

Atmos. Meas. Tech. Discuss., referee comment RC1
<https://doi.org/10.5194/amt-2022-94-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on amt-2022-94

Anonymous Referee #1

Referee comment on "TROPOMI/S5P Total Column Water Vapor validation against AERONET ground-based measurements" by Katerina Garane et al., Atmos. Meas. Tech. Discuss., <https://doi.org/10.5194/amt-2022-94-RC1>, 2022

Review: "TROPOMI/S5P Total Column Water Vapor Validation against AERONET ground-based measurements" by Garane et al.

General comments

In this study, Garane et al. validate a new TROPOMI TCWV product by Chan et al. (2022) using 2.5 years of ground-based measurements from AERONET as reference. They also investigate the influence of different input variables (e.g., viewing geometry, surface albedo, clouds) and the retrieval results themselves (i.e., H₂O SCD and AMF) on the retrieval performance.

Although the overall aim of the paper is interesting, I have some concerns about the validation analysis. I also feel that the authors miss an opportunity by not taking advantage of the high density of AERONET stations in certain regions and instead averaging the results zonally, which unfortunately also results in a lot of information being lost.

That being said, I recommend publication if the following points and concerns are addressed.

Major issues

1. It is not completely clear how the collocation of ground-based and satellite measurement was conducted. If I look at the total number of data and roughly calculate, I results in about 2 measurements per day per station ($633000 / (365 \times 351 \times 2.5) \sim 2$). And here all filters are already taken into account ($CF < 0.5$, AMF, RMS, SZA). In my opinion, this number seems unexpectedly high, especially considering that not all stations were able to provide measurements for the most recent months. The authors must explain the collocation procedure more clearly: were several satellite pixels (within 10km to the reference station) compared with one measurement from AERONET? Or is simply a large fraction of the collocated measurements at stations at high latitudes? If the former is the case, then some reference measurements would be used more often than others, making the comparison inconsistent. I would therefore suggest either to take only the closest satellite pixel within 10km or to calculate the mean value of all satellite pixels within 10km and compare it with AERONET.

Furthermore, negative TROPOMI TCWV values also appear in the comparisons (for instance in Figure 3b and 6b). I would ask the authors to clarify where these negative values come from.

2. The use of zonal means does not make much sense, as TCWV has a high variability along a longitude. This also negates the great advantage of AERONET, namely the network's high station density. I suggest the authors to restructure Section 4 as follows:

- One should carry out the analysis of TROPOMI vs. AERONET (i.e. the regression analysis) for each station individually and then present the fit results on a world map.
- For regions with a high station density (e.g. Europe and North America), separate plots could be shown separately.
- In the corresponding regions, one could interpolate the regression results of all stations and then analyse how well the performance of the retrieval depends on geophysical parameters, i.e. performance in humid/arid areas, influence of albedo, ground elevation, etc. This would also lead to an overall better understanding of the retrieval.

3. It is not really clear why AERONET is used at all when the authors themselves say that AERONET most likely underestimates the actual TCWV content (see Section 2.2) and the quality of the measurements also strongly depends on the calibration of the instrument on site (see line 166f). What is the great benefit of AERONET compared to other measurement networks? GPS measurements from SuomiNet or IGS, for example, have a much higher accuracy than AERONET and can also measure under all-sky conditions.

Other

Overall, the quality of the figures needs to be significantly improved: Instead of point clouds, 2D histograms should be used (e.g. Figure 5 or 6b). The numbers in Figures 10-13 are hardly readable. It might be better to show the amount of data points using a colorbar with coloured dots.

Moreover, the language should be improved. Here and there the wording is not really appropriate. For example, "quantity" should be replaced by "variable" and "percentage difference" by "relative difference" throughout the paper.

Specific comments

L11: "blue wavelength band" to "visible blue spectral range"

L12: MetOp

L18: -3%: Table 2 shows much higher values (-4 to -10% in NH, +2 to +6% in SH).

L21: "low cloudiness" --> low cloud heights

L23: "-4 +- 4.3 % with the ground-truth": In Section 5 it is written that it is -9 to -13%. Accordingly, one should write here that it is -4% in relation to AERONET, but probably -9-13% to the "truth". The slope of the linear fit in Figure 6b (Section 4) gives a value of 0.89. So the retrieval actually underestimates by about -10%?

L28: Water vapour does not have to form clouds to be transported around the globe.

L40: It should be mentioned that a major source of stratospheric H₂O is methane rather than tropospheric H₂O.

L43: Here some exemplary instruments (and corresponding papers) should be mentioned. GPS radio occultation is missing in the list.

L50: It is a bit strange to mention Schneider et al. (2020) and not refer to the other TCWV retrievals from TROPOMI by Borger et al. (2020) and K uchler et al. (2021). In particular considering that Borger et al. (2020) also retrieve TCWV in the visible blue.

L70: TROPOMI was launched in October 2017.

L80: DOAS: reference missing (e.g. Platt and Stutz, 2008)

L84: Which improvements have been implemented in the spectral analysis and the AMF calculations?

L92: Not really necessary to mention the data format.

L114: Replace "utilized" by "used"

L128: Is the reference source now Martins et al. (2019) or Smirnov et al. (2004) and Alexandrov et al. (2009)?

L131: "coverage of all continents": looking at Figure 2 only North America and Europe are covered well. Please rephrase.

L140: How does the "in-house quality control" look like? Please clarify.

Section 3: Here it would be more interesting to show an example from another climate zone rather than showing two similar stations. Replace one of the examples with another one (maybe from the northern mid-latitudes, where most of the AERONET stations are located).

Figure 3b: Is the value shown for the correlation the Pearson's correlation coefficient R or the coefficient of determination R^2 ? And is the linear fit based on ordinary least squares (OLS) or total least squares (TLS)? Since the uncertainties of TROPOMI and AERONET are

of comparable magnitude, a TLS might be more appropriate.

L165: $0.788 < 0.79$

L166: Shown is the monthly mean bias, right?

L168: The period of 2.5 years is much too short to speak of a high temporal stability, especially if the time series also has some gaps.

L185: With the high number of measured values, the standard error does not provide any additional, relevant information and should therefore be removed from all comparisons.

L196: A slope of 0.89 basically means underestimation of more than 10%? Has a TLS been used? Negative TCWV values in TROPOMI comparison. Please clarify, where they come from.

L207: "temporal stability": see comment above

L212: "mainly representing the latitude belt 0° to 60° S": redundant information, as only there is only one station in Antarctica.

Table 2: What has been the reason for the different latitude binning?

L265: "The performance mainly on the aspect of the surface albedo parameter credibility appears to be sound.": Considering that the albedo is an elementary input parameter of a satellite retrieval, the performance should not be called "sound". Rather acceptable?

L285: Instead of looking at the VZA dependence, it might be better looking at the row dependence so that one can see if a West-East dependence exists in the TROPOMI swath.

L298ff: Please clarify how these dependencies compare to the findings in Borger et al. (2020)?

L315ff: So why do you investigate surface pressures lower than 900hPa, when you only have a limited number of measurements?

L318ff: Since there are hardly any measurements for albedos > 20%, this should also only limit to 20%, but make the binning finer.

L360: "10-19%": this is only valid for TCWV in the tropics. Likely much higher in the mid-latitudes.

L375ff: Since there are almost no measurements in the polar regions at high latitudes, the statistics are not meaningful.

L382: "cloudiness" --> cloud top height

L386: "temporally stable": see comment above

L386f: "product of high quality and precision, temporally stable and not affected by any other parameters, except from clouds": This is a contradiction in terms, considering that a few lines earlier it is mentioned that the TCWV retrieval likely underestimates by 10-13%, the time series is far too short for analyzing temporal stability and clouds are by far the most important input parameter in satellite retrievals.

L403: Competing interest: To the best of my knowledge, DL is an editor of AMT, which, according to journal's publication guidelines, should also be mentioned here.

L448: Apparently, the reference from Kleipool et al. was mixed with another from Köehler et al.