We would like to thank the referee for reviewing our manuscript. Below we reply to the issues raised by the referee, where
blue repeats the reviewer's comments,
black is used for our reply,
and green italics is used for modified text and new text added to the manuscript.

Review of manuscript essd-2021-319 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

This manuscript presents a total column water vapour (TCWV) data set derived from 16 years of OMI observations. The retrieval method was developed in a previous publication (Borger et al., AMT, 2020). It has been slightly improved in order to meet the long term stability requirements for a climate data record. The manuscript describes briefly the modified aspects of the retrieval algorithm and gives additional details on the data quality control applied for this specific purpose. The latter seems to reject a significant fraction of the raw data, although this number is not indicated. Most of the manuscript is devoted to a comparison/validation analysis of the OMI TCWV results with respect to satellite microwave radiometer data (SSM/I), reanalysis data (ECMWF’s ERA5), and the ESA/CCI/CDR-2 water vapour product. The comparison results are fairly documented, including scatter plots, Hovmoeller diagrams and maps, although some synthetic statistics are missing (see the specific comments below). However, 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. Two main reasons are hypothesized: too low land surface albedo and incorrect cloud information, both leading to an underestimation of the AMF. These paths should be further explored in order to achieved a more reliable product meeting the climate data quality requirements. Although it is shown that the results are improved when a special could mask is used, this is only an artificial way to improve the quality of the product.

Regarding the temporal stability, it is not clear how the significance of the global mean bias, RMSE, and trend differences are established. It seems to me that the numbers are beyond the limits usually required for water vapour climate data (e.g. an error of 0.1%/decade in the global mean TCWV trend represents nearly 20% of the signal). Moreover, the uncertainty due to different time and space sampling with the different reference products should also be quantified.
Many thanks for pointing out the GCOS requirements! For our value of 1.0%/decade we have followed the User Requirements of the ESA CCI WV (https://climate.esa.int/media/documents/Water_Vapour_cci_D1.1_URD_v3.0.pdf). If we understand correctly, the value of 0.3%/decade in the GCOS document refers to radiosonde measurements or their WVMR measurements (see requirement tables in Appendix 1 of the GCOS document suggested below in the Specific Comments). Please also note that the requirement mentioned refers to the global mean. However, we are also interested in regional trends, which usually have significantly higher magnitudes.
In addition, we determined the sampling error (and the clear-sky bias) and added the following text to the revised version:
Although satellite observations enable the analysis of trace gas concentrations on global scale, a fundamental problem is that typically a satellite measurement is only taken once a

day for one location. Furthermore, satellite measurements are usually only available under cloud-free conditions, especially in the visible or infrared spectral range and thus no continuous time series is guaranteed. Consequently, they cannot provide a complete picture of geophysical variability, which leads to sampling errors in the calculation of averaged values (e.g. monthly means).

Moreover, the question arises to what extent the limitation to cloud-free pixels influences the monthly averages determined from the OMI satellite measurements, i.e. whether in the OMI TCWV data set a so-called "clear-sky bias" exists. Gaffen and Elliott (1993) investigated this bias using radiosonde ascents and found that the TCWV is about 0-15% lower under cloud-free conditions than under cloudy conditions. Similarly, Sohn and Bennartz (2008) found a clear-sky bias between MERIS and AMSR-E of about 10%.

To estimate the sampling errors, we follow the methods of Xue et al. (2019) and Gleisner et al. (2020): we choose hourly-resolved ERA5 data with a spatial resolution of 0.25°x 0.25° as reference data and collocate the ERA5 data with OMI overpass times. These data are then resampled to the 1°x1° resolution of the OMI TCWV data set and the monthly averages are calculated (TCWV_sampled). We then take the complete, original ERA5 data, resample it to the same spatial resolution and calculate monthly means from this data as well (TCWV_true). The difference between the two data sets then represents the sampling error:

$$e_{sampling} = TCWV_{sampled} - TCWV_{true} \quad (4)$$

With this definition, the sampling error summarises the uncertainties due to gaps in the swath, temporal differences or missing data (e.g. due to clouds) (Xue et al., 2019).

Figure 3 shows the annual mean absolute and relative sampling errors for the year 2006. Overall, it can be seen that most deviations are negative, i.e. the actual TCWV is underestimated. Regarding the absolute deviations, the strongest deviations can be seen in the area of storm-tracks in the mid-latitudes (e.g. North Atlantic) and the polar regions with values around -5 kg m-2. The smallest deviations are found in the quasi-permanent cloud-free regions in the subtropics. As expected, the relative differences increase from the equator towards the poles due to the decreasing TCWV values and reach values stronger than -30%.

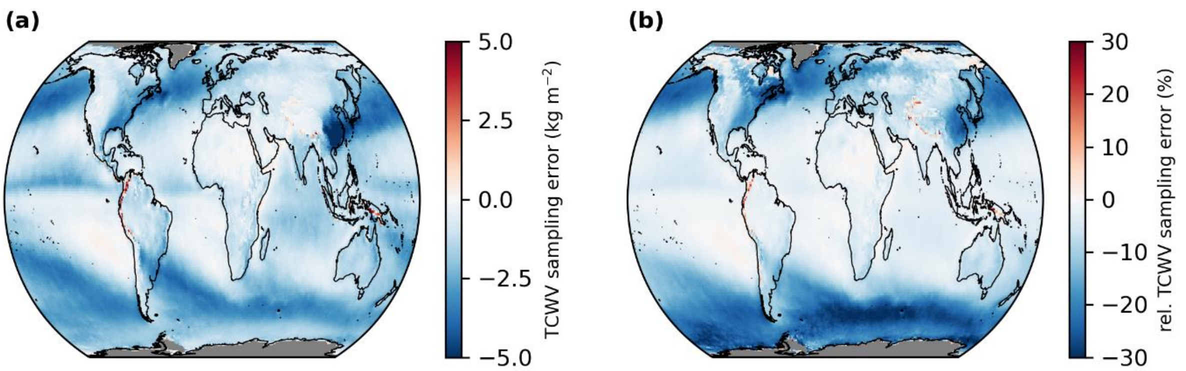

Figure 3. Global distributions of the mean sampling errors derived from monthly mean sampling differences for the time range January 2005 to December 2020. Panel (a) depicts absolute sampling error (i.e. "sampling) and Panel (b) relative sampling error (i.e. "sampling=TCWVtrue). Grid cells for which no data is available are coloured grey.

To investigate to what extent these deviations are related to the clear-sky bias, we proceed similarly to the calculation of the sampling error: we collocate the ERA5 data to the OMI overpass time and once apply a cloud filter (effective cloud fraction < 20%) and once not. Then we resample both data sets to 1°x 1° and calculate monthly means. The difference of both data sets then represents the clear-sky bias:

e_clear = TCWV_clear – TCWV_all (5)

To determine seasonal structures, the global distributions of the absolute and relative clear-sky bias for the different seasons were determined from the monthly differences (see Fig. 4). Overall, the distributions of the clear-sky bias correspond very closely to the distributions of the sampling error, both in strength and in pattern. Moreover, the absolute and relative deviations show only slight changes between the different seasons.

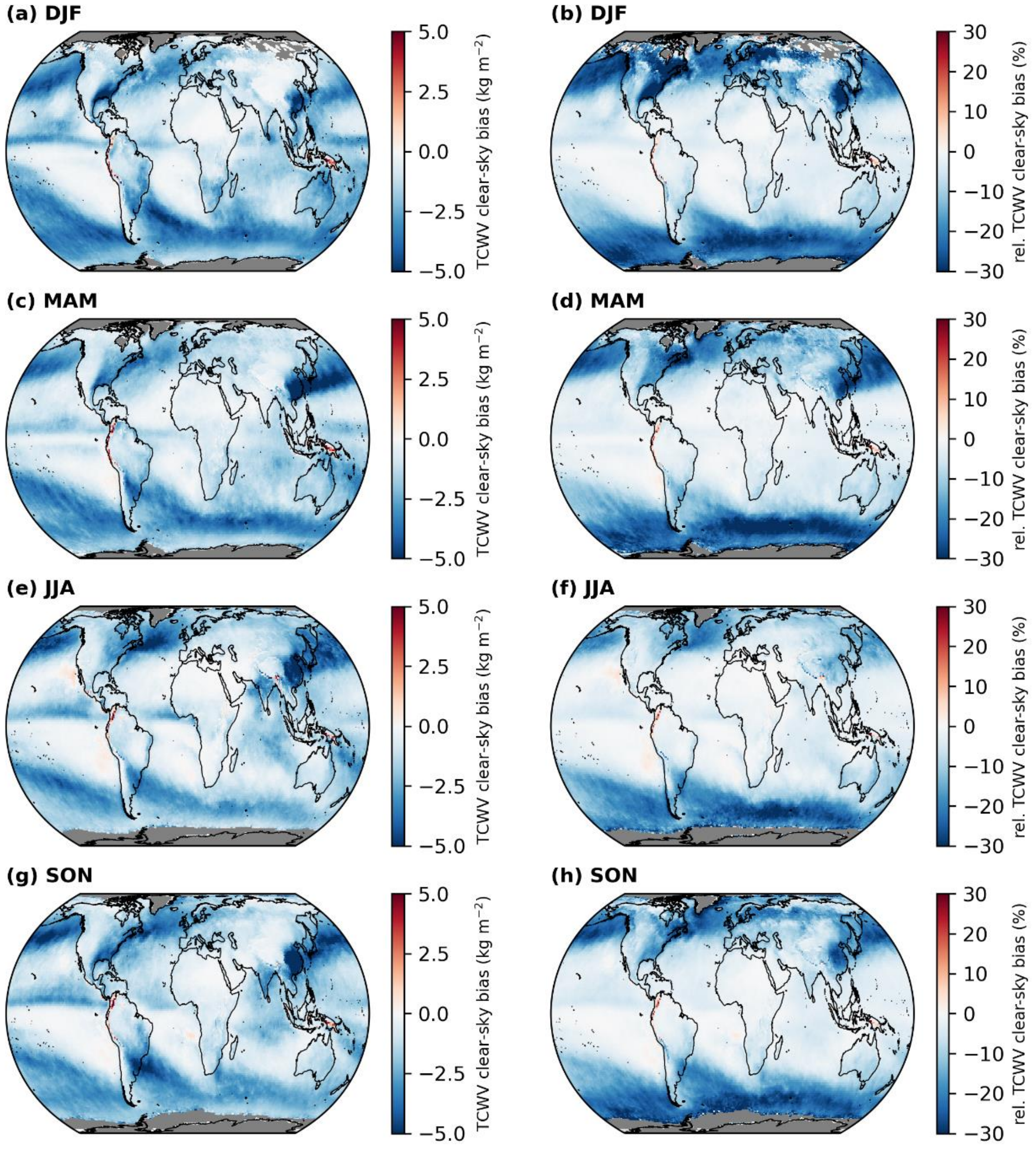

Figure 4. Global distributions of the absolute differences ("clear; left column) and relative differences ("clear=TCWVall; right column) of the monthly mean differences between clear-sky and all-sky ERA5 based on the OMI cloud information for winter (DJF; a & b), spring

(MAM, c & d), summer (JJA, e & f), and autumn (SON, g & h) for the time range January 2005 to December 2020. Grid cells for which no data is available are coloured grey.

Figures 5 and 6 summarize the sampling error and clear-sky bias distributions, respectively. For the sampling error we obtain a mean absolute deviation of -1.6 kg m-2 (median -1.4 kg m-2) and a mean relative deviation of -9.5% (-6.2%) and for the clear-sky bias we get a mean absolute deviation of -1.7 kg m-2 (median -1.3 kg m-2) and a mean relative deviation of -10.0% (-6.0%). However, the distributions of the absolute and relative deviations for the sampling error and the clear-sky bias are highly left-skewed and thus the mean value in particular is influenced by the long tails of the distributions. Nevertheless, for the clear-sky bias the obtained values agree well with the findings of Gaffen and Elliott (1993) and Sohn and Bennartz (2008). Since the effect of the clear-sky bias is already included in the sampling error and the results for both errors are very similar, it can be assumed that the spatial and temporal sampling errors play only a minor or negligible role compared to the clear-sky bias.

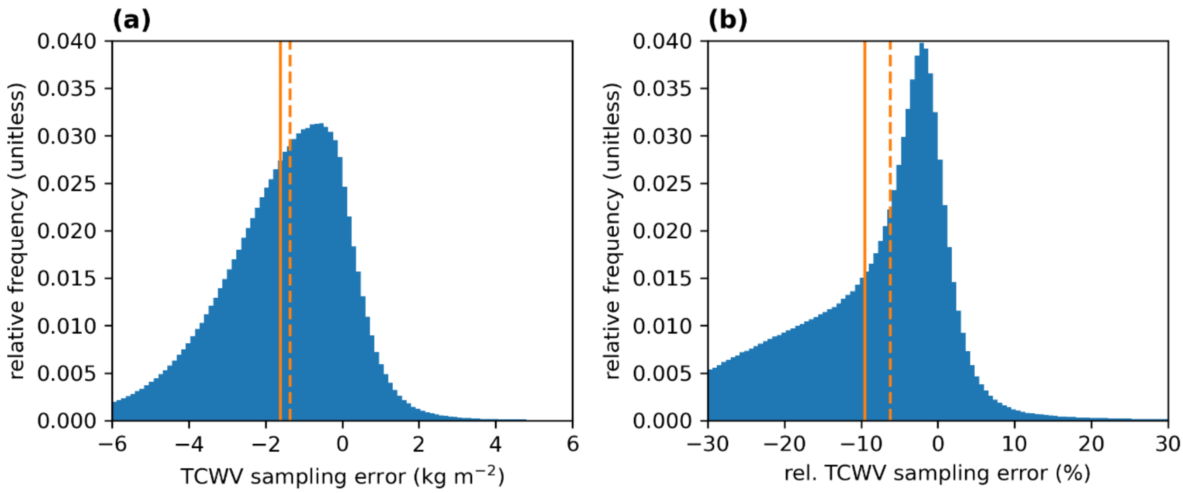

Figure 5. Distributions of the absolute differences ("sampling; Panel a) and relative differences ("sampling=TCWVtrue; Panel b) of the monthly mean differences between clear-sky and all-sky ERA5 data based on the OMI cloud information. The solid and dashed orange line indicate the mean and the median of the distributions, respectively.

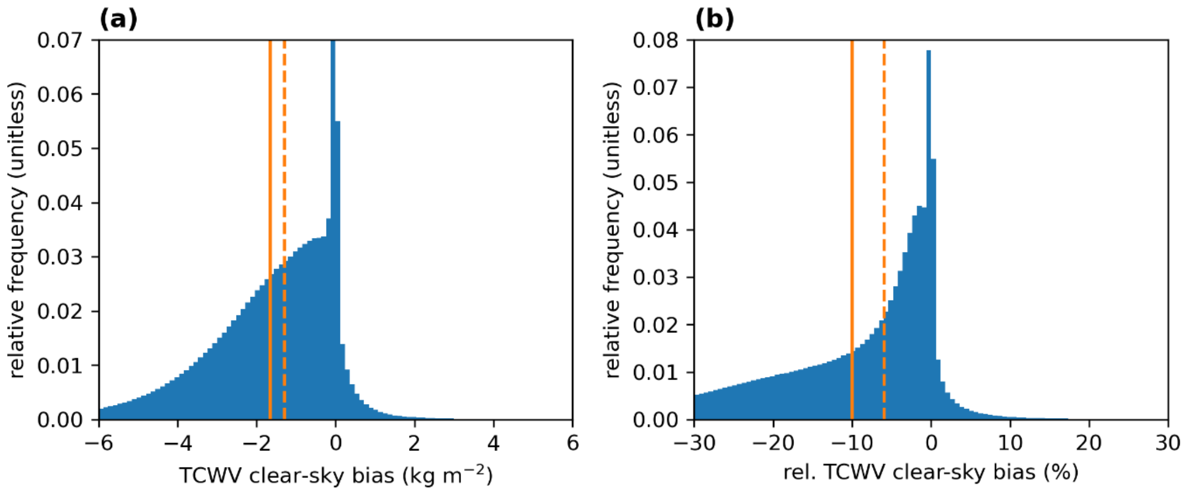

Figure 6. Distributions of the absolute differences ("clear; Panel a) and relative differences ("clear=TCWVall; Panel b) of the monthly mean differences between clear-sky and all-sky ERA5 data based on the OMI cloud information. The solid and dashed orange line indicate the mean and the median of the distributions, respectively.

In addition to the sampling error and the clear-sky bias, we also examined in Appendix C to what extent the monthly means would change if no RA-filter is applied, i.e. if all data of the complete OMI swath were available. It turns out that although deviations arise due to the RA-filter, these deviations are almost an order of magnitude smaller than those of the clear-sky bias and the global distribution of the deviations is mostly noisy. Due to this small influence of the RA-filter, we conclude that the filtered OMI TCWV data are a good representation of the actual TCWV values.

Moreover, we also investigated trends in the clear-sky bias (which has the largest impact on the sampling error) and obtained absolute trends between +-0.04 kg/m^2 per year (and -0.002kg/m^2 per year on global average), which is one order of magnitude smaller than typical TCWV trends (see e.g. Borger et al., 2022).

In conclusion, it is my feeling that the proposed data set has significant defects that are not well understood. 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.

Specific comments

More should be said about the "row anomaly" which affects the OMI observations throughout almost the whole period analysed in this paper. Figure A2 shows that a large fraction of "rows" are discarded. Is it sufficient to discard these rows or could adjacent rows also be affected in some way? What is the impact of this screening on the representativeness of the final observations?

The row anomaly is a dynamic artefact and initially spread from a few isolated rows over a large area of the detector, affecting about 50% of the swath. However, it is observed that it seems to have stabilised or not changed much for a few years (see e.g. Figure 22 in

Schenkeveld et al., 2017). Based on the daily monitoring of the instrument and the rigid row-anomaly screening, we can therefore at least assume that we are filtering the very largest part of the row anomaly to the best of our knowledge, although we cannot say one hundred percent that other rows are slightly affected.

We investigated the extent to which the monthly means would change if the full swath had been taken into account and concluded that the impact is almost an order of magnitude smaller than other uncertainties (e.g. clear-sky bias). The following section is added to the appendix:

Due to the row anomaly filter, approximately 50% of the complete satellite swath of OMI is not considered in the TCWV data set. This raises the question of how much the monthly mean values would differ if the data of the complete swath were available. To investigate this, we follow the same scheme as in Sect. 3 and use the same ERA5 data as a reference. We select the ERA5 data to match the OMI overpass, once applying the row-anomaly filter and once not. However, in both cases the clear-sky filter based on the OMI cloud information is applied (effective cloud fraction < 20%).

Compared to the clear-sky bias, the deviations are much weaker and no particular spatial patterns are discernible in the global distributions except in the deep Pacific tropics and parts of Southeast Asia (see Fig. C1). Furthermore, the histograms for the absolute and relative deviations in Fig. C2 show a normal distribution for both cases with mean values of -0.30 kg m-2 and -2.1% (and for the median -0.23 kg m-2 and -1.1%). Considering the much larger uncertainties of the OMI TCWV retrievals of typically 20% and more and that the clear-sky bias is almost one order of magnitude larger, the obtained deviations are negligible and thus the monthly means from the RA-filtered data are a good representation compared to the monthly means from the data for a full swath, even though only half of the satellite data is actually used.

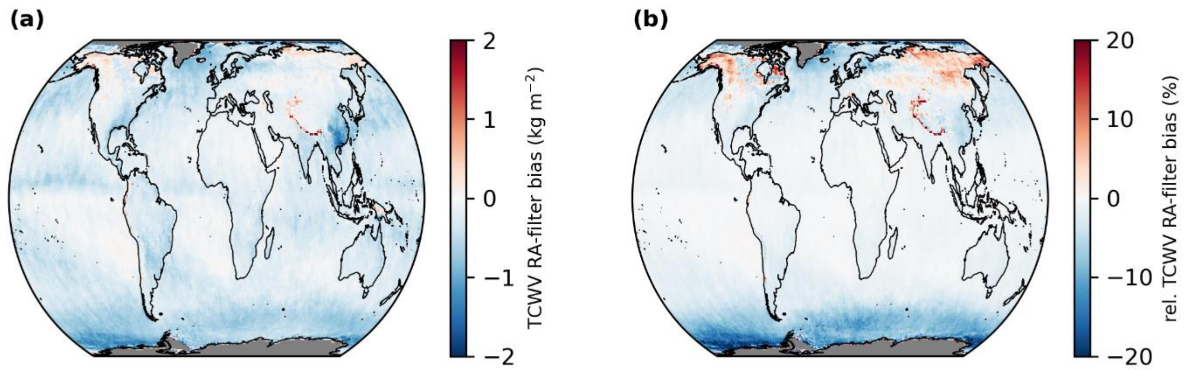

Figure C1. Global distributions of the monthly mean differences between row-anomaly (RA) filtered and full swath ERA5 based on the OMI cloud information for the time range January 2005 to December 2020. Panel (a) depicts absolute differences (i.e. RA-filtered minus full swath) and Panel (b) relative differences (i.e. (RA-filtered minus full swath) / full swath). Grid cells for which no data is available are coloured grey.

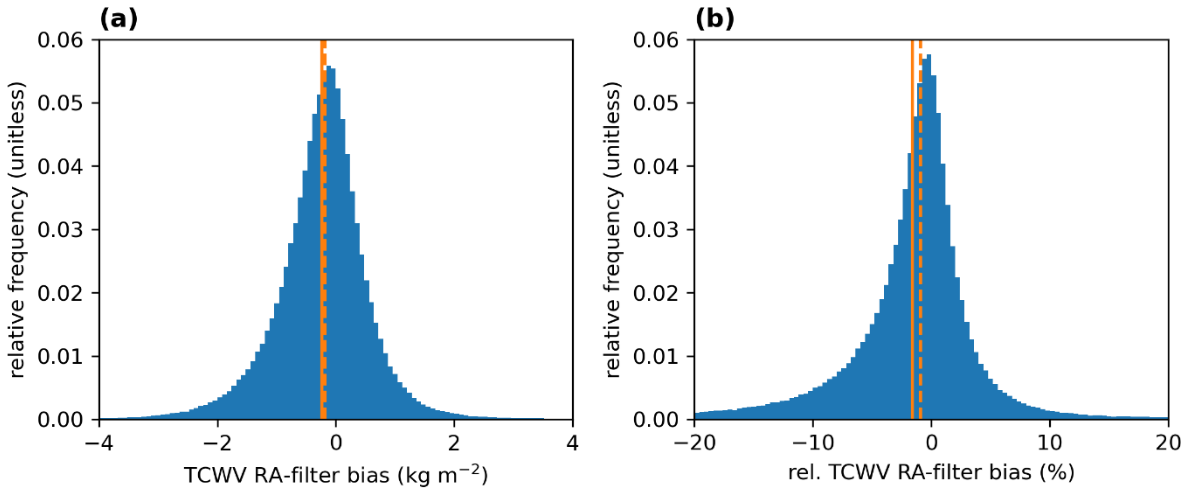

Figure C2. Distributions of the absolute differences (RA-filtered minus full swath; Panel a) and relative differences ((RA-filtered minus full swath) / full swath; Panel b) of the monthly mean differences between RA-filtered and full swath ERA5 data based on the OMI cloud information. The solid and dashed orange line indicate the mean and the median of the distributions, respectively.

So, although about 50% of the orbit is missing, this still covers a swath of about 1300km and is thus still larger than the swaths of GOME-1, SCIAMACHY or GOME-2A (all around 960km) or in the order of magnitude of SSMI (about 1394km). Thus, OMI still achieves a complete coverage of the Earth about every 2-3 days, which should provide enough observational data for a good representativeness in the case of a monthly mean (see also the good agreement with the reference data).
We added this information to the revised manuscript:
So while about 50% of the orbit is missing because of the RA-filter, the remaining data still cover an "effective" swath of about 1300 km and is thus still larger than the swaths of GOME-1, SCIAMACHY, or GOME-2A (all about 1300 km) or of the order of SSM/I (about 1394 km). Thus, OMI still achieves complete coverage of the Earth about every 2-3 days, which should provide enough observational data for good representativeness in case of a monthly mean (see also Appendix C and the good agreement to the reference data in Sect. 4).

Why are two regression methods (OLS and ODR) used? In principle, a single statistic is sufficient, unless the difference of results from the two are discussed, but this is not done in this manuscript. I suggest either to choose one or to better justify the choice of two and analyse the obtained differences.
We decided to use only the ODR, but also to show the results of the PWLF regression in the scatterplots instead.

L77: replace "I and I0" by "I0 and I"
We replaced the terms accordingly.

L129: define also Delta_SCD
We have added that Delta_SCD is the offset between Earthshine and normal SCD.

L144: indicate which fraction of raw data is remaining
If we only take the data that is already filtered according to the row anomaly as a basis, approx. 30% remains. If we take all the data of an orbit as a basis, approx. 12% remain.

However, this also includes pixels above the polar regions for which a spectral analysis is not possible or does not make sense due to the high noise.
We added the following text:
In total, this leaves about 30% of data from an RA-filtered orbit and about 12% of data from a complete orbit.

L154: "ESA Water Vapour CCI climate data record CDR-2" needs a reference
At the moment, no reference for the data set is available yet.

L154: "For the correlation analysis" is misleading or incorrect if referring to regression analysis. Please reword (e.g. For the intercomparison...)
We have changed the phrase accordingly.

L155: add a reference for the ODR method
We have added Cantrell (2008) as a reference.

L155-156: "In the case of the ODR it is necessary to use reasonable ratios of the relative errors of the compared data sets instead of using absolute errors in order to obtain meaningful results". This statement needs to be justified by an adequate explanation or reference.
Based on the descriptions of Cantrell (2008), one sees in equation (5) in his paper that the slope depends on a parameter W_i, which relates the uncertainties w_x and w_y of x and y to each other (see formula below).

$$
\begin{aligned}
& b = \overline{y} - m\ \overline{x} \quad m = \frac{\sum W_i \beta_i V_i}{\sum W_i \beta_i U_i} \\
& \overline{x} = \sum W_i x_i \Big/ \sum W_i \quad \overline{y} = \sum W_i y_i \Big/ \sum W_i \\
& U_i = x_i - \overline{x} \quad V_i = y_i - \overline{y} \quad W_i = \frac{w_{xi} w_{yi}}{w_{xi} + m^2 w_{yi} - 2 m r_i \alpha_i} \\
& \beta_i = W_i \left[\frac{U_i}{w_{yi}} + \frac{m V_i}{w_{xi}} - (m U_i + V_i) \frac{r_i}{\alpha_i} \right] \quad \alpha_i = \sqrt{w_{xi} w_{yi}}
\end{aligned}
\qquad (5)
$$

In the case that the error in y is significantly larger than in x, the ODR approaches ordinary linear regression.

Cantrell, C. A.: Technical Note: Review of methods for linear least-squares fitting of data and application to atmospheric chemistry problems, Atmos. Chem. Phys., 8, 5477–5487, https://doi.org/10.5194/acp-8-5477-2008, 2008.

L156-159: these sentences sound in contradiction with the previous statement. Moreover, the sensitivity of the regression results to the relative errors should be discussed in more detail (e.g. in an Appendix) and the choice of 5%, 10%, and 20% for the three dataset (which appear quite arbitrary) should be clearly motivated/discussed.
To motivate our choice, we have added the following text to the revised version:
Mears et al. (2015) found that the uncertainty of daily microwave TCWV observations for TCWV=10 kg m-2 was around 1 kg m-2 and for TCWV = 60 kg m-2 around 2-4 kg m-2. Hence, we assume that the uncertainty of the RSS data set is 5% or at least 1 kgm−2. For

ERA5 and ESA CDR-2 we can assume similar uncertainties over ocean, since the TCWV values there are also mainly based on microwave observations. Unfortunately, no uncertainties are provided for TCWV over land. Thus, for the sake of simplicity, we assume that the relative errors of the reference data sets over land are twice as high as over ocean, i.e. 10% or at least 2 kg m-2. For the OMI TCWV data set we assume an uncertainty of 20% (Borger et al., 2020), but at least 2kg m-2. We also tested other error assumptions and it turned out that the exact choice of errors is negligible for the regression results as long as the ratio of uncertainties remains similar.

L169-171: “In general the deviations are quite low with values between +/- 2.5 kg/m2” be more specific in quantifying the differences here, e.g. indicate which fraction of data lie in the range of +/-2.5 kg/m2, or use quantiles or other statistics (mean, standard deviation, etc.). Note also that the correlation coefficient is not much relevant when the seasonal variations are included.

We added the information of the mean bias together with the standard deviation for the comparison to every data set and also provide this information for the tropics (-20°N – 20°N) and for the extratropics.

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.

L173-178: Be more quantitative again, here in the comments on Fig. 4. I would also suggest to include the coastlines of Africa and Indonesia in the list of regions with significant positive deviations.

As mentioned above, we now provide the mean bias and the standard deviation. Moreover, we rephrased the sentence:

Consistent with the findings from Fig. 7 highest positive deviations can be found in the tropical Pacific ocean and near the coastlines of South America, Africa, and Indonesia whereas […]

L176: Be more specific on the impact of the “cold tongue” and “too low albedo” on the observed deviations.

The area of the "cold tongue" is often affected by low maritime clouds (cloud top height at approx. 1km). Since the highest water vapour concentration are found in the lower troposphere or boundary layer, deviations in the AMF of the order of 10% can occur even with slightly deviating cloud heights of a few 100m.

In the area of Central America and the west coast of Africa, the albedo is influenced by the absorption by phytoplankton (Kleipool et al., 2008), which may not have been optimally corrected during the creation of the LER or ensures that already low albedo values can lead to further small deviations, which are then again large in relative terms (e.g. with albedo values of 0.05 to 0.04).

We rephrased the text as follows:

In the case of the tropical Pacific ocean the distribution of the systematic positive deviations matches quite well regions of cold water or of the so called "cold tongue" which is frequently affected by low clouds. Since the highest water vapour concentrations occur in the lower troposphere, small deviations of a few 100m in cloud height can have relatively large effects on the AMF. In the case of Central America or Atlantic ocean, a too low albedo due to additional absorption by phytoplankton (Kleipool et al., 2008) could explain the systematic positive deviations.

L182: “the slight overestimation of 3-5%”: it is not clear what these numbers represent exactly. Is it a mean difference (bias)? Is it computed over all data or only a fraction? (Note that a slope of 1.03 does not mean that all the values are 3% higher, this depends also on the intercept value).
Many thanks for this hint! Indeed, we have not expressed our approach clearly enough. By 3-5% overestimation, we are referring to the slope of the fit line. Regarding the y-axis intercept, we will explicitly mention it if it is larger than the minimum assumed uncertainty (1kg over ocean, 2kg over land). For the ocean comparisons, the offsets are less than +-0.25kg/m^2, so they are negligible and thus the slope is sufficient as the sole indicator of over- or underestimation. For the comparisons for the data over land, however, they are systematically higher than the minimum uncertainty, so we have revised the text of the respective comparisons:
For data over land, the picture is different: although the ODR gives similar results for the slope as for data over ocean, the distribution in the 2D histogram (Fig. 9c) shows particularly strong positive deviations of approximately +10 kg m−2 at high TCWV values and an overall systematic offset of around +1.43 kg m−2.Within the PWLF analysis we find a good agreement to the reference data for TCWV values up to about 25 kg m−2 (which represents approximately 74% of all data points) with slopes of around 0.96. However, for higher TCWV values we find distinctive positive overestimations of up to 24%. Nevertheless, even for low TCWV values a systematic offset of approximately +2.52 kg m−2 is obtained.
[…]
Similar to the intercomparison of ERA5, the intercomparison over land (Fig. 11c) shows roughly similar ODR fit results as over ocean, but here we also find striking positive deviations for high TCWV values and an overall positive offset of 2.41 kg m−2. Again, when applying a piecewise linear regression analysis we obtain good agreement with slopes of around 0.95 for TCWV values to about 25 kgm−2 but still a distinctive positive offset of 3.73 kg m−2 for low TCWV values and distinctive overestimations of up to 33% for higher TCWV values, which is even higher than for the comparison to ERA5.

L194: how is the change-point at 26 kg/m2 selected in the piecewise linear regression?
The change point is automatically determined by a non-linear least-squares fit.

L210: satellite measurements in the thermal infrared are NOT available/reliable in cloudy conditions.
We have reworded the sentence as follows:
[…] satellite measurements (or none at all in the thermal infrared) […]

L209-215: I’m not convinced that the ERA5 uncertainty over tropical land areas contributes much to the huge bias observed in the differences (above 10 kg/m2). This idea should be further documented or discarded (also in the Conclusion).
The regions in question are highly affected by quasi-permanent cloud cover, so observations are systematically missing and there may be a clear-sky bias, which can be in the order of a few kg/m^2 (see also Sect. 3 in the revised manuscript). And even if radiances are assimilated into cloudy-sky scenarios, their uncertainty is still large, as the radiative transfer of cloudy pixels is highly complex (e.g. Li et al., 2016). Especially even in the ESA WV_cci CDR these regions are flagged, although MODIS should have enough observations available for good statistics. We conclude that the large deviations in the tropics cannot, of course, be completely attributed to the uncertainties in ERA5, but they are not so small as to be negligible either.

Li, J., Wang, P., Han, H. et al. On the assimilation of satellite sounder data in cloudy skies in numerical weather prediction models. J Meteorol Res 30, 169–182 (2016). https://doi.org/10.1007/s13351-016-5114-2

L227: is there any update on the publication of the ESA CDR-2 data set?
To the best of our knowledge, no publication is available at the moment.

L225-254: Similar comments as for ERA5 apply here to the CDR-2 comparison (lack of statistics, etc.).
See comment above about added statistics.

L261: More details are needed on the linear regression method and significance tests.
For the analysis, we use an ordinary least-squares fit, with the significance test or p-value based on a two-sided Students t-test (see also https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.linregress.html). We added this information in the revised manuscript as follows:
[…] and then calculate temporal trends of these deviations using linear ordinary linear least-squares regression following the approach of Danielczok and Schröder (2017) and Beirle et al. (2018) and assess the significance of the results based on a two-sided Student's t-test.

L274: The stability requirement for water vapour climate data is rather at the level of 0.3 %/decade (GCOS – 112, April 2007).
See comment above in the General Comments section.

Figure 3: the fit results would be more understandable if given as an equation: y = 1.03 x + 0.18 rather than just two numbers.
We have changed the legends in the figures accordingly.

Figure 3: indicate that the OMI results here are over ocean (it is only obvious if one knows that SSM/I data over only over the oceans).
We added in the Figure caption that the results correspond to data over ocean.

Figure 4: add the piecewise linear regression lines (mentioned L194) on the plot.
We have added information of the PWLF regression results in all relevant figures.

Figure 9: the red dashed lines are not visible in the plots.
We revised Figure 9 and removed the dashed red lines.