REPLY TO REVIEWER #3

The authors highly appreciate the constructive comments. They are very useful contributions that will certainly help to improve the revised manuscript. In the following, the authors reply point by point to all Reviewer comments, which are written in italic while our replies are in standard font. Within the manuscript all changes from the submitted version are highlighted in red.

MAJOR COMMENTS OF REVIEWER #3:

The RL is obviously a superior methodology to the MWR for humidity retrievals. Although the authors use the MWR measurements to improve the RL retrievals where retrievals from the latter are not reliable, the MWR retrievals themselves have very little information above the first kilometer. It is possible that the RL profiles could be improved above the boundary layer just by choosing a better climatology or may be a model output without going to the extent of doing an optimal retrieval estimation. The only place where I could see some real advantage of using MWR measurements is in the lowest 200-500 m although in the case shown the results appears mixed.

Actually, for water vapor remote sensing using the 22.235 GHz water vapor line, the weighting functions show only a weak dependence on height (see http://cfa.aquila.infn.it/wiki.eg-climet.org/index.php5/MWR_Fundamentals). However, this weak dependence caused by pressure broadening of the line, makes it possible to obtain limited information (2 degrees of freedom for signal (DOF)) on the vertical distribution of water vapor in the first place. Note, that the weighting functions near the line center actually increase from surface upwards. Thus, MWR measurements between 20 and 30 GHz (K-band) can provide water vapor information throughout the troposphere and are not limited to the boundary layer (as in case of temperature profiling). This is quantified in terms of DOF in Fig. 9, which shows that information from the MWR (accounting to roughly 1.5) is added up in the region above 2.5 km. This information is most prominent for the daytime profiles, when the lidar presents a weaker performance (Fig. 7). Furthermore, the effect is illustrated in real measurements for an individual radiosonde ascent in Fig. 2, with the superior performance of the joint retrieval at 3000 m height.

It is also important to note that the full benefit of the method will be the application to cloudy scenes where the lidar is strongly limited.

The second reservation is about the conclusions as I am not sure that the results entirely support the conclusions.

For the sake of clarity, the conclusions have been modified as follows:

"The improvements of merging both instrument systems have been consistently analysed in terms of both the reduction of the theoretical error and the increase of DOF. Significant advantages of instrument synergy are clearly shown above the highest valid lidar signal. For example, when applying the combined retrieval to the complete HOPE period, the absolute humidity theoretical error above ~3 km is reduced by a factor of 2 with respect to the case where only lidar is used. The addition of the MWR information to the RL results in 1.6 additional degrees of freedom per signal, which are mainly distributed in the layers above the lidar noise threshold. The synergy presents its strongest advantages in the regions where RL data is not available, whereas in the regions where both instruments are available, RL dominates the retrieval."

SPECIFIC COMMENTS:

Page 8: Line 225, Eq. 6: Can the author provide a reference for this definition of vertical resolution? Usually the Backus-Gilbert technique is used to define the vertical resolution from the spread function. An example of application of this technique to determine the microwave radiometer

vertical resolution can be found in Westwater and Snider and Carlson (J. Appl. Meteor., vol. 14, pp. 524–539, 1975).

There are many different ways to define vertical resolution. In addition to the example given by the reviewer, Liljegren (2004) exploited the interlevel error covariance. Here we follow the approach based on the optimal estimation as presented by Rodgers (2000) in pages 52, 53 and 54:

"*Resolution*, like *information*, is a word with a multiplicity of meanings, and tends to be used differently in different contexts. […] Possibilities include characteristics of the averaging kernel or state resolution matrix, such as the width of the averaging kernel or the point spread function, where *width* has many possible interpretations, the response of the retrieval to sine wave perturbations in the state, and the range of heights covered divided by number of independent quantities measured.
[…]
Possible characterisations of resolution are […]
(iv) the degrees of freedom for signal, d_s is the trace of the averaging kernel matrix. Consequently the diagonal of A may be thought of as a measure of the number of levels per degree of freedom, and thus a measure of resolution. "

This is the definition of equation (6). Other authors have previously used this definition for vertical resolution, e.g. Liu (2014).

Page 16 Fig. 6. The results in Figure 6 are mixed. In the upper troposphere the RL seems to have the lowest bias up to 4 km. Above 4 km the combined retrievals show a very small improvement, however what the standard deviation is considered I am not sure that the improvement is clear.

We included section 5.2 to show the retrieval performance in comparison to an independent measurement of humidity profiles. As we mention in the paper, this comparison is not conclusive because it is based on a limited number of samples and the agreement in terms of the atmospheric volume sampled by the radiosonde and the remote sensing systems is not perfect. In the manuscript (line 429-433) we only state that there is, on average, in region c, a small improvement of the combined retrieval with respect to RL retrieval both in term of bias and standard deviation. We are very careful to not state any improvement for the combined retrieval in the other regions

Page 17, section 5.3 and 5.4 I am not sure what is intended by theoretical error shown in Fig. 7. I think the author means the "a posteriori" covariance. However this measure of uncertainty, although necessary, represents a partial picture. A better estimate of "error" intended as RMS Error is the one you provide in the comparison with radiosondes in Fig. 6. The authors should probably change "theoretical error" with "covariance" if this is what they meant. Otherwise they should explain what they mean by "theoretical error". It is not clear how the uncertainty shown in Fig. 7 and 8 relates to the error bars shown in Fig. 2 (the text says they are both computed from Eq. 4), however the values seem considerably different.
The reviewer is right. A clarification is needed and introduced in the new manuscript as follows:

"From S_{op}, the theoretical error (in kg/m^3) associated to each altitude of the retrieved profile x_{op} is calculated as the square root of the main diagonal elements in S_{op}. The word *theoretical* emphasizes that it is an a posteriori estimate, and not a direct difference to a given reference".
In particular the error bars above 2 km in Fig. 4 seem to be ~0.5 g/mˆ3, but they seem smaller in Fig. 7. Or it is just due to the different scales of the plots?

The values on Fig. 2 represent the theoretical error for one specific profile, while in figure 7 and 8 the values correspond to the **mean** theoretical error, **averaged over more than 600 profiles**. That is the reason why the values for the particular example in Fig. 2 might differ from the ones presented in Fig. 7 and 8.

I am not sure I entirely understand the difference in what is plotted in Fig. 7 and 8, besides the classification between daytime and nighttime. Could you please explain that more clearly?

On the one hand, Fig. 7 is presented in order to illustrate the benefit of the synergy separating day and night measurements. The reason for that is that the lidar presents a better performance during nighttime periods. During these periods, the vertical contribution of the MWR is reduced to the lidar overlap region in the lowest atmosphere. Nevertheless, during daytime periods, the MWR contribution to the retrieval in higher atmospheric layers can be stronger, due to the lack of lidar data. The authors wanted to highlight this difference.

On the other hand, Fig. 8 is presented as an average over the complete HOPE period, allowing us to provide an overview through the complete campaign. Here, an artificial clipping of the lidar data up to 2.5 km has been performed. Its aim is to describe the behavior of the instrument combination with a large number of profiles, in a situation where the regions with and without lidar data availability are clearly defined. In contrast, in Fig. 7, no artificial clipping for the lidar is performed.

Several clarification sentences have been included in different parts of section 5.3. (Please, see red sentences in new manuscript version).

Page 25 line 576: "The improvement of the synergy have been analyzed in terms of several parameters like the reduction of the theoretical error or the increase in DOF, showing significant advantages. . ." I am not entirely sure about the accuracy of this statement. The theoretical error (which is the a posteriori covariance) is related to the DOF. The two metrics are not independent and essentially convey the same information in different form. Although it is true that the analysis shows the reduction of the covariance after the retrieval, the comparison with the radiosondes conveys mixed messages about the actual usefulness of the MWR measurements.

We do not claim that theoretical error and DOF are independent measures. They state two measures with which we are able to characterize the benefit of instrument synergy. We show both measures, which underline the consistency of our results. The main advantages of MWR and lidar synergy are shown to occur above the height of the minimum lidar sensitivity. This fact becomes clear in Figs. 2, 3, 6 (right panel), 7, 8 and 9. Indeed, Fig 6 (left panel) does not show a clear improvement of the combined retrieval considering bias error. However, it does not allow to draw a conclusion from where the bias originates. We have modified the text and now it states:

"The improvements of merging both instrument systems have been consistently analyzed in terms of both the reduction of the theoretical error and the increase of DOF. Significant advantages of instrument synergy are clearly shown above the highest valid lidar signal."

ENGLISH CORRECTIONS:

There are a few English corrections needed:
Page 24 line 542 "is chosen kind of arbitrary" can be rephrased: "The increase in RL measurement uncertainty is arbitrarily chosen based on. . ."
Page 25 line 569: ". . .synergy of different sensors has become come more. . ." remove come Page 576: "several parameters like" "like" can be replaced with a colon.

The English corrections have been corrected in the manuscript.