

# **Author's response to comments of the three reviewers on the manuscript “Total Column Water Vapour measurements from GOME-2 MetOp-A and MetOp-B” by M.Grossi et al.**

## **Anonymous Referee #1**

We thank the Anonymous Referee #1 for the careful reading of the manuscript and the helpful and constructive comments. We have incorporated them in the revised version of the manuscript. Below, we list the referees' comments (in black) individually followed by our answers (in blue).

**Total column water vapour measurements from GOME-2 MetOp-A and MetOp-B, Grossi et al.  
2 June 2014**

### **General :**

The manuscript describes the improved GDP v4.7 operational retrieval algorithm for total water vapour columns derived from GOME2. It also compares -for the 8 month overlap period- the H<sub>2</sub>O total columns measured by GOME2 on Metop-A and Metop-B. In addition, GOME-2A H<sub>2</sub>O total columns are validated/compared with ECMWF ERA interim data, SSMIS, and a new combined SSM/I-MERIS H<sub>2</sub>O total column product.

In general the topic is well suited for AMT. However, the manuscript needs some serious improvements to improve the readability. This relates both to the English used as to better/more accurate descriptions/explanations and argumentations of statements made. I have listed below what was not clear to me and/or needs improvement.

Also it is not clear to me why the GOME2 satellite products are not also validated with ground based measurements as a starting point. Are the satellite data used (SSMIS, SSM/I, MERIS) more accurate ? But to what extent have these been validated ? It is clear that on a global scale comparisons/validation is only possible with other satellite measurements. But one would expect to first start with validation with more accurate ground based data (e.g. radiosondes, FTIR,...) and then extent to comparisons on the global scale. Please explain.

We agree with the referee that ground-based measurements are the preferred source for validating satellite data because of their documented quality and stability. Therefore we have performed a validation of the GOME-2 satellite products against ground-based measurements in a separate paper “Comparison of GOME-2/MetOp total column water vapour with ground-based and in-situ measurements” by Kalakoski et al. (2014). This paper provides a detailed description of the global validation of the O3M-SAF total column water vapour from the GOME-2 sensors (for the period January 2007- July 2013) using radiosonde data from the Integrated Global Radiosonde Archive and GPS data from COSMIC/SuomiNet network. We added in the Introduction a reference to this paper, with a short summary of the results. Also the publication by Anton et al. (2014) is dedicated to the validation of the GOME-2 data set here analyzed against six reference atmospheric soundings data sets obtained from GCOS Reference Upper-Air Network (GRUAN).

In this work, we decided to focus only on global scale comparisons with other satellite measurements and ECMWF data. The algorithm for the retrieval of H<sub>2</sub>O VCD, even though proved itself to be

sufficiently precise also on single measurements, was designed to have a good accuracy on a global scale.

The combined SSM/I + MERIS data set was already compared with a large variety of reference observations in the framework of the ESA DUE GlobVapour project. Considered data included GUAN radiosonde as well as ARM microwave radiometer and radiosonde observations, AIRS and ATOVS water vapour product. The technical specification of the bias ( $1 \text{ kg/m}^2$ ) was met in the majority of the comparisons (see Schneider and Schröder, 2012).

The SSMIS data produced by the Remote Sensing Systems are subject to daily quality verification and they have been extensively studied in Calibration/Validation campaigns (Kunee et al., 2008). Match up comparisons between co-located SSMIS Environment Data Records (EDRs) and observations by operational radiosonde based on the Integrated Global Radiosonde Archive (IGRA) and ECMWF analysis fields have also been performed. Moreover, very large buoy and radiosonde data sets were used to calibrate the retrieval algorithm.

### **Detailed comments :**

Abstract l5,l16,l23 not clear to me why one comparison is called a comparison and the other is called validation. Is there a clear difference ?

In our updated version of our manuscript we use the terminology “comparison” (or “inter-comparison” when comparing GOME-2A and GOME-2B data sets) to refer to the analysis we perform and not “validation” anymore.

The term “validation” is defined as “the process of assessing, by independent means, the degree of correspondence between the value of the radiometric quantity derived from the output signal of a calibrated radiometric device and the actual value of this quantity” (see J.C. Lambert, 2013). According to this definition, the ECMWF data can not be considered a reference data set suitable for “validation” because include modeling. At the same time, the SSM/I and SSMIS microwave measurements, which can detect water vapour below clouds, cannot be used as a precise reference data for water vapour data from GOME-2 satellites spectral range, where clouds block the radiation. Therefore, we now use the term “comparison” in all cases we examine.

Section 1 (introduction) It is not clear to me if earlier versions of the GOME2 H<sub>2</sub>O total column product have been validated before and what the outcome of those validations was ?

The validation of the GOME-2 H<sub>2</sub>O total column data produced with an earlier algorithm version (GDP 4.5) was already presented in Kalakoski et al. (2011). From the comparison with SSM/I data it was found that GOME-2 generally overestimates the SSM/I data and typical biases between 1.7 and 2.  $\text{kg/m}^2$  were found for monthly global averages. These values are about one order of magnitude larger with respect to the bias we obtained with the SSMIS or SSM/I+MERIS sample in our analysis (see Tab.2).

Following the referee suggestion, we have clarified and commented this point in the Introduction.

p.3027,l4 ENVISAT was not launched in 1995

Corrected : “SCIAMACHY sensor launched in 2002 on ENVISAT”.

p.3031, l3-20. It is not 100% clear to me on which parameters the LUT for correction factor depends. I assume also cloud fraction (l13)

but it is not mentioned where that information is taken from.

The LUT with the correction  $C_{\text{ratio}}$  in the AMF computation depends on the solar zenith angle, on the viewing geometry (line of sight angle and relative azimuth angle) and on the surface albedo (see Valks et al., 2012). These correction factors are derived from radiative transfer calculations based on an average  $\text{H}_2\text{O}$  profile, an  $\text{O}_2$  profile from the US standard atmosphere and assuming a fixed surface albedo of 2% and cloud free conditions (similar results were obtained also assuming 3% surface albedo over ocean and 5% cloud fraction). We have now clarified in the text that we assume cloud free conditions.

p. 3035,14 please provide precise information on which ground pixel(s) is(are) used as a reference, and explain why this observation is used as reference to normalize. How can we be sure the  $\text{H}_2\text{O}$  column is more accurate for that viewing angle ?

Is it not possible to use the comparison with ECMWF, SSMIS, SSM/I or MERIS to determine which of the ground pixels is best use to base the SAD correction/normalisation on ?

For the scan angle dependence correction we used pixel numbers 9, 10 and 11 near the center of the swath. We indeed determine which ground pixel selection is best to use based on comparisons with ground-based and satellite observations.

According to the dependence of the  $\text{H}_2\text{O}$  VCD measurements on the pixel number (Figure 1) the WV obtained averaging the pixel number 9-10 and 11 is very close to the  $\text{H}_2\text{O}$  VCD averaged over pixel numbers 0-23. However, the normalization introduced in this way is easy to calculate and rescale.

Also, the choice of pixel numbers near to the center of the swath should produce more accurate measurements (smaller AMF).

p.3035, 11-2 I don't think I understand '(we require ...). In this way we avoid ....  $\text{H}_2\text{O}$  columns.'

We compute the correction taking into account only regions where we have a sufficient number of measurements in order to reduce the effect of natural variability. We have reformulated this sentence in the revised manuscript.

More in detail, the criteria we used to select those regions is the following. We require that, for a given latitudinal bin, the ratio between the average value of the water vapour measurements with a given pixel number  $x$ , and the averaged value of the water vapour measurements with pixel number used for the normalization (9-10-11), does not vary more than 20%.

p.3035 It would be interesting to see the spread on the datapoints in figure 1 to give an impression of how well this SAD 'correction' can be determined. Please also provide some information on what you think is the accuracy of the correction and why.

We have added the spread of the  $\text{H}_2\text{O}$  VCD data points in Figure 1. The spread of the water vapour data points (defined as the standard error of the mean :  $SE = S/\sqrt{N}$ , where  $S$  is standard deviation of the sample mean,  $N$  is the sample size) is very small due to the large number of data which we used to compute the correction (we sum up 6 years of GOME-2A measurements).

Because of the large natural variability in the spatial distribution of the water vapour data, the standard deviation is quite large, but the statistical bias as function of scan angle is well determined, due to the large number of measurements. We use a 10<sup>th</sup> order polynomial to fit the normalized measurements and compute the LUT for the SAD correction. With our procedure, residuals are of the order of few percent.

p.3035 do I understand correctly that the SAD correction is also depending on the month in the year ? Please state clearly on which parameters (latitude, month, pixel across swath, land or sea, ...) the SAD correction depends on.

Yes, the SAD corrections tables we have determined depend on the surface type (land or ocean), on the month in the year, on the latitude and on the scattering angle (pixel number). We have now clarified that in the text.

p.3034 do I understand correctly that 6 years of GOME2A data is used to derive the correction that is then also applied to GOME2B ? Please add in caption figure 1 that this is GOME2A data.

Yes, we used 6 years of GOME-2A data to derive the correction that is then also applied to GOME-2B. We have updated the caption in Figure 1.

p. 3036, l8 what is the 'operational offline water vapour product' ?

With 'operational offline water vapour product' we refer simply to the operational GOME-2 water vapour product, which is retrieved offline, i.e. within few days after the GOME-2 measurements (in contrast to near-real time data). We agree with the referee that this terminology could be ambiguous. Therefore, we avoid this expression in the updated manuscript and replace it with 'operational water vapour product'.

p. 3037, l19 does it make sense to add a stdv to the mean bias ?

Done. We have now added the stdv to the value of the mean bias (bias=-0.05g/cm<sup>2</sup>, stdv=0.5g/cm<sup>2</sup>). The standard deviation is relatively high because of the natural variability in water vapour.

p. 3038, l7 what is asymmetric cloud screening ?

With 'asymmetric cloud screening' we intend the cloud screening of the H<sub>2</sub>O VCD which, as you can see from Figure 1, is asymmetric between the east and west part of the maps. As already mentioned in Section 3.2, the O<sub>2</sub> cloud screening criterion rejects mostly GOME-2 measurements over the west part of the scan, since these are measurements with small AMF and low GOME-2 sensitivity for H<sub>2</sub>O. We now wrote "the asymmetric cloud screening (due to the O<sub>2</sub> cloud flag indicator, see Section 3.2)"

p.3038 it should also be clearly stated somewhere after the SAD correction is explained that in principle always SAD corrected data is used unless otherwise stated (or is this not the case ?)

Yes, in the course of our analysis we always employ SAD corrected data, except in section 4.1 where we analyze the effect of the correction. We have now clearly stated that in the end of Section 3.2.

p. 3038, l9 is this for individual collocated observations ? or how is this done ? Fig. 4 it would be useful to also include the spread (1 sigma) on the monthly averages

The global averaged monthly bias is obtained using the gridded monthly mean data and not from co-

located GOME-2A and GOME-2B daily data. However, in this case the two procedures produce very similar results (slightly larger scatter in the case of co-located measurements because of smaller data set), since the GOME-2A and GOME-2B data sets are processed with the same algorithm and cloud screening procedure.

We have included the spread on the monthly average bias shown in Figure 4. The spread in the case of water vapour data is dominated by natural variability and so it is quite large.

p. 3038, l21 is there a way to get access to Hovila et al. ? I tried to look up what was stated here, but did not manage to find the document.

What is the stdv on the mean difference and to what extent does that influence the statement on ‘can be used for scientific purposes’, or is that not relevant ?

We thank the referee for pointing out that the reference to Hovila et al. (2008) is not available for download. Therefore, we now refer to the updated O3M SAF Service Specification Document by Hovila et al. (2013). Both documents contain a table with the values of the Threshold, Target and Optimal Accuracy value for the Offline Total H<sub>2</sub>O product (Accuracy: Threshold 25%; Target 10%; Optimal 5%) .

The stdv on the mean difference is dominated by natural variability of the water vapour content and does not influence our statement on the reliability of the data set.

p. 3038,l25 again, how is this done ? is first the monthly mean determined per grid cell for GOME2A and GOME2B separately and then the differences taken per grid cell ? Or are first only individual collocated datapoints per grid cell taken and then the monthly mean determined, and then the differences per grid cell plotted ?

Consistently with what we have done for the bias computation (see comment above), we use monthly gridded mean data from the GOME-2A and GOME-2B satellite to perform the orthogonal analysis and not co-located data points.

Fig. 6 From the plots it looks like the red points indicate lower H<sub>2</sub>O VCD than the green points. According to the figure caption the red points are the GOME2B values, so then the difference plot would be mostly negative. So I get the impression the red and green is interchanged.

We thank the referee for pointing out this mistake. Indeed, red points are the GOME-2A values and green the GOME-2B, as now explained in the caption of Figure 6.

Fig. 6 also here stdv/spread on points would be appreciated. You can make light grey or something so we don’t miss the main message.

The spread on the value of the bias GOME-2B – GOME-2A is consistent with the one computed in Figure 4 and is dominated by natural variability.

p. 3039, l14-15,  
largest deviations -> largest absolute deviations  
low latitudes ( $\pm 10^\circ$ ) -> near the equator (10S-10N) (?)

Corrected as recommended.

p.3040, 13 ECMWF is not only modeled data, please rephrase. Also, I think ECMWF uses radiosonde data which is possibly also used by the SSM/I dataset for calibration ?

We have now rephrased the sentence and specified that for our comparison we used the ECMWF forecast data. Also, we agree with the referee that the compared data sets are not completely independent of each other. Observations assimilated in ERA-Interim consist of a large variety of in situ and satellite data, including clear-sky radiances from SSM/I. We added a comment in the revised manuscript.

p. 3040, 14 what does that mean ‘at all available time steps’ ? please provide numbers and what that means for the comparisons.

We use ERA Interim forecast fields, which are produced from forecasts beginning at 0000 and 1200 UTC.

p.3040, 123-27 I think this was already defined earlier in the manuscript, then it can be removed here.

We removed the definition of the H<sub>2</sub>O flag.

p.3040, 128-29 please make clear if collocated data is being compared or monthly averaged data or ..

In the new version of the manuscript we made our comparison more consistent by using the same gridding and filtering methods for the three analyses described in Section 5. As now clarified in the text, we have re-gridded single day measurements in a 1.5° grid and then determined the bias from the daily co-location between GOME-2 measurements and the reference data sets. We also computed the statistics separately for land and ocean surface (see also answer to Major Comment 3 from Anonymous Referee #3).

p.3041, 16 how is the RMSE determined because I don’t see such a large RMSE in the fig.7 data ?

The RMSE is computed in the following way:

$$RMSE = \sqrt{\sum_N [(H2O_{GOME} - H2O_{ref})^2] / N}$$

Expressing the formula in words, the difference between the GOME-2 sensor and the reference data set (H<sub>2</sub>O<sub>GOME</sub>-H<sub>2</sub>O<sub>ref</sub>) are each squared and then averaged over the sample (N). Finally, the square root of the average is taken. The RMSE for the full time series (see Table 2) is just the mean of the RMSE computed from the monthly comparisons values. We add this definition in the updated manuscript.

Since the deviations are squared before they are averaged, the RMSE gives a relatively high weight to large deviations. This means the RMSE for the water vapour measurements is relatively high due to the high water vapour natural variations. Most useful is the bias analysis and the seasonal variation of the bias from which we computed the uncertainty margins result.

Fig. 7 why is the SSMI+MERIS vs GOME2A comparison limited to 2007-2008 ?

Unfortunately, SSMI+MERIS products from the GlobVapour project are available only for the time frame 2003-2008 and so our comparison must be limited to the period 2007-2008.

p.3041, 124 what does ‘cloud corrected’ mean ? Or is cloud-free meant here ?

In this sentence 'cloud- corrected' means 'cloud-free', i.e. we remove (part of) the cloud contamination by flagging cloudy pixels.

p.3042, 125-28 I do not understand on what evidence the explanation of (changing) cloud conditions is based to explain the observed seasonal differences between GOME2(A) and SSMIS. Please explain more clearly. I mean is the much better agreement between GOME2(A) and ECMWF not suggesting the problem is with SSMIS ?

Since the microwave instrument can measure the water vapour also below clouds, we expect some residual differences between GOME-2 data (based on visible observations, where clouds block the radiation) and SSMIS data, which deliver results also in cloudy condition. Therefore, we claim that the seasonality of the cloud properties can have an effect on the bias, and amplify differences in different period of the year. However, we observe a pronounced seasonal cycle of the bias distribution also when comparing GOME-2 data with ECMWF data only over ocean.

The better agreement between GOME-2A and ECMWF data could suggest that the problem is within the SSMIS data, but differences in the retrieval procedure have to be taken into account to explain the bias. Also, in general the value of the mean bias decreases when comparing monthly means data (with respect to the value obtained when doing daily co-location) because of the different resulting cloud selection (more cloudy data enter in the comparison, since they are not rejected on a daily basis).

p.3042, 110-12 you mention the diurnal cycle in H<sub>2</sub>O. To what extent does that affect the comparison between MERIS, SSMIS, SSM/I with GOME2 ? I mean for ECMWF you take –I presume- the output closest in time with GOME2A. For the others you only have fixed observation times. Please elaborate to what extent this can explain the observed differences (or not).

In the revised manuscript, we used daily co-location between GOME-2 and ECMWF data in order to compute the bias. The mean daily data were computed by averaging the ERA Interim forecast fields produced from forecasts beginning at 0000 and 1200 UTC.

When repeating the comparison for outputs closest in time with GOME-2A we found differences in the mean bias up to 0.02 g/cm<sup>2</sup>. However, the global distribution of the bias between GOME-2 and the other data sets is similar in both cases.

p.3042, 118-20 do you mean that the microwave measurements are affected by clouds and these can not be flagged ? therefore their TCWV is biased low compared to e.g. GOME2 ?

In the microwave spectral range, measurements are not affected by clouds, which means that the water vapour vertical column can be measured also below clouds, while in the visible regions clouds block the radiation. Therefore, SSMIS measurements in cloudy conditions are biased low compared to GOME-2 measurements (we compute the bias as GOME-2 – data; microwave data in cloudy conditions have in general higher water vapour VCD, since they see also below clouds). As already commented before, we found that the value of the mean bias decreases when comparing monthly mean with respect to comparing daily mean, because in the former case more cloudy data enters in the comparison, since they are not rejected on a daily basis.

p.3043, 110 can something be said and referenced as to the quality of ECMWF data ? I mean we need to have some idea to judge its value for the comparison provided here.

We added a comment about the accuracy of ECMWF data.

p. 3043, l4 would another advantage not possibly be that there is output at several moments during the day ? which moments in time have been used here in comparison with GOME2, and what does that mean of average and stdv in differences in temporal sampling between the two ? (and what can that mean for observed differences ?) in l12 it is stated that 12 h forecast values have been used to derive a daily mean, why is not the ECMWF data closest in time with GOME2A used ?

In the revised manuscript we used the same gridding and filtering procedure for the three data-set (see answer to the Major Comment 3 from Anonymous Referee #3). Therefore, even though an advantage of ECMWF data is that they are available at several moments during the day, we used only daily mean data. This is necessary because one of our data set (SSM/I + MERIS) is available only as daily and monthly means. In the same way, we selected a 1.5° grid size because this is the resolution of the ERA Interim data set (even though we have better resolution available for the other data sets).

p. 3043, l21 inner -> inter

Corrected.

p.3043, l25 tropical ? total column H2O

In the tropics, removed.

p.3043,l26-27 should it not say monthly mean differences between GOME2A and ECMWF ?

Yes, corrected.

p.3044, l2 good accuracy is a bit too strong I would say. I would say it confirms that GOME2 captures the overall spatial variability in the H2O total column values quite well.

Changed.

p.3044, l15-19 if undetected clouds are the problem should this not lead to a systematic underestimation of the H2O by GOME2 ? is that what we observe and where you are referring to here ? that is not clear to me.

Yes, as stated by the referee, if clouds are the problem this should lead to a systematic underestimation of H<sub>2</sub>O by GOME-2. We refer here to “The differences over ocean, e.g. along the ITCZ and the Pacific Warm Pool region”, where we observe a small positive bias (light red regions) since we compute the bias as GOME-2A – ECMWF (see Figure 11, top panel).

p.3044, l23 right panel -> lower panel

Done.

p.3044, l26-... in fact the region with largest deviation (e.g. over central Africa) does not seem to correspond with the actual desert regions or highest surface albedos (I did a quick check with the



surface albedos I found on the TEMIS website). So I don't understand this hypothesis.

We have reformulated our comment in the revised manuscript. What we claim is that in most places the bias is associated with high surface albedo values or a high TCWV.

Dry bias is observed in most extremely arid areas, like southern parts of the Sahara desert, the coast of Somalia, the Arabian Desert in the southern part of the Arabian Peninsula, the Thar desert in the northwestern part of the Indian Subcontinent. Regions with relatively high surface albedo values (in the range 0.3 – 0.5) which present dry bias include northern Africa, the Arabian Peninsula, India and part of East Asia and Central America.

We agree with the referee that the dry bias we observe in the central Africa should be rather related to land areas with a very high humidity in the Northern Hemisphere summer months (see ECWF monthly mean H<sub>2</sub>O map for August 2008). From a correlation analysis we found indeed that the bias between GOME-2A and ECMWF data over land areas decreases with increasing humidity.

See attached Figure 1-3.

p.3045, 18-11 can we then explain why this results in always higher GOME2 values compared to ECMWF ? or is this not the case ?

Indeed, we always observe a higher total column water vapour for the GOME-2 values compared to ECMWF values in the northern latitudinal regions. This result is consistent with what observed also in the comparison against the SSMIS and SSM/I data sets. While the bias between the SSM/I and SSMIS data set and the ship data used for the calibration of the retrieval algorithm might be responsible for the differences, we could not identify the source of the observed bias in these regions.

p.3045, 122 what is a physically based algorithm ?

The algorithm used for the retrieval of SSMIS water vapour data is based on a physical model for the brightness temperature of the ocean and intervening atmosphere. Radiative transfer theory provides the relationship between the Earth's brightness temperature and the geophysical parameters (surface temperature, near-surface wind speed and vertically integrated cloud liquid water) which are used for the retrieval (see also Wentz, 1997).

p.3046, 15 How is collocated defined here ? Later only the temporal collocation is specified.

In this analysis we were selecting SSMIS data with spatial resolution within a GOME-2A pixel (0.5°). In the revised manuscript, we used a gridded map of 1.5°, since this is the resolution of the ECMWF data set and we used a consistent methodology for all three data sets we analyzed.

p.3046, 19 what is the max. difference in time ? is that as much as 7 hours ? What is the effect of this large temporal difference (even 4.5 hrs is quite large) ? or is this issue of minor importance ?

Analyzing the results obtained in this comparison with the one obtained in the revised manuscript using daily co-location, we found that the total mean bias is smaller in the co-location case (0.006 g/cm<sup>2</sup> instead of 0.028 g/cm<sup>2</sup>). However, the distribution of the differences is very similar.

p. 3046, 119-25 I would first remark that these differences were also observed in the comparison of GOME2 with ECMWF and are thus likely related to GOME2, is that correct ? Or is the calibration of SSMIS with radiosondes and the use of radiosondes by ECMWF causing a dependence between the

two datasets making it impossible to draw this conclusion ?

We added this remark. While the bias between the SSM/I and SSMIS data set and the ship data used for the calibration of the retrieval algorithm might be responsible for the differences (and ECMWF uses SSM/I data for assimilation), we cannot exclude that the reason for the discrepancies resides in the GOME-2 data set.

p.3046, l24-25 I do not understand that a cloud residual would cause GOME2 to give higher values than SSMIS (or ECMWF), should we not expect that this gives lower values. Please clarify why this would give a positive bias for GOME2.

In the revised manuscript we only state that the cloud fraction retrieved by the GOME-2 instrument is typically higher than 0.5 in these regions, but we do not claim that one effect causes the other. We agree with the referee that in general clouds shield (and therefore decrease) the fraction of H<sub>2</sub>O VCD below them. On the other hand, since clouds are bright, they enhance the sensitivity towards trace gases directly above the clouds with respect to the clear sky scene with a low albedo.

p. 3047, l8 how should I interpret the bias of 0.25 g kg<sup>-1</sup> in terms of g/cm<sup>2</sup> as used here ?

The value  $Q_a=0.25$  g/kg is the surface layer specific humidity, which is related to the SSMI temperature. This quantity cannot be directly translated in terms of vertical columns without specific assumptions on temperature, pressure and water vapour profile at the location/time of the retrieval, which would affect our estimate of the bias. Using the typical concentration of specific air humidity values (Bentamy et al., 2003), we estimate that this value corresponds to an average bias of about 2%. The required accuracy for water-vapor retrievals in SSMIS over ocean is 3 kg/m<sup>2</sup> over tropics, 2 kg/m<sup>2</sup> over mid-latitude regions, and 1 kg/m<sup>2</sup> over polar regions.

p.3047, l10 where is the estimate of 1% for 1 hr time difference coming from ? please provide reference or explanation.

We provided the reference to the O3M SAF Validation Report by Kalakoski et al. (2011), where the diurnal variation of the H<sub>2</sub>O VCD from ship borne GPS observations at the tropical Indian ocean is reported.

p.3047, l20 does this now mean different SSM/I data is being used compared to section 5.2 ? Please explain in text to avoid confusion.

Yes, the water vapour data used in this comparison are SSM/I data from the F13 and F14 satellite and they are obtained using a different retrieval algorithm. We now clarify this in the text.

p. 3047, l22 I understand from this that anyhow a different retrieval algorithm was used for SSM/I compared to the SSM/I data used in section 5.2 ?

Yes, the SSM/I total column water vapour product for the GlobVapour project were retrieved using an improved 1D-Var scheme developed at ECMWF by Phalippou (1996).

p.3047, l26 why was the L3 GOME2 dataproduct used here ? and not first daily collocated data and then averaged as in all previous comparisons ?

In the revised version of the manuscript we indeed use daily co-located data for this comparison. Combined SSM/I + MERIS daily and monthly Level 3 products for 2003-2008 are available within the GlobVapour project.

p.3048, l5-7 I don't understand is the SSM/I data cloud contaminated?

Since the microwave instruments can see below clouds, the SSM/I and SSMIS data sets can retrieve the water vapour also in cloudy conditions. Comparing the GOME-2 measurements with the SSM/I on a monthly bases means to use also SSM/I observations with large cloud cover. If we do a daily co-location, on the other hand, the results of the two satellites will be closer because in this case we reject all SSM/I measurements in regions flagged as cloudy by the GOME-2 instruments.

p.3048, l10-11 I would add that the differences show large resemblance to what is seen in the comparison wit ECMWF.

Added.

p.3049, l14-16 please add that similar patterns were observed in the differences with ECMWF hinting at a problem in the GOME2 data. Wrt remark on surface albedo see my previous comment on that in section 5.1. If you think there is a correlation with high albedo please provide corresponding plot (in reply) to illustrate that.

Added. Attached you can find a map of the distribution of the Surface Albedo for GOME-2A land data on August 2008, and the correlation plot between bias and Surface albedo value for the comparison GOME-2A/ ECMWF (see attached Figures 1-3).

p.3049, l8 I don't understand 'and consequently the different cloud selections ...'. Caption fig. 11 mentions only cloud free data used. Do you mean here remaining cloud contamination ? But I don't really see why that is more of a problem when using daily collocated or not as SSM/I does not suffer from cloud contamination (if I understood correctly)

In this section we indeed refer to remaining cloud contamination, which is a major problem when comparing gridded monthly data since we do not remove strongly cloud contaminated pixels on a daily basis. In the revised manuscript, we compute the bias between GOME-2 and SSM/I+MERIS from daily co-located data and we obtain very different results, as you can see by comparing the old and new analysis.

p.3049, l29 is this then due to the effect of temporal mismatches ?

Yes, it is due to the large natural variability in these regions.

Fig. 12 Figure captions mentions GOME2A – GOME2B, incorrect

Corrected.

p.3050, l. 13 unclear to me what other satellite retrieval methods are meant here, or was this mentioned already elsewhere ?

As mentioned in the introduction, here we refer to other satellite retrievals, e.g. GPS measurements, microwave and near infrared measurements (SSM/I, SSMIS and MERIS).

p.3051, l10 four or three ?

In this manuscript we analyzed only three data sets. Corrected.

p.3051, l19-21 see earlier remarks on the correlation with surface albedo

See attached Figures 1-3 and explanation provided before.

In section 4 it is concluded that GOME2B is biased high compared to GOME2A. In section 5 GOME2A is compared with other water vapour datasets. Would it not be useful to say something about the fact that GOME2B is also biased high compared to the other datasets and to first order could be corrected based on the comparison with the GOME2A data ?

In the revised manuscript we included this comment in the conclusions.

p.3061, Table1 The mentioned time period does not apply to all comparisons, please correct.

We now specified the time period used for the different comparisons in the caption of Table 2.

**Major point :** The English in the manuscript should be improved. I advice to ask a native speaker to read the manuscript. The following is just my first suggestions w.r.t. improving the English :

p.3022,l24 are => is

p.3023,l2 are also present => are observed

Done.

l23 However, .... Sentence needs rephrasing

Rephrased.

l25 explicitly state what 'it' is

Done. 'It' is water vapour.

p. 3024, l3-4 something is missing in the sentence

Rephrased.

l6 over => under

Done.

l8 what are spurious changes ?

Sources of possible random error and bias in computing statistics of radiosonde data, including: sampling problems, bias introduced in using monthly average data due to the nonlinear relationship among moisture variable, inconsistent reporting of moisture variable, daytime versus nighttime soundings.

L22 on => over

Done.

L26 timescale

Done.

L29 this sentence is now very unclear. I think you want to say two things. One is that in NIR you can not see through clouds, and the other thing is related to observations over oceans. Please rephrase.

Rephrased.

P3025, L16 was => is

L18 current => current operational

Done.

L20 take out 'which ...predecessor GOME' sentence is much too long

Done.

P3026, L9 out, but the results pointed out => out pointing to large....

L13 against => with

L17 type instruments => instruments

P3027, L3 launched in

L8 at => using

Done.

L13-14 strange formulation '.. improved version.., but we can identify ...'. Please rephrase.

L20 Finally, ... I think that is a consequence of the statement in the sentence before on the swath, please rephrase.

Rephrased.

L23 the different GOME-type sensors.

Done.

L23 why is there a hard return here ?

Different paragraph related to GOME-2.

P3027, L3 At => On

L4 operating a reduced swath width of 960 thereby increasing its spatial resolution....

L4 if you use GOME-2B in that sentence I would also use GOME-2A.

Please introduce somewhere this terminology, so GOME-2/MetOp-A (GOME-2A) ....

Corrected.

P3028, L8 improves => improve

L11 and it is foreseen to guarantee => guaranteeing

L18 variety of methods .... has been ... => Various retrieval methods ..

have been

Done.

P3029, l6 what does it mean to take into account the cross sections of O<sub>2</sub> and O<sub>4</sub>, I presume you mean taking into account absorption by ...

It means taking into account absorption by O<sub>2</sub> and O<sub>4</sub>, and therefore also their cross sections for the DOAS fitting.

L7 what is meant by a single H<sub>2</sub>O cross section is used ?

It means that we used a single GOME-convolved H<sub>2</sub>O cross section calculated for a fixed temperature and pressure of 290 K and 900 hPa.

In Wagner et al. (2003), the authors investigated the temperature and pressure dependence of the H<sub>2</sub>O absorption structure by varying the temperature by +/- 20 K and the pressure by +/- 100hPa. The analysis of the GOME-2 measurements using these different H<sub>2</sub>O spectra yielded H<sub>2</sub>O SCDs varying by only +/-3%.

L12 similar spectra, please rephrase

L22 we therefore apply

L23 correction

Corrected.

L25 and they are larger, sentence does not make sense. Rephrase.

The correction factors are larger for higher H<sub>2</sub>O SCDs.

L3031,l20 what does this sentence mean ? why is this product –contrary to others apparently- especially valuable .... ?

The GOME-2 H<sub>2</sub>O VCD is especially valuable for long-term time series and climatological studies because the water vapour retrieval here described does not rely on external information (except a global surface albedo map used for the computation of the correction factor).

L24 another example of connecting singular with plural, this and errors, I will no longer explicitly list these errors, just too many.

L26 may be also => may also be

Done.

I also suggest to try to avoid the use of 'we' in the manuscript. I stopped on page 3031 identifying faulty English. Please have a native speaker read your manuscript.

p.3032, l21 variable scenes => inhomogeneous scenes

Done.

l22 largely separated detector channel, unclear what this means

We intend that the separation between the detector channel is large. Now rephrased "measurements far away from the O<sub>2</sub> band".

l26 begin (?) of channel 4 => beginning

Done.

Table 1. The main ... are clearly visible. Maybe say Summary ... characteristics illustrating the main improvements of GOME-2 compared to its ....

We have revised this sentence as suggested.

Table 1 what does LT mean ?

LT stands for Local Time. The abbreviation is now introduced in Section 2.

## **Figures**

- Most figures (fig 1, 2, 3, 5, 8, maybe also 6 and 12) should be increased in size as they are too small to see any details. Also delete all titles of figures if information is already in the figure caption as in a lot of cases the title text is too small to read.

We increased the size of most of the figure in our new manuscript version. Moreover, we remove the redundant titles in Figures 5 and 12.

## Anonymous Referee #2

We would like to thank Anonymous Referee #2 for the thorough review and the many useful comments and suggestions. We believe by addressing these, we can improve our manuscript.

Below, we list the referees' comments (in black) individually followed by our answers (in blue).

This paper describes the operational total water vapour column retrieval algorithm for GOME-2, including various corrections applied to new algorithm version (GDP 4.7). Retrievals from GOME-2 instruments aboard Metop-A and -B satellites are compared for 8 month overlapping months. Finally, GOME-2A retrievals are validated globally (for period of January 2006 - August 2013) against ECMWF model outputs and SSMIS and SSM/I + MERIS satellite data sets.

Topic is well suited for publication in AMT. Methods and results are presented well and in logical order, although the paper needs further proof reading, as well as some clarifications to figures. Some suggested improvements on the figures are listed below. Beyond these clarifications, there are no major problems with the paper.

### Detailed comments:

FIGURES: Clarity of the figures would be greatly improved by different colour scale, especially for figures 3 and 9-11, where difference between observations is shown.

Rainbow colour scale makes it very hard to see differences and areas with zero bias are not clearly distinguished from small negative biases. I would suggest a divergent colour scale in these figures for clarity.

We agree with the referee that the use of a divergent colour scale will help the reader in the visualization of the differences and areas with zero bias between the different data sets. Therefore, we have now adopted a divergent color scale in Figure 3 and Figures 9-11 of our paper (we have substituted the jet color scale with a reversed RdBu color map).

Size of some figures (At least figs 2, 3 and 8) should be increased.

Done.

Figure 2 should include a subplot showing the difference between observations with SAD correction and without. Differences are hard to spot from two very similar maps.

Following the referee suggestion, we also include a new plot (Figure 3) showing the difference between GOME-2B observations with and without SAD correction.

In figure 5, the length of the x-axis should probably be shortened to better show any skewness in the histogram.

Done. We have shortened the length of the x-axis in the right panel of figure 5.

Figure 7 should have a clear line at zero difference. Bias, or any possible drift is hard to see in the figure.



Done. We added a black line at zero difference in Figure 7 and in Figure 4. In order to facilitate the reader in the estimate of the bias ranges from these plots we have also superimposed a grid.

LANGUAGE: Please improve the English throughout the manuscript, possibly have a native speaker to read the text. Mostly the errors are relatively small (singular/plural confusion or similar), but in some cases they also make the manuscript hard to follow, especially p.3024 L29-31: Sentence very unclear, not sure what was meant. Please clarify.

In order to clarify this aspect we have reformulated the sentence adding the following statement :  
“Sensors operating in the near infrared, like the Medium Range Resolution Imager Spectrometer (MERIS) on ENVISAT (Li et al., 2006), can derive water vapour also over land, but cannot retrieve the water vapour in cloudy conditions. Moreover, the very low albedo of the ocean-surface in the NIR limits the retrieval in these areas. However, measurements are possible in sun-glitter or above cloudy regions over ocean, since these two conditions increase the surface albedo.”

Outside the sun glitter, water surfaces show a significantly lower surface reflectivity in the NIR with the result that the measured signals over water surfaces are governed by aerosol scattering.

## **Anonymous Referee #3**

We thank the Anonymous Referee #3 for the time spent on reading and commenting the manuscript. Below we give answers to all comments in blue. In black is the text of the original referee report.

### **Review of AMTD paper by Grossi et al. entitled „Total column water vapour measurements from GOME-2 MetOp-A and MetOp-B”**

#### **General**

This paper introduces an improved version (version 4.7) of the GOME Data Processor (GDP) algorithm for deriving total column water vapor (TCWV) from observations of the GOME-2 instruments aboard EUMETSAT's MetOp-A and MetOp-B satellites. The GDP 4.7 water vapor algorithm is used by the EUMETSAT's Satellite Application Facility on Ozone and Atmospheric Chemistry Monitoring (O3M SAF) for the operational generation of GOME-2 TCWV products and improves upon the previous version by the introduction of an empirical correction of scan-angle dependency. The TCWV estimates derived from GOME-2 on Metop A and B are intercompared for the overlap period of 8 months. Additionally, TCWV estimates from both GOME-2 instruments are compared to other satellite-based TCWV estimates and a reanalysis product (ECMWF Interim Reanalysis, ERA-interim).

The topic and the results presented are in principle interesting and well suited for AMT. There are several issues, however, due to which the manuscript is proposed to undergo a major revision before it can be accepted. Major points of criticism:

- Imprecise and misleading usage of terms,
- “real” validation (involving collocated – in space and time - data pairs) using ground-based measurements such as radiosonde observations is missing
- Comparison methods applied differs between comparison data sets used which makes it difficult to compare and interpret the results:

More details are given below.

#### **Major comments**

##### **1. Imprecise and misleading usage of terms**

The imprecise and misleading usage of terms makes the text hard to follow. For example, the terms “validation”, “comparison”, “cross-comparison” are used without any clear distinction. “Comparison” is surely the most appropriate term for the results presented so far. The term “validation” should be reserved for comparisons against a reference of known accuracy, here for example ground-based observations such as radiosonde observations, based on collocated (in time and space) data pairs.

Following the referee suggestion, in the new version of the manuscript we used the terminology comparison (or inter-comparison when comparing GOME-2A and GOME-2B data sets) to refer to the analysis we perform and not “validation” anymore (see also answer to Detailed comments Abstract from Anonymous Referee #1). We indeed agree that modeled ECMWF data cannot be considered as

reference data set suitable for “validation” since they include a lot of modeling and therefore they cannot be considered as “real” data as intended in the definition of validation. While MERIS data have been fully characterized, both theoretically and through comprehensive comparisons with real measurements, microwave measurements by SSM/I and SSMIS are not suitable for “validation”, because of the different retrieval in cloudy conditions.

The term “collocated” is also often used in an imprecise way. Both for data pairs matched in time and space (SSMIS) and when just referring to a grid box by grid box comparison of averages (ERA-interim).

In the comparison between ERA-Interim and GOME-2 sensors data sets, we did not compare monthly “grid box by grid box”, but we did first a daily-colocation and from that computed the monthly means. In the new version of the manuscript all comparisons use the same gridding and filtering method, and we use in all three cases a daily-colocation between GOME-2 and reference measurements.

Within the text several terms are also used to describe the same thing. This leads to confusion as it implies that there are some distinctions where there may really be none. Therefore better only use one term consistently e.g. H<sub>2</sub>O columns, H<sub>2</sub>O VCD, TCWV, WV are used to refer to total column water vapor estimates and ECMWF data, ERA-Interim data to refer to the ERA-interim data set. For the latter, ERA-interim would be more suitable as that’s the official name of the data set.

Please be more precise and use clear definitions. Additional work concerning the language/phrasing is also needed. (see Editorial Remarks) in order improve the readability.

## **2. “real” validation (involving collocated – in space and time - data pairs) using ground-based measurements such as radiosonde observations is missing**

The authors aim at performing a “ sound assessment of the quality of the (presented) satellite products” but the necessary validation using ground-based observations such as radiosonde data as a reference is missing. I assume this kind of validation was done but it is not clear why it is not added to the manuscript. Please clarify! Such a validation would ideally include both previous and improved algorithm version. The presented comparisons against other satellite- based and reanalysis products are still very useful and valuable to check the consistency of the spatial patterns and can serve as an additional means to check the performance of the product.

We agree with the referee that ground-based measurements are the preferred source for validating satellite data because of their documented quality and stability. Therefore we have performed a validation of the GOME-2 satellite products against ground-based measurements in a separate paper “Comparison of GOME-2/MetOp total column water vapour with ground-based and in-situ measurements” by N.Kalakoski et al. (2014). This paper provides a detailed description of the global validation of the O3M-SAF total column water vapour from the GOME-2 sensors (for the period January 2007- July 2013) using radiosonde data from Integrated Global Radiosonde Archive and GPS data from COSMIC/SuomiNet network. We added in the Introduction a reference to this paper, with a short summary of the results. The validation of GOME-2 data based on the previous and the improved algorithm versions against IGRA radiosonde data set and a detailed analysis of the differences is described in Grossi et al. (2013).

## **3. Comparison methods applied differs between comparison data sets used which makes it difficult to compare and interpret the results.**

So far for the three comparison data sets three different comparison methods are applied:

- Era-interim : daily mean (Era-interim) vs. single day-time measurement, on 1.5° grid
- SSMIS: collocated data pairs (5 a.m. vs. 9:30 a.m.) , no information about regridding
- SSMI-MERIS: Level 3 daytime only, but for SSM/I no cloud screening is applied, 0.5 ° grid

The comparison should be made more consistent by using same gridding and same filtering, e.g. using not only spatially but also temporally matched observations (like for the SSMIS comparison) which would allow to apply a common (GOME-2 like) cloud screening to all comparison datasets.

In any case, the statistics should be calculated separately for ocean, (maybe also coasts) and land parts (for Fig. 7 and Table 2). So far the bias time series (Fig. 7) and statistics (table 2) are not comparable as the data sets have different spatial coverages (ocean-only, land+ocean).

In order to make the comparison more consistent, in the revised manuscript we use the same gridding and filtering procedure for all three comparisons. All data set are now gridded on a 1.5° grid, since this is the smallest grid available for the ERA-Interim data set (we use 1.5° and not anymore 0.5° for the SSMIS and SSM/I+MERIS datasets). We used daily co-located data to compute the mean monthly bias in all three cases. In fact, for the SSM/I+MERIS GlobVapour data set we have only monthly and daily mean data available. This procedure reduces the bias between microwave observations (SSM/I), which can retrieve water vapour below clouds and GOME-2 measurements in the optical region, where clouds block the radiation. In this way pixels flagged as cloudy are removed on a daily basis also in the reference data sets.

As suggested by the referee we also compute the bias statistics separately for ocean and land surfaces (see Table 2 and Figure 9).

The bias time series including all three comparisons (Figure 8) is kept as a reference for the reader. The aim of this figure was not to compare the bias results within each other, but to have all results included in the same plot for reference.

### **Minor comments, questions, critical remarks:**

P 3022,L6:“long-term stability”: the term seems inappropriate as the data period studied covers 6.5 years only which not really be considered to be long long-term. Reformulate.

Corrected. In the revised manuscript, we rephrased it as “temporal stability”.

P 3022,L 28: “independent”: As 2 of the 3 comparison datasets are also satellite-based products using similar types of instruments (SSM/I and SSMIS) and SSM/I and SSMIS radiances are also assimilated in ERA-interim Therefore, the comparison data sets are not completely independent.

Corrected. In the revised manuscript we removed the word “independent” since, as the referee pointed out, it could be misleading. Here we intended simply that the three data sets are independent with respect to GOME-2A and GOME-2B measurements.

P3024, L13: “only satellite observations”: what about reanalysis data sets, weather models..?

Agreed. In the revised manuscript, we insert this remark in the following paragraph and remove “only”.

P3025, L 26: “input quantity”: input to what?

Corrected with “input quantity for climate models”.

P3025, L27f: “useful for numerical weather predictions”: how? For assimilation? Give more details.

We have now added: “useful for assimilation in numerical weather prediction models, e. g. for following dynamical structures in water vapor when a high absolute accuracy is not required.”

P3025, L 28: “in contrast to other satellite data sets”: What about HIRS, IASI, AIRS, ATOVS based data sets? These are also available over land and so far completely missing in your list of existing satellite-based water vapor products.

Following the suggestion of the referee, in the introduction of the updated manuscript we have introduced a short description of these satellite based data sets. Satellite infrared observations have the disadvantage of being less sensitive to the surface emission from the lowest layers, where most of the water vapour is present and they require model input.

P3026, L 9: “pointed at large deficiencies”: Also for GOME-2 retrievals? Give more details about the results! Which retrieval was used for GOME-2, a previous GDP version?

As stated in Schröder, M. and Bojkov, B. (2012), the variability of the bias of the GOME/SCIAMACHY/GOME-2 product was found to be generally large (of the order of  $2 \text{ kg/m}^2$ ). They found that the GOME product tends to be drier than the compared ground based and satellite data (including GUAN stations, 3 ARM radiosonde sites and ATOVS data) with the exception of AIRS (Aqua) data (bias  $+1.6 \text{ kg/m}^2$ ). “Large differences” on regional basis were observed in the comparison with SSM/I+MERIS with negative bias ( $1 \text{ kg/m}^2$ ) over ocean and smaller positive bias in land regions ( $+0.28 \text{ kg/m}^2$ ). For the retrieval of the GOME-2 product a previous GDP version (4.4) was used.

P 3028, L 7: “monthly narrow-swath day”: does this refer to the change to narrow swath of GOME-2A ? Sounds like it happens only once per month? Reformulate.

We removed this comment to avoid confusion to the reader. What we mean here is that in the TANDEM mode the GOME-2B sensor continues to operate in 1920 km swath mode, but does not operate anymore once per month with a “narrow swath” day. This is not the narrow-swath of GOME-2A, but the narrow swath mode operated by GOME-2B before the TANDEM operation mode was introduced.

P3036L17: “monthly mean distribution”: you mean you look at the spatial distribution of the bias, right? Reformulate, in order to be more clear.

Corrected. The sentence has been reformulated as: “For the monthly comparison, we first analyze the spatial distribution of the bias from gridded monthly mean GOME-2A and GOME-2B water vapour columns. Then, in order to make the data selection in the two instruments as similar as possible, we have also performed the comparison using only co-located measurements.”

P3038,L7: “low statistics”: why don’t you use more data for the GOME-2 intercomparison? You have in total 8 months of overlap.

In the revised manuscript we extended the comparison between GOME-2A and GOME-2B

measurements to include the period January 2013 – June 2014. The daily comparison is nevertheless affected by low statistics, especially in the tropical regions where the orbit of the GOME-2 sensors intersect each other only in smaller regions, e.g. we have less daily co-located regions. In Section 4.1 we analyze the results obtained from daily comparisons in order to show how the different orbits of GOME-2A and GOME-2B influence the bias evaluation. In Section 4.2 we highlight the differences between GOME-2A and GOME-2B measurements based on monthly data.

P3038,L19-22: Here you have to give more details otherwise it is not clear what you want to say. What is the PRD document? How is the optimal accuracy threshold defined and which value does it have?

The PRD document is the “O3M SAF Product Requirements Document” by Hovila et al. (2008). Since we realize that this internal document is not available for download, we now add a reference to the updated O3M SAF Service Specification Document by Hovila et al. (2013). Both documents contain a table with the values of the Threshold, Target and Optimal Accuracy value for the Offline Total H<sub>2</sub>O product (Accuracy: Threshold 25%; Target 10%; Optimal 5%).

The accuracy is defined as the root mean square (RMS) difference between the measurement and the reference 'true value'. Verification methods include the comparison with ground based measurements and satellite to satellite comparison. The threshold refers to one standard deviation.

We agree with the referee that we should clarify this point for the reader and we have rephrased this sentence.

P3039,L17: what do you mean with fractional difference?

By "fractional difference" we mean relative difference - which in this case is the difference between GOME-2B and GOME-2A divided by the GOME-2B values. Corrected.

P3039,L22 “smaller number of data points due to the cloud selections”: why don't you use more months then (see also comment above)? You have in total 8 months of overlap.

As suggested by the referee we now used more than one and a half year of measurements, from January 2013 to June 2014. However, this analysis refers to one month of results, January 2013, and therefore the conclusions we reached are the same. In section 4.2 we focus on monthly differences between GOME-2A and GOME-2B data because the humidity has a strong annual cycle.

P3039,L24: Change title to “Comparison result” (see Major comments)

Done.

P3039L25f: “A sound assessment of the quality” and “validation” would imply for me the comparison against ground-based measurements like radiosonde data (see Major comments). Reformulate!

Done. In the revised manuscript, the sentence reads: “To have a sound assessment of the quality of the satellite products with respect to independent satellite observations and model data, we performed an extensive comparison of both the GOME-2A and the GOME-2B products.”

P3040,L5: “at all available time steps” what do you mean?

We used ERA Interim forecast fields which are produced from forecasts beginning at 0000 and 1200 UTC. We have clarified this point in the revised manuscript.

P3040, L23: “all validations” => all comparisons

Corrected.

P3010,L28ff: The results in Fig. 7 are presented before the data sets and the methodology are properly introduced. I suggest to create a new “comparison data sets” subsection (5.1) within section 5 where all comparison data sets are described.

Following the suggestion of the referee, we have now created a new subsection “Comparison data sets” where we introduce the data sets and methodology for comparison.

P3040,L28ff : The bias timeseries are not comparable as the data sets have different spatial coverages (ocean-only, land+ocean). See major comments above.

As already disclosed in the answer to the Major comment #3, we have now introduced a separate Figure with the bias time-series over land and over ocean separately.

P3043, L12: “forecast 12h values”? does this refer to the 12hour forecast based on 00 and 12 UTC analysis? Why do you use the forecasts and not the analysis data (6-hourly resolution)?

Yes, we used the 12 hour forecast based on 00 and 12 UTC analysis for our comparison. We uses forecast data in order to have a more independent data set, since they include mostly modeling. However, we have redone the same analysis with the analysis data set (step 0 at 0,6,12, and 18 UTC) and we verified that the result is very similar (slightly larger bias, 0.039 g/cm<sup>2</sup> instead of 0.34 g/cm<sup>2</sup>).

P3044,L2: “good accuracy”: I would call it good agreement.

Here we refer to the accuracy of the GOME-2A data itself, which is consistent over the land and ocean regions, e.g. the mean bias has very similar values for land and ocean surfaces. We rephrased this sentence as suggested by Anonymous Referee #1.

P3044, L13: “very low number of collocations”? you compare monthly means, so probably you mean that the GOME-2 monthly mean is based on 8 measurements only? Anyways, striving for a more consistent approach to compare the different datasets (spatial and temporal matched, as done for the SSMIS comparison) would increase the comparability.

In our previous comparison, we did not use monthly mean, but daily co-locations of GOME-2 and ECMWF data. The monthly differences were than obtained from the daily bias in the co-located regions (we were rejecting cloudy regions in both data sets). In the revised manuscript we were adopting the same gridding and filtering procedure for all the data set, which means that we used a grid of 1.5° (this is the smallest grid for ECMWF data) and daily co-location (this is the smallest temporal resolution of SSM/I+MERIS measurements).

P3044,L13ff: “also problems of the ECMWF..”: give more details on input data, preferably also already in the “comparison data set” section”

As suggested by the referee, we added more details on the ECMWF data sets in the new Section 5.1.

P3044,L18: “regions without..”, you mean grid boxes or pixel without severe cloud cover, as the cloud screening is applied on the Level 2 estimates?

The cloud screening is applied on the Level 2 estimates. The single measurements flagged as cloudy are not included in the daily measurements and therefore in the comparison and the calculation of the bias. However, residual regions with some cloud contribution (for example regions with  $CF > 0$  but  $CTA * CF < 0.6$ ) are present.

P3044,L21-P3045,L3: as far as I can see from the maps the region you refer to is actually not the Sahara desert but rather covers the Sahel zone and areas even more south, where the month of August forms part of the rainy season. That’s why the high surface albedo discussion does not apply. Reformulate!

We have reformulated our comment in the revised manuscript. What we claim is that in most places the bias is associated with high surface albedo values (in the range 0.3–0.5) or a high TCWV (i.e. large AMF). Dry bias is observed in most extremely arid areas, like southern parts of the Sahara desert, the coast of Somalia, the Arabian Desert in the southern part of the Arabian Peninsula, the Thar desert in the northwestern part of the Indian Subcontinent. We agree with the referee that the dry bias we observe in the Sahel zone and the more southern part of the African continent (but also India, part of East Asia) should be rather related to land areas with a very high humidity in the Northern Hemisphere summer months (see ECMWF monthly mean H<sub>2</sub>O map for August 2008 and attached Figures 1-3).

P3045,L13: move SSMIS related text general “comparison data sets” section as stated earlier.

Done.

P3046,L3ff: “if we evaluate..”: does not make sense to me. Maybe you mean “mean SSMIS” data instead of “mean ECMWF” data? Reformulate!

Corrected. We indeed mean “ mean SSMIS” data.

P3046,L5ff: “as for ECMWF, we select only daily co-locations”: the term co-location in the context of the ECMWF comparison is confusing. You compare monthly means at each grid box. Anyways what you do for SSMIS is completely different compared to ECMWF (or better ERAinterim), here you use the descending F16 orbits only (i.e. Level 2 data), for ECMWF you use daily averages. In general it would be best to make the comparison methodology among the different data sets more consistent (to reduce sampling related issues)!

In the ECMWF comparison we did not compare monthly means at each grid boxes, but we compare co-located daily data and from that we compute a monthly mean bias map. In the revised manuscript, following the suggestion of the referee we use the same gridding and filtering method for all three data set in order to make the comparison methodology among the different data sets more consistent.

P3046,L10: “5.08p.m.” you probably mean a.m.



We mean indeed 5:08 pm. In the revised manuscript, however, we used daily co-located data also for the SSMIS comparison, which we computed from both ascending and descending orbit. We had to use daily co-location in order to adopt the same comparison methodology for all three data set, because the SSM/I+MERIS GlobVapour data set is available only as daily and monthly mean.

P3046,L10: “drifting”: why don’t you use F17 then instead which is drifting less and has a more close overpass time?

We have verified that we have consistent result using the F16 and F17 satellite (slightly lower bias with F16 data). The analysis in the revised manuscript is based on daily co-locations.

P3047,L2: But in case of intense rain, there will be no valid TCWV estimate and therefore no SSMIS-GOME-2 match-up pair, so these cases do not influence your results, they only decrease the number of available match-ups.

We have rephrased the sentence and specified that in rainy conditions the effect is only to reduce the number of measurements.

P3047,L10: “1% per 1h time difference”: how do you get this number?

We here provided the reference to the O3M SAF Validation Report by Kalakoski et al. (2011), where the diurnal variation of the H<sub>2</sub>O VCD from ship borne GPS observations at the tropical Indian ocean is reported.

P3047,L12: “intradaily” => diurnal

Corrected.

P3047,L20f: provide also the overpass time for F13, and F14

Done.

P3047,L26: Why don’t you use the available daily product? This would make the comparison closer to what was done for SSMIS?

P3048,L6: “values are larger..”: therefore better use daily data and make comparison as done for SSMIS see comment above.

As suggested by the referee, in the revised version of the manuscript we repeated the analysis using the daily product.

P3048,L25: “potentially important differences”: which region are you exactly talking about?

Differences over Europe and north America coasts, which are particularly evident e.g. in the southern part of Sweden, along the coasts of the Baltic sea. We added that in the text.

P3049,15ff: also here the area of large differences is situated south of the Sahara, so the attribution to (problems with) high albedo values is questionable. Please provide details and results of the correlation analysis you did

We specified here that the dry bias is correlated with regions with high surface albedo “and land areas with high humidity.” From a correlation analysis we found that the negative bias is correlated both with high humidity regions and with high albedo values (0.2-0.5). See attached Figures 2-3.

P3049,L20ff: it would be more interesting to see these scatter plots for collocated data pairs (so far only possible for SSMIS comparison) and then also for the other comparisons (GOME-2 vs. SSMIS, GOME-2 vs. ERA-interim).

In the revised version of the manuscript we have done the analysis based on daily co-location. We use the scatter plot here in order to show the differences between the old and the new approach.

P3050,L27: “because of the reduced overlap between the GOME-2A- and GOME-2B orbits is lower in the tropics”. Reformulate and explain why.

Rephrased. “since the orbits of the GOME-2A and the GOME-2B sensors have the smallest overlap in the tropical regions.”

P3051,L1: “Assimilated data” => reanalysis data

Corrected.

P3051,L9: “surprisingly good agreement”: Really? What did you expect? Did you do similar comparisons based on the previous retrieval version? Then it would be interesting to see the improvements.

The validation of the GOME-2 H<sub>2</sub>O total column data produced with an earlier algorithm version (GDP 4.5) was already presented in Kalakoski et al. (2011). From the comparison with SSM/I data it was found that GOME-2 generally overestimates the SSM/I data and typical biases between 0.17 and 0.2 g/cm<sup>2</sup> were retrieved for monthly global averages. These values are about one order of magnitude larger with respect to the bias we obtained with the SSMIS or SSM/I+MERIS sample in our analysis (0.006 and 0.032 g/cm<sup>2</sup>). Following the Anonymous Referee #1 suggestion, we have clarified and commented this point in the Introduction.

P3051,L10: why four data sets? You mean three (ERA-interim, SSMIS, SSM/I-MERIS).

Corrected.

#### Editorial Remarks

1. Additional work concerning the language/phrasing is needed. The authors are strongly urged to carefully check their manuscript on spelling and grammatical errors once again. Some examples are given below, but list is by no means exhaustive :

1) P3024,L13: offers => offer

P3024,L22: constrained on => constrained to

P3028,L3: at the 15 of July => on 15 July; check other dates!

P3035,L29: The net effect...is to reduce => is a reduced bias

P3044,L22: “expected”, rather estimated or similar..

P3047,L8: North Ocean Atlantic => “North Atlantic Ocean”

P3049,L5: higher data => higher data values

...

Done.

2. There are many abbreviations which are not introduced at all, introduced several times or not introduced when used first but later, for example GOME-2A GOME-2B, HOAPS, WV, UVN, O3M-SAF, ERA-interim, TCWV.. This list not exhaustive. Please check carefully.

Done.

3. Figures and Tables

a. Several figures are too small and therefore hard to read, Fig. 2, 3, 7, 8. Their size needs to be increased.

We increased the size of most of the figure in our new manuscript version which uses a different journal class file.

b. Table 2: As the spatial coverage differs between the data sets statistics should be provided separately for ocean (maybe also coast) and land areas. It is not mentioned that SSM/I+MERIS statistics are based on a shorter time period only. How do you calculate the margins you provide for bias and RMSE?

In the revised version of the manuscript, we provided bias statistics also for ocean and land areas separately. The uncertainty margins we provide for the bias and RMSE statistics result from the (non-weighted) spread of the bias and RMSE values in the time series.

c. Fig7: How did you calculate the global means? Did you apply area-weighted average or just averaged over all grid boxes?

For the computation of the global means we simply averaged the bias over all grid boxes. This approach might give larger weight to Arctic latitudes.

## REFERENCES

- Antón, M., Loyola, D., Román, R., and Vömel, H.: Validation of GOME-2/MetOp-A total water vapour column using reference radiosonde data from GRUAN network, *Atmos. Meas. Tech. Discuss.*, 7, 9573-9601, doi:10.5194/amtd-7-9573-2014, 2014.
- Bentamy, A., Katsaros, K. B., Mestas-Nunez, A. M., Drennan, W. M., Forde, B. E., and Roquet, H.: Satellite estimates of wind speed and latent heat flux over the global oceans, *J. Climate*, 16, 637–655, 2003.
- Grossi, M., Kalakoski, N., and Valks, P.: O3M SAF ORR VALIDATION REPORT, SAF/O3M/DLR/ORR/H2O, Issue 01/2013, 2013.
- Hovila, J., Hassinen, S.: Service Specification Document Issue 1.1, SAF/O3M/FMI/RQ/SESP/001, 2013.
- Kalakoski, N., Wagner, T., Mies, K., Beirle, S., Slijkhuis, S., and Loyola, D.: O3M SAF Validation Report, Offline Total Water Vapour, SAF/O3M/FMI/VR/H2O/111, 2011.
- Kalakoski, N., Kujanpää, J., Sofieva, V., Tamminen, J., Grossi, M., and Valks, P.: Comparison of GOME-2/MetOp total column water vapour with ground-based and in-situ measurements, *amt-2014-292*, 2014.
- Kunkee, D. B., G. A. Poe, D. J. Boucher, S. D. Swadley, Y. Hong, J. E. Wessel, and E. A. Uliana. 2008. Design and Evaluation of the First Special Sensor Microwave Imager/Sounder. *IEEE Trans. Geosci. Remote Sens.* 46, no. 4: 863–883.
- Lambert, J.-C., Atmospheric Service Validation Protocol, MACC-II Deliverable D\_153.1, May 2013.
- Schneider N. and Schröder, M.: GLOBVAPOR, Project Validation Report (PVR) for SSMI+MERIS, Issue 3, revision 0, 2012.  
[http://www.globvapour.info/download/GlobVapour\\_D19\\_PVR\\_SSMI\\_MERIS\\_V3.0.pdf](http://www.globvapour.info/download/GlobVapour_D19_PVR_SSMI_MERIS_V3.0.pdf)
- Schröder, M. and Bojkov, B.: GLOBVAPOR, Project Validation Report (PVR) of GOME + SCIAMACHY + GOME-2, Issue 3, revision 0, 2012  
[http://www.globvapour.info/download/GlobVapour\\_D19\\_PVR\\_GOME\\_SCIAMACHY\\_V3.0.pdf](http://www.globvapour.info/download/GlobVapour_D19_PVR_GOME_SCIAMACHY_V3.0.pdf)
- Valks, P., Loyola, D., Hao, N., Rix, M., and Slijkhuis, S. S.: Algorithm Theoretical Basis Document for GOME-2 Total Column Products of Ozone, Minor Trace Gases and Cloud Properties (GDP 4.5 for O3M-SAF OTO and NTO), DLR/GOME-2/ATBD/01, Iss./Rev.: 2/GG, 2012.
- Wagner, T., Heland, J., Zöger, M., and Platt, U.: A fast H2O total column density product from GOME – Validation with in-situ aircraft measurements, *Atmos. Chem. Phys.*, 3, 651–663, doi:10.5194/acp-3-651-2003, 2003.
- Wentz, F.J., "A well calibrated ocean algorithm for special sensor microwave/imager", *Journal of geophysical research*, vol. 102, NO. C4, Pag. 8703-8718, April 15, 1997

## FIGURES

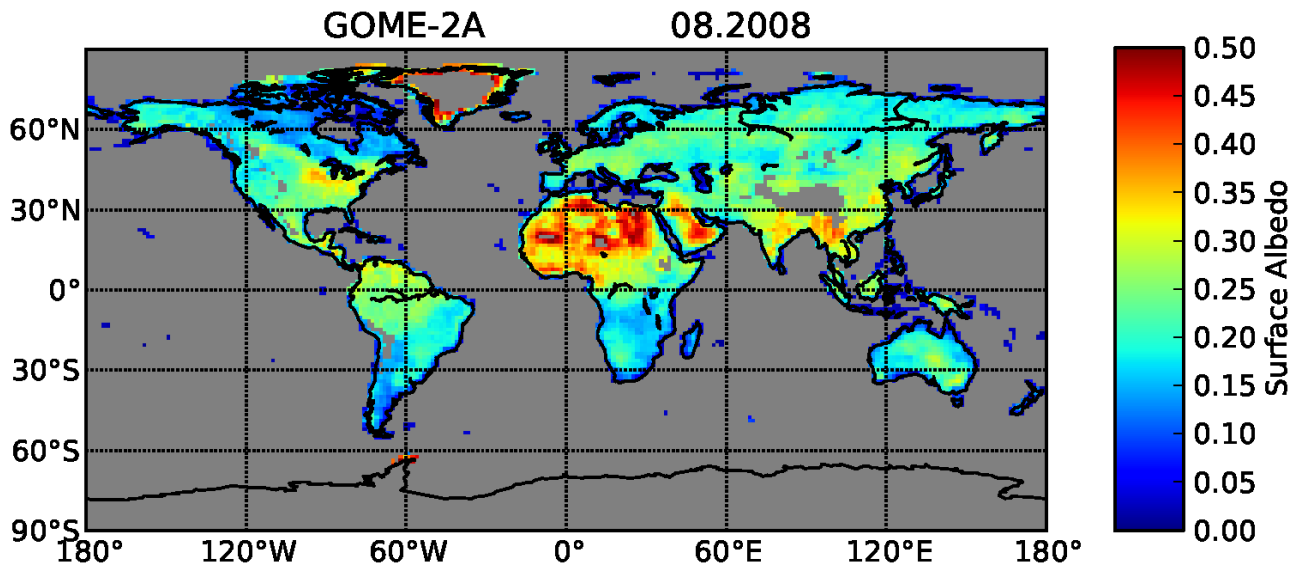


Figure 1: Surface Albedo retrieved with the GOME-2/MetOp-A satellite in August 2008 over land.

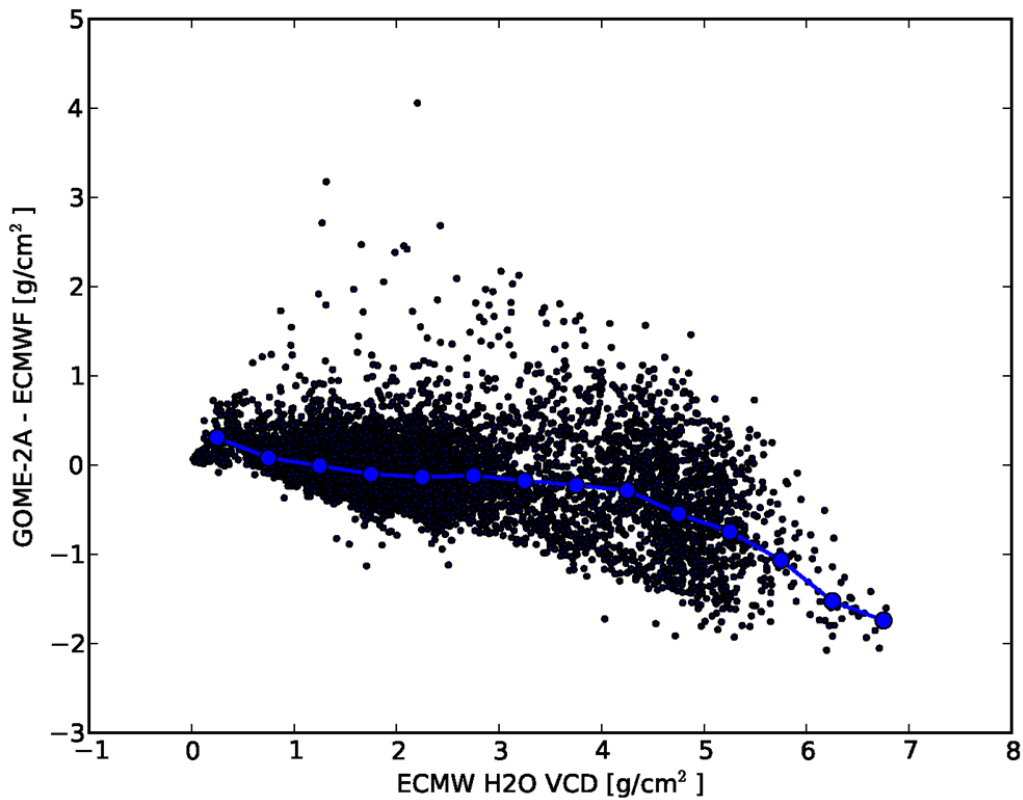


Figure 2: Mean difference between H<sub>2</sub>O VCD data retrieved by GOME-2A and ECMWF as a function of the ECMWF H<sub>2</sub>O VCD values in August 2008 over land.

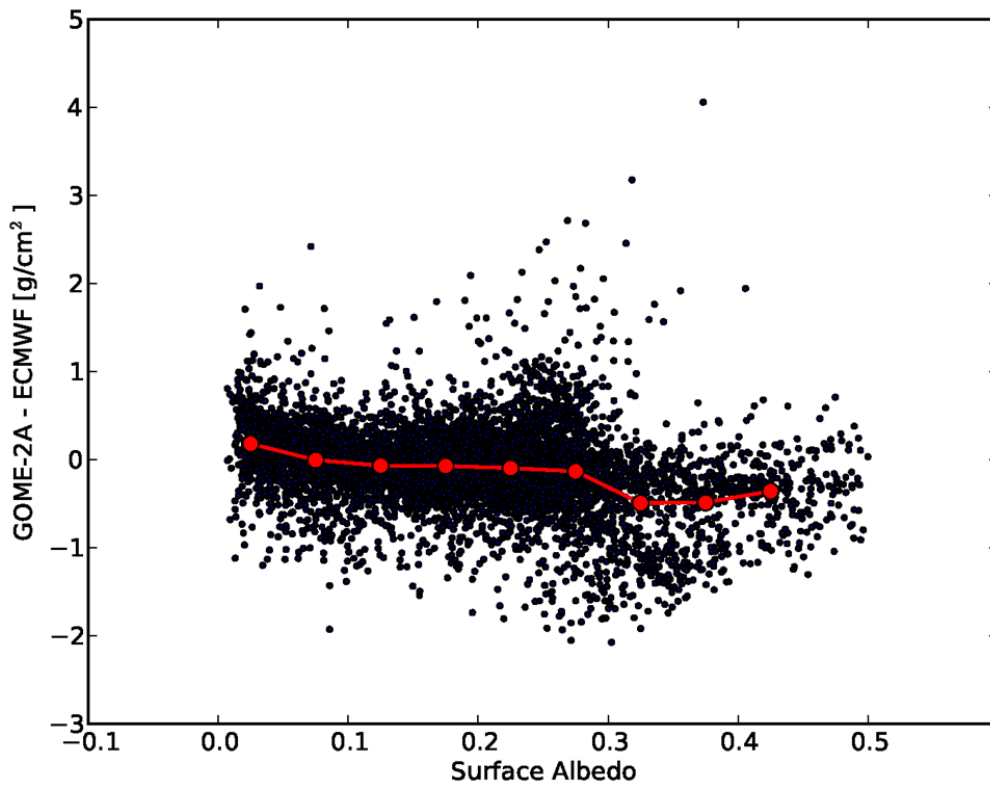


Figure 3: Mean difference between H<sub>2</sub>O VCD data retrieved by GOME-2A and ECMWF as a function of GOME-2A satellite surface albedo in August 2008 over land.