

Atmos. Meas. Tech. Discuss., referee comment RC1  
<https://doi.org/10.5194/amt-2021-141-RC1>, 2021  
© Author(s) 2021. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Referee comment on amt-2021-141

Anonymous Referee #1

---

Referee comment on "Retrieving H<sub>2</sub>O/HDO columns over cloudy and clear-sky scenes from the Tropospheric Monitoring Instrument (TROPOMI)" by Andreas Schneider et al., Atmos. Meas. Tech. Discuss., <https://doi.org/10.5194/amt-2021-141-RC1>, 2021

---

The paper by Schneider et al. presents an extension of the former TROPOMI column data set published by the authors last year that was restricted to cloud-free scenes. It is great that the authors now can also retrieve the cloudy scenes and thus extend their previous data set. This allows now for a much wider scope of scientific application than the previous data set and presents therefore a quite valuable data set.

I think the only drawback is that the structure of the paper is quite similar to the previous paper by the authors (Schneider et al., 2020). It would have been nicer and maybe more interesting to set the focus on the comparison of the old and new data set and the additional gain and application possibilities of the new data set instead of just extending the previous analyses with some more stations (though this of course is also nice and valuable).

I have following comments I would like to ask the authors to consider before publication.

P1, L1: The first sentence of the abstract is not clear without knowing what you have done. I would suggest rephrasing and being more precise.

P1, L4: Clouds are usually present over the oceans and over land. How does it then comes that you in case of cloud-free scenes derive data over land and if you consider cloudy scenes over oceans?

P2, L36-38: Same as for P1, L4. An explanation or more information would be quite helpful.

P2, L40-41: This sentence is not easy to understand, please rephrase.

P2, L42: "Section 2....." this comes a bit suddenly. Add an introductory sentence to begin this paragraph of paper structure description as e.g. "The paper is structured as follows:" or write "In the next section we describe the retrieval set-up...."

P3, L59 and 61: fitted to what? Please be more precise.

P4, L99: "except if the data are assimilated using averaging kernels" I would suggest to make an extra sentence since also the data of all scenes can be used when the averaging kernels are applied. This does not only hold for data assimilation.

P6, L101: Retrieve what exactly? A specific gas or trace gases in general? Also here I would suggest to be more precise.

P6, L103ff: How is it guaranteed that the missing data that is filled in is reliable? Is there another filter used to check the quality or to filter out the non realistic filled up information?

P6, 110: How was this correction factor derived? By comparing FTIR to TROPOMI? If yes, how can you then be sure that the correction factor holds for both data sets (old and new)?

P6, L111: missing calibration of what? Be more precise.

P6, L127: Also in this paragraph you could be a bit more precise. So it seems the altitude difference is generally a problem and as higher the station as higher the problem/error gets? You could e.g. more clearly state here that this is the reason why a high-altitude stations are considered separately.

P7, L135: Extended to 0 hPa? This is really high. Is this realistic? Are measurements made that high up?

Figure 4: Why using an average? Why is not the correction factor for each station used? How large is the introduced error by using an average?

P10, L160: Why should one fill the null-space with MUSICA-NDACC profiles if this data set is then used for validation? TROPOMI would then not be an independent data set. However, for validation rather an independent data set should be used.

P10, L173: This is a very complicated sentence and very hard to follow. Please rephrase and distribute the content over several sentences.

P11, L181: add "which is then used for the validation"

P12, Figure 5: The figure could be improved by using lines and symbols that are a bit thicker/larger. Looking at the number of data points it looks like there are almost no data points from TROPOMI.

P13, Figure 6: Why is TCCON corrected? I thought MUSICA-NDACC is the data set that needs to be corrected?

P14, L228: Why is this connected to the different conditions in cloudy and clear-sky weather? Isn't this simply due to the nature of the relative differences? These are usually more severe (higher) for lower values. The differences of course (i.e lower values and larger differences) may be related to cloudy or clear-sky conditions.

P16, L241: Refer here again to Figure 8 .Add also a table with the biases? The approximate values can be derived from the figure, but if one needs the exact value a table would be quite useful.

P17, Figure 9: As for Figure 5, the lines and symbols are too thin and the colors hard to differentiate. For the clear-sky data the corrected and uncorrected data are shown, but for cloudy only the corrected data are shown. Why? Shouldn't also here both been shown, corrected and uncorrected?

P18, L278: What is the "scaled" and what the "depleted" prior?

P18, Figure 10: Also here the figure could be improved by using a thicker line style.

P21, L294:"The data reduction method is described in Sect. 3.3"? You mean the collocation criteria is described in Sec. 3.3?

P21, Figure 13: Use also here colors that are better differentiable and line styles that are thicker and better visible.

P22ff: I really appreciate that you demonstrate the application of the data set, however, since your paper is already quite long and complex I wonder if it wouldn't be better to have an own, more sophisticated application paper where you also could include model simulations. You could only show here Figure 14 and give a short description of where exactly you gain more information and in which areas thus more sophisticated studies are possible. Another option would be to keep this example short. In that case, I would suggest skipping the details on Figure 14 (the discussion of the dD distribution, L307-L333, it does definitely not need to be that detailed) and put the focus only on the case study itself.

P23, L305: Here you jump from data coverage to different dD amounts. Of course, different weather conditions result in different dD amounts, but that relation should be better explained.

P23, L310: "most of the scenes in the area" is rather confusing and I would suggest to rephrase to "dominate the scenes".

P23, L311: "slight underestimation of total column dD" Why? Shouldn't the cloudy retrieval cover these values?

P27, L375: "or"? I though the cloudy data set contains all data? Or then do you use all data? This should be better explained throughout the paper.

General question: If a scientific study is done is then the cloudy data set enough or does one need both data sets (the cloudy and the non-cloudy data set)? Then it would be interesting to see how a combination of both data sets can be used.

P28, L396-397: I am not sure if you are here a bit too optimistic. This could be just coincidence.

### **Technical corrections:**

P1, L4: Add "new" so that it reads "The new data set....."

P1, L12 and several other occasions in the manuscript: "prior", shouldn't that read "a priori"?

P2, L24: I would rather write here "useful" instead of "promising"

P2, L44: move the reference in line 46 to line 44, otherwise these are a bit lost in line 46.

P9, Figure 4 bottom panel: In "Karlsruhe" the "l" is missing.

P10, L155: In the Figure 4 caption the average value is given as 1.1527

P10, L166: two full stops, thus one obsolete.

P10, L166: List with comma: "of cloud water, total water amounts and isotope ratios."

P10, L167: Deployment -> campaign or field mission

P10, L127: 27 Sep 2018 and 21 Oct 2018 -> 27 September and 21 October 2018

P11, L186: "above" obsolete

P11, L188: in loc. cit.????

P11, L192: Pearson coefficient -> Pearson correlation coefficient

P11, L199: by small sensitivity -> by the small sensitivity

P13, Figure 7, second panel: put -1 should be in superscript

P16, L254: ..procedure applied.. -> ...procedure as applied....

P18, Figure 10 caption: Sep -> September

P18, L281: Pearson coefficient -> Pearson correlation coefficient

P23, L335: add "still", so that it reads...is still very sparse.....

P23, L309: moistest -> moisture

P23, L336: add also here "still" so that it reads "is also still sparse"

P25, Figure 15 caption: Jan -> January