Reply to Referee #2 (R. Lang)

We would like to thank the referee Rüdiger Lang for the very constructive and helpful comments and questions.

Below we reply to the issues raised by the referee point by point, where
blue sans serif type repeats the reviewer's comments,
black serif type is used for our reply, and
black italic type indicates text from the manuscript, whereby modifications are highlighted in red.

Water vapour retrievals from the visible region (here the red part) of the spectrum contribute to the overall suite of available water vapour remote sensing retrievals by offering a very good accuracy over land and a very acceptable accuracy over ocean under day-light conditions. Like their companion retrievals from the thermal infrared region, VIS-NIR WV retrievals are sensitive to clouds, which have to be screened out, at least to a large extent. The suite of available retrieval methods for UV-VIS-NIR instruments like GOME-1 and 2, and SCIAMACHY (and also OMI) have also the advantage to be relatively free of a priori knowledge of the atmospheric state and can therefore be considered as truly independent measurements both with respect to ground-based measurements and WV concentrations provided by model output.
The provision of a quality controlled and consistently reprocessed merged data-set from the GOME-1 and 2 and SCIAMACHY suite of instruments, already covering two decades of data and (with Metop-B and C) to be continued until the end of the 2020th, therefore provides a very important basis for validation and trend analysis.
The paper by Beirle et al. describes the strategies used to merge the three WV datasets available using a retrieval strategy originally developed by Wagner et al and implemented in the GOME operational processor environment (GDP) by DLR (Grossi et al.). Since the retrieval method, using the 640 nm region, as well as the conversion from SCD to VCDs using AMFs derived from the oxygen B-band, are the same for all instruments, the only remaining issue is to overcome instrument design and operation specific differences and their impact on the climatological WV values derived as “cloud-free” monthly means at global coverage. Cloud screening is carried out by comparing the retrieved ground-pixel averaged mean scattering-height (derived from the O2-AMF) to the theoretical cloud-free height (using pre-scribed “maximum” pressure profiles per instrument above a cloud-free pacific-ocean), while the strategy for merging of the data, overcoming instrument specific design aspects, is to reduce the spatial resolution of GOME-2 and SCIAMACHY to that of the approximately 8 times coarser GOME-1 foot-print and subtract a global (SCIAMACHY) or latitude depended (GOME-2) offset.
The paper is well written and the results are scientifically sound so that I can recommend it for publication in ESS, provided the authors can address the two main issues I have with the presented results and analysis.

We thank the reviewer for his positive general assessment.

I) The reason why the offsets between SCIA and GOME-2 are so different for land and ocean is still very puzzling after having read the paper, since for me it seems that this cannot be explained by the interference of the missing/screened out monthly narrow scan mode (NSM). This is because such instrument operations issues, like most calibration issues, should be largely independent from the surface type, as also Figure A1 - showing the impact of the NSM removal on the mean scan angle - seems to demonstrate. One can only guess that the observed land/ocean difference in the offset between the two instruments WV data might be more related to the SCD retrievals, e.g. in case they use different surface treatments or databases or similar auxiliary data related to different surface types.

The climate product aims to create consistent time series as best as possible. This is achieved by spatial merging of SCIAMACHY and GOME-2 pixels to the coarser pixel size of GOME.
As shown in Fig. 4, the merging of SCIAMACHY pixels result in far better agreement to GOME. But still, there are small (~1-2%), but systematic differences remaining which require an additional offset

correction between the instruments. These offsets have been found to be considerably different for land and ocean, as shown in Fig. 9.
Note, however, that the retrieval settings are identical for the different instruments, i.e. the land/ocean difference cannot be explained by surface treatment or auxiliary data used within the retrieval. Instead, as discussed in section 3.3, the remaining offsets might be caused by different instrument characteristics (like polarization sensitivity or spectral resolution) or imperfect spatial re-sampling (since the SCIAMACHY pixels cannot be merged to GOME pixel size exactly).
Most of all, however, we consider the difference in local time to be the most likely cause for the remaining offsets, as we discuss in section 3.3. This is supported by the finding that the offsets GOME-SCIAMACHY and GOME2-SCIAMACHY have opposite signs, i.e. the changes over instruments are monotonous in local time. This is probably caused by a systematic change of cloud conditions between 9:30, 10:00, and 10:30 local time. Note that (a) even small changes of cloud conditions might affect the TCWV retrieval, and (b) the actual magnitude and sign of such a change is difficult to quantify, as a change in O_2 affects both the AMF correction and the cloud flag.
We have extended this discussion in section 3.3 of the revised manuscript to clarify the line of argumentation.

The discussion of scan angle effects and the need for a smoothed TCWV over ocean was obviously not clear enough in the manuscript. We have thus largely revised the respective section (formerly 4.4, now 3.4, as the smoothed TCWV is part of the Climate product, thus its description should be part of section 3 on "The climate product") and Appendix A on mean scan angles.
In short, (a) mean scan angles vary in monthly means, but (b) are about 0 in the long term average, except (c) in cases where the orbital patterns are not regular, such as for GOME over the Himalaya or for GOME-2 where NSM orbits are periodically missing.
The effect of such a systematic scan angle bias is now very different over land (where it has almost no effect) and over ocean (where it has a quite strong effect). This has been shown in Figure 1 of Grossi et al. (2015). Thus, orbital patterns can be recognized in the mean difference map between GOME-2 and SCIAMACHY, but only over ocean. For the offset correction, this effect is eliminated by applying just a latitude-dependent offset over ocean for GOME-2. In the final product, $TCWV_{smooth\_ocean}$ is provided, where the remaining (very small, but) systematic orbital structures over ocean are removed.

II) Statistical effects from monthly averages in combination with cloud screening are not considered. This is an oversight which is often made when evaluating monthly mean data-series, especially for species with such high temporal variability. If the cloud screening per instrument, and during the course of one month, is different due to differences in spatial sampling or due to the different coverage (e.g. by the SCIAMACHY alternate limb/nadir sounding), this can lead to significant monthly differences in the averaged result, simply because the monthly average covers different days, and since the day-to-day WV variations at the same point can be very large. This may in particular play a significant role for the gaps analysis (Section 4.2), and in particular in the tropics where the temporal variability (in absolute terms) is very high. So a comparison of the WV distribution around the mean in selected months (or for each month for a selected year) should be carried out, for analysing this effect on the presented monthly mean values.

We agree that sampling effects are important for the evaluation of the representativeness of the monthly means provided in the climate product.
Part of the effects described by the reviewer, i.e. differences between instruments caused by different spatiotemporal sampling, are shown in the comparison between GOME and SCIAMACHY for all available measurements (Fig. 3a) versus the comparison of coincident measurements only (Fig. 3b). Note that the final offset between GOME and SCIAMACHY with reduced resolution (Fig. 3c) is determined for coincident measurements during the overlap period, i.e. there are no sampling effects (except the small time shift of half an hour). For the offset between SCIAMACHY and GOME-2, on the other side, a very long overlap period of more than 5 years is available so that statistical sampling effects cancel out.

In order to provide additional information on the representativeness of the provided monthly means, we have processed a new version (v2.2) of the climate product which now also includes the monthly standard deviation (std) of TCWV for each 1° pixel and month. This reflects the day-to-day variability of TCWV within a month.

In addition, the standard error of the mean can now be derived which represents the precision of the climate product.
The standard deviation and standard error of the mean are introduced and discussed in a new section (3.5) of the revised manuscript. Additional figures of std and standard error are provided in a new Appendix D.

The systematic impact of cloud screening is a general problem of satellite remote sensing, and particularly important for water vapor. We thus add the following to section 4.3 ("Known issues: accuracy"):
In addition, the selection of cloud free observations corresponds to generally dryer atmospheric conditions, which likely results in low biased means. This effect is unavoidable for water vapor retrievals from satellite measurements in the visible range, where clouded scenes have to be masked out.

Specific comments:
Section 2.4, p.4, last paragraph: But this "consistency" could also be achieved by, e.g., using the same independent pixel approach (e.g. FRESCO+) applied to the O2-B band.
We agree that consistent treatment of cloud effects might also be realized by other methods, but the simple O_2 threshold approach does it implicitely.
We modify the respective sentence to
The advantage of this approach ~~(and the reason to stick to it for the climate product)~~ is that ~~is allows for~~ it directly provides a simple, but consistent ~~cloud~~ treatment of cloud effects across the different satellite instruments, ...

And it is not clear to me why a DEM is not used for clearly separating a mountain or a cloud (using the O2-AMF approach)?

We agree that the very simple procedure (using the measured O_2 SCD in relation to the maximum O_2 SCD over the Pacific for both AMF correction and cloud masking) might generally be replaced by a more complex method which takes the surface elevation into account. This would probably improve the accuracy of the TCWV over elevated terrain and would overcome the persistent gaps over high mountains.
However, within this study, we follow the simple approach, in order to be consistent with the GDP algorithm documented in Grossi et al. (2015), and because a modified treatment of the O_2 threshold would require a major reprocessing of the complete data set, including updated testing and validation of the resulting product.

In the revised manuscript, we now point out at beginning of section 2 that the current retrieval includes some simplifications, which however affect all instruments likewise and thus do not impair trend analyses.

Figure 2: Why is the corresponding GOME-2 situation not shown here? I think this would be very useful for the reader and for the user of the data-set as well.

We appreciate the reviewer's comment and acted on his suggestion:
We have changed Fig. 2 by considering the 1st of June 2009 now (instead of 2003), where measurements from GOME, SCIAMACHY, and GOME-2 are available over the Atlantic.
This figure illustrates the merging patterns of small pixels into GOME-like pixels for both SCIAMACHY and GOME-2. In addition, it shows the reduction of the GOME-2 swath width, and illustrates why a direct comparison of coincident measurements between GOME-2 and SCIAMACHY (with shifted orbital patterns) is not as meaningful as between SCIAMACHY and GOME.
The reference to and discussion of Fig. 2 is revised accordingly in the manuscript.

Editorials:
Section 1, p.2, l.7: I would add "ground-based" here since GPS can also be satellite.

We have modified the sentence accordingly.

Section 1, p.2, l.9+10: Radio occultation (upper troposphere/stratosphere) - should be mentioned here, since RO is meanwhile a key-contributor to NWP for WV.

We thank the reviewer for this hint and have added the following sentence to the introduction:
In addition, radio occultation (RO) is an accurate method to determine water vapor concentrations in the upper troposphere/lower stratosphere region and is a key contributor to numerical weather prediction.

Section 1, p.2, l15: I would add here that therefore these measurements are "complementary to the MW, RO and IR derived climate data-sets, which are sensitive only to specific surfaces or altitude ranges".

We appreciate the reviewer's proposal. We have added the following to the introduction:
Thus, TCWV products from satellite observations in the red spectral range are a valuable complement to MW, IR and RO water vapor products, which are sensitive only to specific surfaces or altitude ranges. TCWV products derived from GOME, SCIAMACHY and GOME-2 have already been used to investigate the water vapor evolution over time on global scale, e.g. the effects of El Nino (Wagner et al., 2005; Loyola et al., 2006) or trends (Wagner et al., 2006; Mieruch et al., 2008, 2011, 2014).

Section 1, p.2, l19: Isn't GDP in the version 4.8, meanwhile?

We have modified the respective sentence to
The TCWV retrieval implemented in the operational GOME-2 data processor (GDP) (from version 4.7 on) ...

Section 1, p.2, l29: I would indicate here already what the difference in the footprint is, by adding "to the smaller swath with larger ground footprint" ... or similar.

We have modified the sentence as follows:
This consistency is reached by
a) spatial merging of the smaller SCIAMACHY and GOME-2 pixels (60/80 km across track) to the GOME pixel width (320 km) and
b) limiting the broader GOME-2 swath (1920 km) to that of GOME and SCIAMACHY (960 km).

Section 1, p.2, l31: I would add that this is why this makes it also a truly independent data-set for model evaluation and evaluation of other WV climatologies.

We thank the reviewer for this suggestion.
We have added the sentence
The resulting climate product is a valuable, independent data set for model evaluation, comparison to other water vapor products, and trend analyses.

Section 2.2, p.3, l23-24: Revise for better reading.
General: "saturation effect" should be explained somewhere, for the non-retrieval DOAS expert as being another name for the "non-linearity" effect in the spectral absorption.

We have completely revised the respective section 2.2 as follows:.
2.2 Correction of nonlinearity in spectral absorption
The spectrally fine structured absorption bands of water vapor are not resolved by the considered satellite instruments. Consequently, the relationship between the actual TCWV and the retrieved H_2O SCD becomes nonlinear. The same holds for O_2.
This effect can be simply modelled based on synthetic spectra as described in Wagner et al. (2003, 2006) for H_2O and O_2, respectively. For the GDP and the climate retrieval, the H_2O and O_2 SCDs resulting from the DOAS analysis are corrected accordingly for nonlinearities in spectral absorption. This correction is also denoted as "saturation correction" in Wagner et al. (2003).

Section 2.3, p.4, l5: "by" Wang et al.

Corrected.

Figure 4: The caption misses a description of the different SCIAMACHY data-sets displayed.

We have revised the caption to Fig. 4 to
Top: Zonal mean TCWV for GOME and SCIAMACHY (in original as well as reduced resolution) as function of latitude.

Bottom: Differences of zonal mean TCWV between GOME and SCIAMACHY in original (light) and reduced (dark) resolution, separately for land (orange) and ocean (blue).

References: Danielczok, A. and Schröder, M. - The reference is not complete here, although provided in the text.

The validation report by Danielczok and Schröder is now publically available via the ESA GOME web page:
https://earth.esa.int/documents/700255/1525725/GOME_EVL_L3_ValRep_final/db7e72c3-044d-4236-9dee-d88405b89ef0
We have added the respective link to the reference.

Figure A1: Caption should state that it is a "monthly" mean.

Figure A1 actually displays the scan angles averaged over the full available time period for each instrument. This is now specified in the figure caption.