2nd review of manuscript "essd-2021-319-manuscript-version4.pdf" entitled "A 16-year global climate data record of total column water vapour generated from OMI observations in the visible blue spectral range" by Christian Borger, Steffen Beirle, and Thomas Wagner.

General comments

I acknowledge for the additional work done by the authors to address the questions of time and space sampling and clear-sky bias. However, I cannot find the answers to my main criticism and recommendations, which are still pending. Let me recall them:

"the conclusions sound far too optimistic to me, given the poor agreement found between the OMI data and the validation data. Especially, the large positive biases over land and near the coastlines in the tropics are striking and not sufficiently commented or explained."

"I recommend first a more insightful analysis of the error sources, especially over land and, if possible, the elaboration of an improved version of the data set, and second, a more comprehensive discussion of the validation results in a revised version of the manuscript.

The biases in the proposed data set are striking, see the reddish patterns in Figs. 3,4,5,6,7,8 of the first submission, and discredit the produced data set, to my opinion. Consequently, I don't think the intercomparisons provided in this manuscript are relevant as long as these biases are unexplained and uncorrected.

Specific comments

Once the biases have been corrected, the authors may also take the following specific comments into account for a revised submission:

1) Reference data sets

There is not reference provided for ESA WV_cci CDR-2 and it is indicated that this is a beta-version data record. In this case, I recommend to discard it from this study until a dedicated assessment report or publication for the data set is available. Reviewer #2 made a similar comment. The intercomparison would be more relevant based on ERA5 and RSS data which are well-established data sets.

2) the impact of time and space sampling and row anomaly are relevant to discuss in the manuscript but the details of the study would better fit into a supplemental material.

3) On the stability requirements (comments to the answer from the first review)

You are right that the GCOS-112 report that I quoted refers to radiosonde or profile measurements. However, since TCWV is the integral of profile data, the same requirements may be applied also to TCWV. This feeling is actually corroborated by the URD (2021) report that you are mentioning. Table 3.1 from this report gives indeed a 0.3%/decade stability requirement in TCWV based on 2 reference documents: GCOS (2016) and G-VAP (2013). The 1%/decade requirement which you cite from this report actually comes from a survey study conducted within the scientific community, but based on 38 answers only. I'm not sure to which number one should give more credit. At least, both numbers may be cited and properly acknowledged.

URD (2021) User Requirements Document of the ESA CCI WV (https://climate.esa.int/media/documents/Water_Vapour_cci_D1.1_URD_v3.0.pdf).

4) Linear regression method

In the first submission you used OLS and ODR without justification. You decided to remove OLS and replaced it with PWLF but still without justification and no details or reference citation (and you define the acronym as “piecewise linear regression”, but I guess the F is for fitting). I guess your motivation for including PWLF is to account for the non-linear behaviour of the scatter plots in some cases (e.g. Fig. 9c) but you don’t actually comment/interpret the change in the slopes. What is the physical explanation behind? I suspect that it just reveals the huge positive biases in the OMI TCWV data set over land. Moreover, comparing one ODR slope and two PWLF slopes does not make much sense (you are comparing the parameters of two different models). I also doubt that the PWLF model is relevant in some cases: Fig. 7, 9a, 10a where the break points are found at very small x values (even negative in Fig. 10a, although I cannot understand how this is possible).

5) Trend estimation

L350: The estimation method is said to be OLS and the following citations are quoted, but the details cannot be found actually:

Danielczok and Schröder (2017) is not available online

Beirle et al. (2018) does not explain how the trends are computed

Saunders et al. (2010) cited in Beirle et al. (2018) is not available online

Grossi (2017) cited in Beirle et al. (2018) used “a simple linear least-square regression analysis” and applied an augmented Dickey-Fuller (ADF) test, but this procedure is flawed. The ADF test only says if the series is stationary or not (null hypothesis). In order to take the serial correlation into account and derive properly scaled standard errors, a generalized least-square regression must be used instead.

6) Specification of the relative errors for the ODR

Following a comment both from myself and the 2nd reviewer, some justification of the specification has been added. However, they are still debatable. For SSMI, you cite Mears (2015), but I cannot find the numbers you are giving: 2-4 kg/m2 at 60 kg/m2. In their Table 4, the GPS-satellite difference reads 2.0-2.7 for 60 kg/m2, and this numbers obviously include errors from both GPS and satellite. For the other data sets, your choices are still very arbitrary. A comparison of the used data sets to a common ground truth (e.g. GNSS or radiosondes) may help to set proper values (it should be possible to find some in published work).

7) Correlation coefficients

Your answer to my comment on the correlation coefficients:

“With regard to the correlation coefficient, we cannot fully agree, as it includes spatial variation in addition to temporal variation: namely, if we reverse the latitudes, we only obtain a correlation of R=0.63 for RSS and R=0.45 for ERA5 over land.”

I cannot undestand why the latitude reversal has an impact. Once the data are paired, the correlation coefficient does not depend on their sorting.

My comment was to say that if x and y include the seasonal signal (which is the dominant temporal variation for monthly time series), the correlation coefficient will always be very close to one. For monthly data, it is more relevant to compute the correlation coefficient of anomalies.

As an alternative, you may indicate the coefficient of determination (R^2) or better the adjusted R^2, as a measure of the goodness of fit.