the associated "Tabular intercomparison" figures (B1, B2, etc.)

In line 535: the authors state, "...relatively small data samples may also have across-track variations .." I strongly agree with this statement, and I think this calls into question the use of the regional data subsets for the purpose of examining across-track biases. All of the cross-track plots of regional subsets appear too noisy to be of any use, and I recommend removing them from the manuscript. The plots could be retained in the supplement, but I would not include detailed discussion or analysis of features seen in the regional subset plots.

For the cross-track plots, the differences between algorithms generally quite small and subtle (other than the results from the VIIRS algorithm), and the plots are very hard to interpret. I cannot discern the color difference between many of the lines: it would be better choose a smaller subset of more clearly separate colors (perhaps 4) and use solid/dashed line styles in addition. The data would also be more clearly displayed with the reference algorithm "cf_fit" displayed as cloud fraction, and then the other algorithm data could be displayed as differences in CF relative to cf_fit. (and equivalent for the CH: plot ch_fresco and then the differences relative to ch_fresco.)

There is lack of discussion at the end of the manuscript for how a potential S5P data user should interpret all this information. The main conclusion seems to be "things are more in agreement, but there are still a lot of differences, and quality flagging matters" - which is not particularly useful for a user. Can the authors provide some sort of guidance here, or would that be considered 'out of scope'?

The organization of the supplemental figures is tough to follow. It would help to have a table of contents or similar at the beginning of the supplement document. It also would be helpful to group all the version 1 plots into a separate section. Many readers will not care about the old products: it would be more useful to be able to skip over those plots entirely.

Specific minor comments:

There are multiple places where the explanation is very jargon heavy, specific to terms related to S5P. (some are noted below)

The "OCRA" algorithm is described as the cloud fraction a priori. Clarify which algorithms actually use this as their a priori - is it just ROCINN, or do others? If OCRA is intended as an a priori, does it make sense to even include in this comparison? Shouldn't the algorithm that uses OCRA as the a priori be an improvement over the OCRA result?

Line 216/Equation (2) - this is missing a minus sign? or should be $(1 - (CP/Ps) \dots)$?

Line 229 - 233: this is very unclear. Why does FRESCO have a different number of crosstrack pixels? Table 1 says that FRESCO, ROCINN and MIXCRU all use the NIR wavelengths.

Line 236: how many VIIRS cloud mask pixels fall within one TROPOMI field of view, on average?

Line 239: These all look like density histograms, not 'scatter plots'. (these are matplotlib hexbin plots, correct? then they are histograms, not scatter plots)

Line 279: what is "BD3"? (S5P jargon, I assume?)

Line 280: instead of "fixed pixel shift", do you mean "integer pixel shift"?

Line 285/Figure 2 (and other figures): it is confusing that the algorithm names are not consistent between the text and the plot annotations. (specifically, "ROCINN" and "OCRA" are not used in the figures. "NO2" became "fit")

For Figure 2, and all similar figures: the colormap appears to be deeply saturated in all cases (meaning, there are many grid cells with values larger than the highest value in the color bar.) These colormaps should be switched to a logarithmic scale.

Line 291: how could a degradation correction create the subpopulations in figure 2e? Shouldn't a degradation correction apply to all points, making the slight less-than-1 slope seen in 2e?

Line 300 caption: Typo, this says "v1 - v1", should be "v2 - v1"?

Line 312: would suggest saying "scatter" (the spread of values in the graph) instead of "scattering" (physical process of light interacting with particles).

Line 315: sentence "In addition, the error in the FRESCO cloud height..." is very unclear.

Line 317: I do not see how you can argue that the FRESCO matches ROCINN better by what is presented in Figure 4. We are only presented with comparison between each algorithms' version 1 and version 2, not between the algorithms.

Figure 5, and other later figures: the subplot labels "summer day", "winter day", etc, are ambiguous. Replace these with the actual dates used.

Line 380-382: unclear, what is 'polynomial and Ring term'? more jargon?
"This leads to a bit higher values for cf_fit than for O2-O2" - I see almost no data points in figure 7e, with cf_fit larger than O2-O2.

Line 400: figures 8d and 8f show negative differences that are spatially correlated to the snow-ice mask from Figure 3, but 8c and 8g show negative differences over much different spatial domains. In fact the large region centered at 50E, 60N, in 8g is snow covered but the CF values are nearly identical.
In figure 8 the display range appears to be too narrow, in particular 8g appears to be mostly out of range. A nonlinear level spacing might help - perhaps [0.1, 0.3, 0.6] instead of [0.1, 0.2, 0.3] for the major levels in the colorbar?

Figure 10 & Line 447 - Line 452, it is concluded that MICRU is doing better over the ocean glint, but this is not demonstrated by the difference plot. From the difference plot, we only know that the two algorithms respond differently in the glint region, not which is more correct. It would help to add another panel that shows the actual cloud fraction reference map (from cf_fit), and there is room in the upper right. Presumably, the cf_fit should show a false excess cloud fraction that lines up to the glint region shapes we see in the difference plot (8f). The color map is again saturated (same as Fig. 8)

Figure 13 - same comment as Figure 8, 10, the color scale is heavily saturated. A nonlinear scale should help.

Figure 15: why does panel (d) show such good agreement compared to 14d? I think Fig 15 should have the same data as 14, just "zoomed-in"? Is this a trick of the way the data is 'saturating' the color scale? (see earlier comment)

Line 530: "No observational effects for full cloud cover ..." - I do not think this is true. For fully cloudy pixels there are still anisotropies in the scattered radiation, that will conflict with the Lambertian cloud models used in many of the retrievals.

Line 560 "... very similar run of the curves ..." what does this mean?

Line 573-577: if the "dip" is due to the change in the binning scheme, shouldn't there be a similar feature on the other side of the cross track scan?

Line 580 "no overall indication for a systematic across-track problem ... except for the above-mentioned FRESCO issue over Africa."

I do not understand this claim, nor the FRESCO issue. In Figure E1, isn't the green line (cf_apriori) the outlier here (ignoring VIIRS)?

Note that these claims are relative to the regional subset data, where the data is too noisy to make these conclusions (see earlier major comment) - in any of one of the subsets the actual mixture of clouds observed across track could be very different. If one algorithm tends to have a bias for low clouds but not high clouds, this would then manifest as a cross-track bias.

Line 595: doesn't the yellow line (ch_base_cal) also show a 'step' due to the interpolation (is this the feature at cross track pixel 22?) If not, I am not sure what 'steps' are notable in Figure 19.

Line 612 - 630: I do not see the value in comparing the not-matched data: the sampling effects (meaning, different algorithms will include different cloud populations) would be so strong that the comparison is not meaningful. If the authors wish to retain these plots I recommend reducing or removing the analysis discussion.

Line 699 - 703: These conclusions are not convincing without more detailed analysis. If such additional analysis is 'out-of-scope', then at minimum, the authors should modify the text since these conclusions in current form are very speculative.

If more analysis can be done, here is other literature on this topic that should be cited (e.g. Maddux, B. C., Ackerman, S. A., & Platnick, S. (2010). Viewing Geometry Dependencies in MODIS Cloud Products, *Journal of Atmospheric and Oceanic Technology*, 27(9), 1519-1528). There are other mechanisms which could imprint a cross track

variation in the CF, for example increased optical path at the scan edges.
It would also help to stratify the CF data by CH to refine the analysis: is the increase in CF at the edges similar for low-level and high-level clouds?