

Authors' response to Interactive comments of Anonymous Referee #2 on "Comparison of GOME-2/Metop total column water vapour with ground-based and in situ measurements" by N. Kalakoski et al.

We thank the referee for careful reading of our manuscript and for the helpful comments. We will incorporate these comments to the revised manuscript. Below, we list the referees' comments followed by our answers (in blue).

The manuscript "Comparison of GOME-2/Metop total column water vapor with ground-based and in situ measurements" submitted for publication in AMT confronts co-located radiosonde and GPS measurements of total column water vapor with satellite observations for a period from 2007 to 2013. The authors find a general good agreement. A wet bias for the GOME-2 data for small water vapor amounts and a dry bias for large amounts of water vapor are revealed. The presented analysis is an important contribution to the monitoring of the quality of water vapor measurements from satellite instruments. The applied methodology is well explained and sound. The results and conclusions are clearly written and meaningful figures are shown. However, some slightly more detailed explanations at some points could help to improve the readability. Thus, I recommend the publication in AMT after addressing my comments below.

General comments:

1. I would like to suggest that you add some sentences to the introduction on the importance of the work, i.e. the monitoring of the quality of satellite water vapor measurements, which is also important for climate and trend studies. Further, it is worth to mention that until now there is a lack of confronting GOME-2 with radiosondes and GPS.

Introduction will be expanded accordingly.

2. Regarding the differences between GOME-2 and radiosonde / GPS data you did not perform any significance analysis in the sense of statistical hypothesis testing. Except for the drifts (p. 12525, l. 24) you performed a significance test. Can you give a justification for your approach not to perform hypothesis tests. Important literature about this discussion is e.g. von Storch and Zwiers (2013).

Significance analysis was performed in some cases, but omitted from discussion and figures since due to large sample sizes biases were found to be statistically significant and error bars made figures even more confusing. However, we agree that the statistical significance of the results should be discussed explicitly. We will improve this area in revised manuscript.

3. You decided to focus on relative differences, do you have an explanation for that? Relative differences emphasize the polar regions with very low water vapor columns. A large relative difference of e.g. 50% could be related to only 2.5kg/m², whereas at the equator a small relative difference of e.g. 10% could be related to 6kg/m² (assuming 5kg/m² in polar regions and 60kg/m² at the equator).

We chose to focus on relative differences to better present the differences seen at low water vapour abundances. The large absolute differences observed at high water vapour abundancies tended to mask any differences seen at lower abundancies. However, we agree that this is a delicate trade-off.

Specific comments:

1. p. 12525, l. 6: How is the correlation between GOME-2 and ground based measurements computed? Is it based on fitting the histogram in Fig. 2? Is the fraction of hits used as a weighting? Or did you estimate the correlation from time series? A little explanation would be helpful.

Correlation is calculated from the original co-locations, not from the histogram in Fig. 2. We will add explanation to the revised manuscript.

2. p. 12527, l. 13: It is very interesting to see this large dependence on the albedo. Can you specify how these “Further developments ...” on the GOME-2 algorithm and albedo database could look like?

Possible further developments are discussed in paper by Grossi et al. (2015). They point out, for example the effect of absorbing aerosols on observing the surface albedo. They have plans to further study the effect of the surface albedo on retrieval.

3. p. 12528, l. 14: Regarding Fig. 7, you speak about peculiarities in the GPS data and the exclusion of these “artifacts”. I wonder how you differentiate between such a “peculiarity” and a “real” bias? Do you have a filtered version of Fig. 7, where these events have been removed?

As it is, we have not explicitly differentiated between “peculiarities” and “real” biases. Peculiarities are caused by the poor temporal and spatial representativeness of the GPS retrievals. We will consider screening the Figure 7. for significant biases, as per your comment 2.

4. p. 12529, l. 9: I absolutely agree that a combined use of GOME-2A and GOME-2B might be beneficial. However, one should be careful. While there are negligible differences of the means of GOME-2A and GOME-2B there might be larger differences on grid pixel basis. Regarding e.g. trend studies, a combination of two or more time series could be accomplished by using a “level-shift”-model as done for the combination of GOME and SCIAMACHY data (Mieruch et al., 2014).

Agreed, we will add a caveat regarding the possible complications in combined use.

Technical corrections:

1. p. 12519, l. 22: troposheric → tropospheric
2. p. 12529, l. 2: ... dependence on GOME-2 ... → ... dependence of GOME-2 ...

Corrected.

References:

Grossi,M., Valks,P., Loyola,D., Aberle,B., Slijkhuis,S., Wagner,T., Beirle,S., and Lang,R.: Total column water vapour measurements from GOME-2 MetOp-A and MetOp-B, *Atmos. Meas. Tech. Discuss.*, 7, 3021--3073,