Third 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

The data set described in this work clearly suffers from a major limitation, the large moist biases over land and near the coastlines. This has already been emphasized in the previous review reports but has not yet been solved and no directions to solve it are proposed in the manuscript. I do not see much utility of this data set as long as the intrinsic causes of these biases are not eliminated.

However, solving the moist bias problem may require substantial work on the retrieval method. If this problem is properly acknowledged and the directions to solve it are discussed (i.e. not just as a list of "possible" error sources), then the publication of an "interim" data set may be considered, but this data set should be limited to quality-assured data only (e.g. based on the effective clearing "masks").

Specific comments

1) Intercomparisons taking masks into account

What are the criteria and thresholds used to create the "common mask" of the ESA data set?

Illustrate on a map which grid points are filtered out by the threshold of 100 measurements of the "dynamic" monthly flag described in Appendix B. How does it connect to the percentage of observations shown in Fig. B3?

2) Reference data sets

The ESA "COMBI" product is still described as a beta-version data record with no validation results published. As such, it should not be considered as a "reference data set". The comparisons to this data set belong rather to a supplemental information section than to the main text.

What is the accuracy and stability of the IGRA-2 data set included in the new manuscript? It is well known that radiosonde data suffer from various biases in inhomogeneities and should not be used for trend analysis.

3) Sampling errors

Although the impact of time sampling is a relevant question, it is not currently a major limitation of the data set (positive biases). It should just be summarised with a few numbers in the text with the details (6 figures) provided as supplemental material.

It is not clear if the TCWV data from the different reference data sets (RSS, ERA5, etc.) are time-matched with the OMI data or if monthly mean products are used. Should be clarified.

4) Linear regression method

The use of two different regression methods (ODR and PWLR) is still not justified. What are the motivations?

Moreover, since both methods use different linear models and assumptions on the errors, it is difficult to compare their results.

In addition, the estimation of the break point x0 in the PWLR is an ill-posed problem, how is it handled?

The PWLR results where x0 is very small are not relevant (e.g. Fig B4). Instead use either a fixed x0 (e.g. 25 kg/m2) or just use a model without break (in this case ODR is preferable).

Standard errors on the estimated slopes should be considered when results are compared.

5) Trend estimation

I am afraid the GLS method introduced in the new manuscript is not correctly implemented. Firstly, identifying the AR(p) order from the ACF is not robust. There is extensive literature on this subject.

The previously cited unavailable references have been removed, but one of the new reference is not available either (Prais and Winsten, 1954).

Secondly, the cited literature deals only with AR(1) models, not AR(3).

What is the formulation of the transformation matrix P in the case of AR(3)?

Why are the new trend estimates are very different from those in the previous manuscript?

Standard errors should be given on the estimated trends, and statistical significance should be quantified by a p-value or a significance level.

The comparisons to the COMBI and the IGRA-2 data sets show large trends, which do not fulfil the stability requirements. These results should be preferably removed or at least moved to a supplemental material and interpreted with caution.

6) Correlation coefficients

My suggestion to use the of determination (R^2) rather than the correlation coefficient applied of course to TCWV anomalies (not to TCWV) since both R^2 and R suffer from the same limitation (being close to one when the seasonal signal is included).