Interactive comment on "Validation of GOME-2/MetOp-A total water vapour column using reference radiosonde data from GRUAN network" by M. Antón et al.

<u>Author's answer to Anonymous Referee #2</u>

The authors greatly acknowledge the anonymous reviewer <u>(Referee #2)</u> for carefully reading the manuscript and providing constructive comments.

General comments

Referee Comment (RC): The authors have made a serious attempt with the writing of this paper however I remain unconvinced as to the results they have chosen to present. Since the main goal is to validate, the main side-goal is to provide with a clear numerical assessment of this validation, which is unfortunately not the case, as is seen from particular comments I have annotated the pdf with. The abstract is well-written and the text well-structured, but the conclusions need to be completely re-written. Also, a clear message should be given out: is the cloud treatment in GDP4.6/4.7 at fault or the SZA treatment in GDP4.6/4.7 at fault for the differences seen? Contrary to what one usually sees, the authors have concluded that the medium to low SZA-associated GOME-2 TWVCs are the worst. Maybe a consultation with the GDP4.6/4.7 algorithm people might shed some light into the reasons behind this effect.

Author's response (AR): We have addressed all reviewer's comments annotated in the pdf as can be seen in the next pages. The conclusions have been completely rewritten giving first the actual numerical findings. Additionally, a clear message is reported for the potential users of GOME-2 TWVC data. Thus, the GOME-2 TWVC data show the smallest differences against the reference sounding data for cloud-free satellite scenes (cloud fraction <5%) with low solar zenith angles (SZA<50º). We are sorry for the misunderstanding with the results given in the subsection 4.4 which could suggest a worse behavior of satellite data for low SZA values. This issue has been widely clarified.

Comment in the annotated text

* Page 2: Abstract

RC #1: Which cut-off value did you use?

AR: For our study, we have considered as cloud-free cases those with CF smaller than 5%. We have included this information in the abstract.

RC #2 and #3: I am assuming that the 1400 cases [544 cloud-free] are referring to all six locations together, right? if so, then the comment here about analyzing all six datasets together is misleading. I suggest to remove it.

AR: Following the reviewer's suggestion, we have removed this sentence.

RC #4: How does "in absolute term" and "19%" differences go together? I do not follow what you mean here. What is this 19%?.

AR: This figure (19%) is the mean absolute bias error (MABE) given by equation 4 in the AMTD paper (in the revised version of the manuscript is the equation 5). We have rephrased this sentence in order to clarify this issue.

RC #5: Hence, can the data be depended upon or not? you mention a lot of problems with clouds. Should one even look at the GOME-2 TWVC data or are the 10% to -20% differences normal for TWVC data? compared to other datasets? Sciamachy maybe? Is GOME-2 better? worse?

AR: We have re-phased the last sentence of the abstract to state that the sounding-satellite differences obtained in our validation for cloud-free conditions with SZA below 50º (±10%) can be considered as a good result for satellite retrievals.

* Page 3: Section 1. Introduction.

RC #1: YOU mean Grossi, I assume.

AR: Yes, sorry for the typo. We have corrected it.

RC #2: What is EUMETSAT? acronym needed.

AR: Following to the reviewer' suggestion, we have included in the text the whole definition of EUMETSAT.

RC #3: missions

AR: We have changed "mission" by "missions".

* Page 4: Section 1. Introduction.

RC #1: ..., the GCOS Reference...

AR: We have added the article "the" before "GCOS References….".

* Page 6:

RC #1: Do I understand correctly that your data are analysed with GDP4.6 up to july 2013 and GDP4.7 from July 2013 onwards? did this fact not introduce any jumps/biases/etc in the time series? since you do not have any time series plots, I cannot judge, but should the east-west correction be evident in the new dataset?
This issue is rather serious, i.e. how does your comparison perform if you exclude data from july 2013 onwards? and you keep only GDP4.6?

AR: We understand the reviewer's query. Nevertheless, satellite TWVC data derived from both GDP 4.6 and GDP 4.7 are very similar. The main difference between the two algorithms is the empirical correction applied by GDP 4.7 for removing the scan angle dependency found for the outermost west satellite pixels using GDP 4.6. This correction only affects a limited number of cases and., thus, its effect in the reported results is no

significant. Nevertheless, following also the suggestion given by the reviewer #1, we have decided to remove from the cloud analysis those GDP 4.6 data with VZA higher than +30º. The number of deleted data is small (81 out of 1400) and the new results and plots are very similar to the previous one. We have included in the revised version of the manuscript a brief paragraph explaining this issue.

RC #2 and #3: Where these measurements used in the past for validation purposes? not specifically GOME-2, in general, validation. You should include a paragraph mentioning these other studies. Are these data quality assured by any formal organization [e.g. WMO?]. It is beneficiary to the reader to include such information.
Furthermore, for the reader who has not worked with balloon data before, some more information on how the information is acquired by the sonde is important I believe. Please add a paragraph making note of that within this section. Also, do these data come with any auxiliary information? or do you simply have a TWVC value and its associated error?

AR: TWVC data from Vaisala RS92 sondes have been used in numerous inter-comparison exercises against other techniques such as sun-photometry, microwave radiometry, GPS (e.g, Schneider et al., 2010; Buehler et al., 2012; Pérez-Ramírez et al., 2014), and other satellite instruments such as SCIAMACHY (e.g., du Piesanie et al., 2013) and MODIS (Diedrich et al., 2014). In addition, RS92 sondes has also participated in several radiosonde inter-comparison campaigns promoted by WMO in order to asses its accuracy (e.g., Nash et al., 2011). All these papers highlight the high quality of the Vaisala RS92 sonde. According to the reviewer's suggestion, a comment about these issues has been included in the revised version of the manuscript.
On the other hand, detailed information about how the radiosonde records the measurements and about the data contained in the RS92-GDP has been also included in this section.

--- Buehler et al. (2012): A multi-instrument comparison of integrated water vapour measurements at a high latitude site, Atmos. Chem. Phys., 12, 10925–10943, doi:10.5194/acp-12-10925-2012.
--- Diedrich et al. (2014): Retrieval of daytime total columnar water vapour from MODIS measurements over land surfaces, Atmos. Meas. Tech. Discuss., 7, 7753–7792.
--- du Piesanie et al. (2013): Validation of two independent retrievals of SCIAMACHY water vapour columns using radiosonde data, Atmos. Meas. Tech., 6, 2925–2940.
--- Nash et al. (2011).: WMO intercomparison of high quality radiosonde systems, Tech. Rep. WMO/TD-No. 1580, World Meteorological Organization.
--- Pérez-Ramírez et al. (2014): Evaluation of AERONET precipitable water vapour versus microwave radiometry, GPS, and radiosondes at ARM sites, J. Geophys. Res., 119, 9596–9613, doi: 10.1002/2014JD021730.
--- Schneider et al. (2010): Continuous quality assessment of atmospheric water vapour measurement techniques: FTIR, Cimel, MFRSR, GPS, and Vaisala RS92, Atmos. Meas. Tech., 3, 323–338, doi:10.5194/amt-3-323-2010.

RC #4 and #5: Maybe a website for this network? is the data public? if not, how did you acquire it? under which premise/project etc?

AR: The reviewer is right. We have specified in the text the website for GRUAN. Additionally, we have written a sentence indicating that the GRUAN data are freely disseminated through its website.

RC #6: What does "contrasted" mean here? sentence needs re-writing.

AR: We have changed "contrasted quality" by "proven quality".

* Page 7:

RC #1: 4% constant for all clouds, sza, seasons, times, other parameters in general? I would suggest to expand a bit on the errors and reproduce some of the information in Dirksen et al., 2014 for the benefit of the reader.

AR: Following the reviewer's suggestion, we have included in this subsection the information about the determination of the relative error associated with the sounding TWVC data. In the previous version of the manuscrip, this information was included in the subsection 4.4 "Dependence of the differences on reference TWVC data". We agree with the reviewer in the convenience of moving this information forward in the manuscript.

RC #2: I do not find this equation necessary, the integration of the mixing ratio for obtaining total column is rather standard mathematics. I leave it to the discretion of the authors though to keep or discard.

AR: According to the reviewer's suggestion, we have removed this equation.

RC #3: Does this mean +/- 2h? i.e. a window of 4h in total? isn't this a bit too wide for water vapour? also for the 100 km of the radius of search, isn't it a bit too much for water?

AR: Yes, +/- 2 hours which means a maximum of 2 hours around the satellite overpass time. That is to say, the time difference between the sonde's launch and the satellite overpass must be smaller than 2 hours. We think that it is a reasonable time criterion. Other inter-comparison works between radiosonde and satellite data use an identical time (e.g., Diedrich et al., 2014) and even longer (e.g., du Piesanie et al., 2013).
Regarding the spatial criterion, although the upper limit is 100 km, most cases are significantly below this distance. Thus, the percentiles 50, 75 and 90 are 18, 29 and 51 km, respectively. This information has been included in the manuscript.

RC #4: Are you implying that you have taken completely cloud-free GOME-2 data here? this sentence is mis-leading. Explain exactly which "clouds" the H_2O flag gets rid of.

AR: We are sorry for the misunderstanding. GDP does not provide TWVC data when the product of cloud fraction and cloud top albedo exceeds 0.6 (anomalously high cloud top reflection) or when the O_2 absorption is too small. These cases correspond to heavy cloudy conditions (large fraction of pixel covered by clouds and, simultaneously, with

high cloud albedo). The TWVC data for the remaining cloudy cases are provided by the GOME-2 algorithm. In order to clarify this issue, we have re-phased the sentences.

RC #5: Why did you choose those 6 stations? these are all NH mid- and high latitude stations. You should explain the reasoning.

AR: RS92 GRUAN Data Product is currently available only for 14 GRUAN stations. Six out of these 14 sites present coincident data with respect to GOME-2 overpass following the temporal criteria (+/- 2 hours) applied in our study. This information has been added to the manuscript.

* Page 8:

RC #1: Again, I find all these equations unnecessary and rudimentary. But most of all, I am surprised that the measurement error is not included in your statistics! Since both balloon and satellite data provide an error estimate, how do you justify not using it in the nominal statistical analysis equations? I expect either for you to do so and change your text or to provide examples showing that using the errors does not significantly alter the findings of this validation

AR: The reviewer is right. Thank you very much for this comment. In the revised version of the manuscript, we have used the uncertainties of radiosonde and satellite data to obtain a combined uncertainty for the relative differences (RD) as:

$$\sigma(RD) = \left[\left(\frac{\partial RD}{\partial Sat} \right)^2 \sigma(Sat)^2 + \left(\frac{\partial RD}{\partial Rad} \right)^2 \sigma(Rad)^2 \right]^{1/2}$$

From the relative differences and their combined uncertainties, it is possible to obtain the weighted mean bias error (MBEw) and the weighed mean absolute bias error (MABEw) as:

$$\text{MBEw} = \frac{\sum_{i=1}^{N} RD_i \cdot \frac{1}{\sigma^2(RD_i)}}{\sum_{i=1}^{N} \frac{1}{\sigma^2(RD_i)}}$$

$$\text{MABEw} = \frac{\sum_{i=1}^{N} |RD_i| \cdot \frac{1}{\sigma^2(RD_i)}}{\sum_{i=1}^{N} \frac{1}{\sigma^2(RD_i)}} \, ,$$

The uncertainty of MBEw and MABEw parameters is characterized by

$$u = \frac{\sum_{i=1}^{N} \frac{1}{\sigma(RD_i)}}{\sum_{i=1}^{N} \frac{1}{\sigma^2(RD_i)}}$$

The new results using these weighted averages exhibit some differences wit respect to the previous results, showing a slight improvement (relative differences smaller in

absolute term). Nevertheless, the patterns shown in all plots are almost identical and, thus, almost all comments reported in the paper follow being valid.

* Page 9:

RC #1: This sentence is quite troublesome and provides, in my opinion, a false image of the data. The MBE parameters turn from positive to negative in two of the six cases, i.e. 33,333% of your data sample, depending on your cloudiness situation. Furthermore, two other stations, i.e. a further 33,333% of your data, do not have cloud-free cases. Hence, you reach a conclusion on the over-estimation of GOME-2 data based on only two stations. This point needs clarification and more discussion. It is well and true to use MABE instead of MBE but if there are systematic oscillations in the sign of MBE, using MABE hides this information, and possible reasons behind the differences between balloon and satellite data. In general, Table 2 should be discussed more.

AR: Following the reviewer's suggestion, we have re-phased these sentences and also discussed more the results reported in Table2. We would like to empathize here that MBEw values for cloud-free cases are substantially smaller (closer to zero) than for all-sky conditions and even turns to a positive value for one station. These results indicate that GOME-2 TWVC data underestimate the reference balloon-borne measurements for cloudy conditions as is verified in the Subsections 4.2 and 4.3.

RC #3: Where do those cloudy cases originate from? is the H_2O flag not sensitive enough? don't the balloon data themselves provide a cloudiness marker of some kind? why isn't that used?.

AR: As was commented in a previous point, the "remaining" cloudy cases correspond mainly to those cases where the product of cloud fraction and cloud top albedo is below 0.6. The TWVC data for these cloudy cases are provided by the GOME-2 algorithm, and thus, they are evaluated in our validation. The GRUAN data do not provide any cloud markers.

RC #4: What is the average CLF for the rest of the 61% of the cases?.

AR: The average (+/- one standard deviation) of the CF values for the 61% of the cases is (53±32)%. This information has been included in the text.

RC #5: So, is there a bias in the GOME-2 TWVC with respect to the reference sounding data [see end of your previous paragraph] or not? Your conclusions are unclear and so far, no comment has been made as to the origin of the differences between the two datasets. Are these differences you find explained by the error related to either dataset?

AR: Yes, our results evidence a notable bias in the GOME-2 TWVC with respect to the reference sounding data which can be partially associated with the cloudy conditions during the satellite overpass. Thus, GOME-2 TWVC data significantly underestimate the reference balloon-borne measurements for cloudy conditions. Regarding the uncertainties obtained for the averages, we have included a paragraph in the manuscript with a wide explanation. More than 98% of the relative differences between GOME-2 and GRUAN TWVC data are within the uncertainty reported for each station. From this

result, it must be emphasize that almost all relative differences are explained by the combined errors of sounding and satellite datasets.

RC #6: From this Figure I would assume that GRUAN over-estimates for high TWVC values. Certainly not an across-all-values bias.

AC: Yes, indeed, as the reviewer can see in Figure 8, the GOME-2 underestimate the reference GRUAN data for high TWVC values. We have included a comment about this issue in the revised version of the manuscript.

* Page 10:

RC #1: So, are you suggesting to possible users of the GOME2 and GRUAN water data not to use them if CLF is larger than 5%? an r-squared of 0.7 is not bad at all. But what is your assessment? this should be clearly stated here.

AC: The reference data used in our validation (GRUAN data) are valid for all sky-conditions. By contrast, GOME-2 data show a substantial bias with respect these reference values when satellite scenes are contaminated with some degree of cloudiness. Therefore, potential users of the GOME-2 TWVC data should bear in mind these results. This comment has been added to the revised version of the manuscript.

RC #2: This SZA is, I assume, the satellite SZA and not the GRUAN SZA, right? please state accordingly in the text as well [you note it in the figure caption].

AC: Yes, it is satellite SZA. According to the reviewer's suggestion, we have included "satellite" before SZA throughout the text.

RC #3: And the remainder 30% [100-31-39] is what? maybe this is a typo error and you mean 61% instead of 31%?

AC: We work with cloud-free cases (CF>5%) and cloudy cases (CF>50%). The percentages of cloud-free and cloudy cases are 31% and 39%, respectively. Hence, the percentages of the remaining cases (5%<CF>50%) is 30%.

RC #4:... the standard deviation error....

AC: In the revised version of the manuscript, we do not include error bars in the plots due to the huge uncertainties obtained with the new error analysis.

RC #5-6:...... similar patterns, showing....

AC: We have corrected this typo.

* Page 11:

RC #1: You are referencing here an ATBD report for which you do not give a link for someone to download from. The DLR site, http://atmos.eoc.dlr.de/gome2/documentation.html, includes an ATBD with similar

characteristics to the ones you reference, but from year 2011? from the times of GDP4.5 still. I am assuming that you are referencing a new ATBD which is not included in the DLR site?. Maybe the new ACPD paper is a better source of information? in any case, an active link to the appropriate ATBD is paramount.

AC: We have included the active link for the ATBD report (2011). The ATBD report (2013) is not available in the GOME website.

RC #2: This finding is, in my opinion, very important to show and to stress. Not only do you reveal a SZA dependency of the GOME-2 data but also a seasonal dependence? if you have space issues, delete the equations and add another figure here, I suggest six seasonal lines, one per station you analyse.

AC: Following the reviewer's suggestion, we have added to the manuscript a new plot (please see below) which shows the evolution of the monthly averages of the weighted mean relative difference for five out of six studied sites. These averages are determined only for those months with more than 10 available pairs of sounding-satellite data. It can be seen that the satellite observations remarkably underestimate the sounding data in spring-summer months, while this underestimation clearly decreases (even in some stations turns to overestimation) for the autumn-winter months.
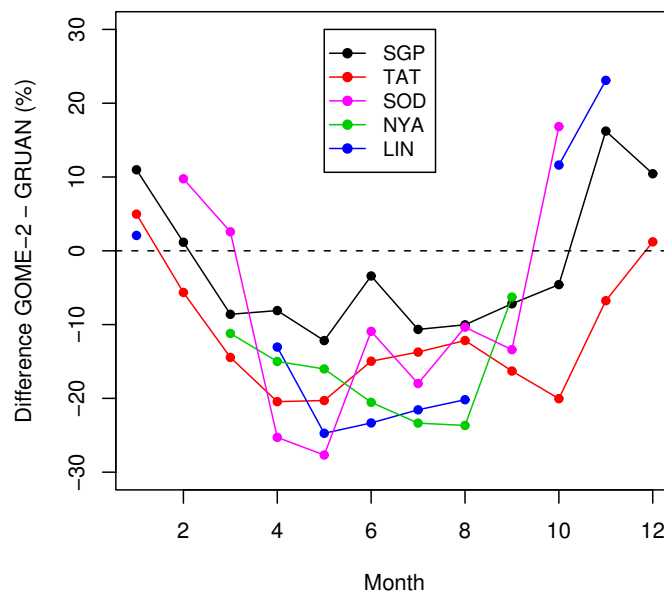
* <u>Page 12</u>:

RC #1 and #2: So, there is a dual problem, both a dependency on CLF and another one of SZA. Are these entirely decoupled? is there a way to decouple them and attribute a percentage of the differences between sonde and satellite instrument to each source of dependency?.

AC: When cloud-free cases are selected, we can evaluate the SZA effects on the relative differences between sonde and satellite data (see Figure 2). To evaluate the

clouds effects, we should select a great number of cases with different sky conditions but with identical SZA. This is not possible in our study due to the limited number of cases. We have tried to analyze the cloudy effects dividing our whole dataset in cases with SZA above/below 50º.

RC #3: you mean "making" maybe?

AC: We have changed "being" by "making".

**RC #4:** you mean "shows" maybe?

AC: We have changed "exhibits" by "shows".

RC #5: You definitely need to make a comment on the whys of this finding!

AC: In the last paragraph of that subsection, we explain that the strong influence of clouds in the relative differences between sonde and satellite data is mainly related to the shielding effect which affects satellite TWVC retrieval.

RC #6: How does this plot change when you use only the cases with CLF<5%? I would like to see that.

AC: The new plot using weighted relative differences (new Figure 7 in the revised version of the manuscript, also left plot in the below figure) has changed with respect to the same plot on the AMTD version using "normal" relative differences (old Figure 6). Now the three curves show an evident decrease of the differences with the decrease of the CTP values. The below figure (right plot) shows that the curve for CF <5% (in red) which displays a similar pattern but with relative differences close to zero for high CTP values.
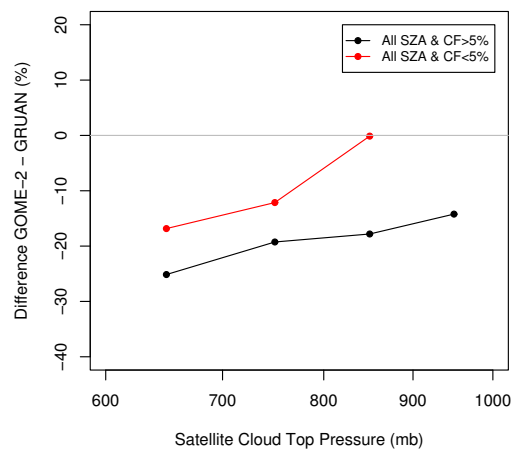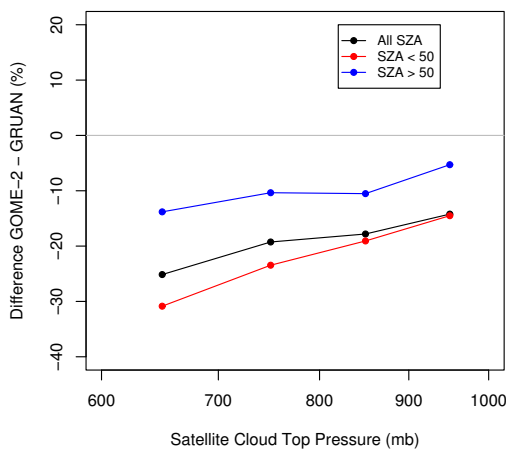
RC #1: Is there an actual value you can quote? from the Mieruch et al papers you reference? or other sources, of course...

AC: We are sorry but we do not find in literature a value to quote the impact of the shielding effect on the satellite TWVC.

RC #2: So, GOME-2 and SCIA TWVC agree in their CTP dependence but not the CLF/CTA/SZA dependence?

AC: Yes, the main reason is that GDP 4.5-4.6 algorithms do not apply any cloud correction method for those TWVC retrieved during cloudy conditions. Therefore, it is expected that the TWVC data derived from the GDP algorithm presents a larger dependence on cloud properties than other satellite retrieval algorithms with some implemented cloud correction method such as SCIAMACHY.

RC #3: I suggest you reference the more appropriate 2014 ATBD paper and discussion therein.

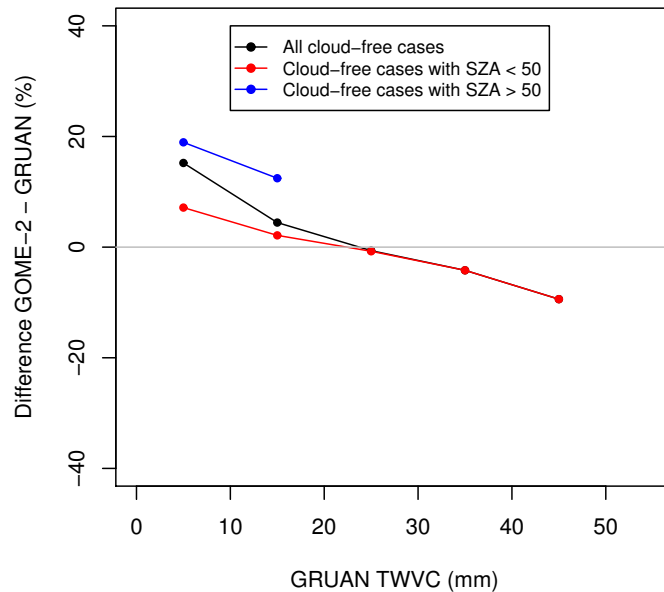**AC:** Following to the reviewer' suggestion, we have included here the reference to the Grossi paper.

RC #1 and #2: Why is that do you think?.

AC: That variation in the relative differences for small and large TWVC values is mainly related to the SZA effect as is explained in the next lines in the manuscript.

RC #3: So, are you advising people to only use GOME-2 TWVC data when SZA>50 degrees and CLF<10%?

AC: No, we are sorry for the misunderstanding. Our advice would be to use GOME-2 TWVC data only for cloud-free scenes (CF<5%) with a satellite SZA smaller than 50º. To clarify this issue, we have determined the relative differences between sounding and satellite data as a function of the reference GRUAN TWVC values for opposite SZA conditions, but exclusively using cloud-free cases (please see below). It can be seen that those cases with SZA above 50º clearly overestimates the reference GRUAN data. By contrast, the relative differences for SZA cases below 50 show values smaller than 10% for the whole range of TWVC values. We have modified the Figure 8, adding this new plot in the revised version of the manuscript.

RC #4: Finally, some comments on errors. Can't this paragraph move higher up in the text? and also something similar for the satellite data to be included?.

AC: According to the reviewer's suggestion we have moved the comments of the GRUAN errors to Subsection 2.2. Additionally, we have included in the Subsection 2.1 a paragraph describing the budget error of GOME-2 TWVC data.

RC #5 and #6: A very strong statement which, however, does not add to our prior knowledge. That satellite data have uncertainties, this was known long ago. How can you quantify them, or suggest improvements to the algorithm scientists based on your findings or suggest an appropriate error characterisation for usage with TWVC data, that should be the key result of such a validation work.

AC: The reviewer is right. We have removed this sentence from the text.

* __Page 15__:

RC #1: I would suggest to first give the actual numerical findings in the conclusions and then to start the various other points. It makes reading the paper by people who wish to focus on the results easier.

AC: The conclusions have been completely rewritten giving first the actual numerical findings.

RC #2: Not so detailed, in my opinion, I suggest to remove this word.

AC: The word "detailed" has been removed.

RC #3: ... between the satellite viewing geomtery and the satellite cloud parameters...

AC: Done

RC #4: .. provided...

AC: Done

RC #5: ... produce...

AC: Done

RC #6: ... it must be noted that strong....

AC: Done

* <u>Page 16</u>:

RC #1: You are comparing different time periods! At least make a clear note of that

AC: This sentence has been removed in the revised version of the manuscript.