REPLY TO REVIEWER #2

The authors really appreciate the constructive comments. There are very useful contributions that will certainly help to improve the revised manuscript. In the following, the authors reply point by point to all Reviewers' comments. In particular, we refined the cut-off analysis of the lidar, introduced new comparisons for the integrated water vapor and removed section 5 and 6 on the temperature and relative humidity case study.

MAJOR COMMENTS OF REVIEWER #2:

The reviewer's comments are written in italic. Changes from the submitted version of the manuscript are highlighted in the revised manuscript in red.

1. The specification of the RL measurement covariance matrix is oversimplified considering only random uncertainty.

In principle, the specification of the RL measurement covariance matrix should also include other error sources affecting RL measurements besides the Poisson uncertainty. However, the overall error affecting water vapour RL measurements (typically lower than 5%) is usually much smaller than the random error, especially for daytime observations. Observation error covariance matrix for lidar systems including only the random error has been considered also by other authors (Wulfmeyer et al., 2006; Dunbar et al., 2014; Adams et al. 2015).

2. The OE retrieval is analyzed in terms of error reduction and degrees of freedom (DOF). The presentation of the averaging kernels (AVK), which is a classical diagnostic of OE retrievals, is missing.

We do agree with the reviewer's opinion: the averaging kernel representation includes good information for understanding the retrieval. Nevertheless, we argue against showing plots of averaging kernel matrices in the manuscript, not to distract the reader. Nevertheless, for the expert reader, we include in the appendix the plots for averaging kernels (using different settings of the a priori covariance matrix), Jacobians and retrieval vertical resolution. We believe that the proposed assessment based on the theoretical error and degrees of freedom for signal provides a thorough insight into retrieval performance. In addition, the manuscript is already quite long and contains a considerable number of plots.

Reviewer #1 shares also this opinion with us: "*I agree that averaging kernels are best placed in supplementary material as, though vital to appreciate the function of an OE retrieval, they are confusing to inexperienced readers*".

3. For all water vapor results shown, the RL data have been artificially clipped at an arbitrary altitude (2.5 km for absolute humidity) for practical reasons. Consequently, the reader does not get any flavor of the nighttime performance, nor of the quality of the optimal product, i.e. using all lidar data available.

With the cutting at 2.5 km of all profiles we aim to investigate the information content of the different altitude regions where lidar data is typically available and where not. We define three regions: (a) from ground to 180 meters, (b) from 180 to 2.5 km and (c) from 2.5 km above. This simplification of the problem allows us to robustly analyze the impact of the MWR in the OEM and thus, assess its potential benefits (see Table 1 and Fig. from 7-9 in the new manuscript).

Nevertheless, we still agree with the reviewer in this point. In order for the reader to get a better flavor of the lidar performance during day and night times, we include and explain figure R-4, new Figure 6 in the manuscript. From the figure one can see that lidar performance is much better during nighttime than daytime. The theoretical error during night is much lower and the profiles reach a higher altitude. During daytime, the highest profiles reach a maximum altitude of around 5.5 km. Note that we have considered meaningful RL data when its relative error is smaller than 100%, same than in the rest of the manuscript.

In addition, this figure allows us to better justify our statistical analysis based on the 2.5 km threshold. We set this value to keep 85% of the profiles in our statistics (see right part of the figure). Lidar profiles which do not reach this altitude are not taken into account for the statistics.

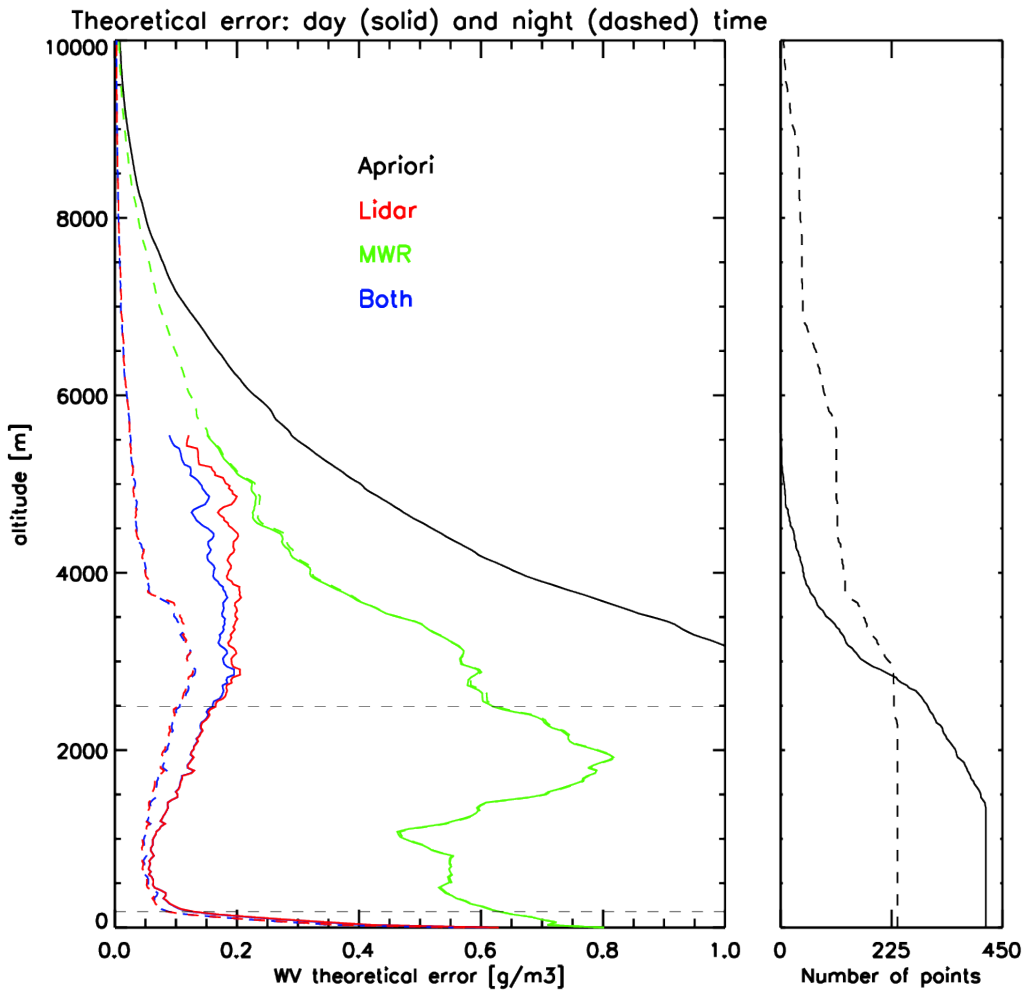


Figure R-4 : Theoretical error for the different cases where only-RL is taken into account (red), only-MWR (green) and the combination (blue). The dataset is studied in two sub-sets: solid line represent daytime measurements, dashed lines represent nighttime measurements.

4. The analysis and discussion of the temperature retrieval and the simultaneous retrieval of temperature and absolute humidity is poor compared to Section 4 relying on one single comparison with radiosonde (RS)

Unfortunately, we lack of more examples for RL temperature profiles. We do agree with the reviewer's comment and, since our dataset is very restricted, have decided to do without these two sections.

5. The authors explain that "the covariance matrix associated with the MWR measurements was obtained empirically by calculating the correlation between the different channels, while constantly viewing an ambient black-body target with known temperature".
- How is the covariance derived from correlation?
- If equation (8) is used, how are s_a and s_b determined?
- Is the same covariance assumed for all elevation angles?
- If so, how is this justified?

The text has been modified, we meant: "*the covariance matrix associated with the MWR measurements was obtained empirically by calculating the covariance between the different channels, while constantly viewing an ambient black-body target with known temperature*"**.**

The value for the covariance is assumed to be independent of the elevation angle. For optically thick channels (above 54 GHz, respectively at low elevation angles) this is easily justified because the brightness temperatures are all within a few Kelvin of the ambient temperature. However at the optically thin K-band we must assume that the uncertainty characteristics of the MWR are similar throughout the whole range of measured TBs (20 - 50 K). Currently, we do not see any reason why this should not be the case (i.e. due to a significant signal to noise reduction with decreasing input power). However, we are currently planning a measurement campaign to test this hypothesis (i.e. via liquid nitrogen views and auto-correlation analysis).

Even less information is provided for the RL part in Sy: "the part of Sy corresponding to the RL is defined as a diagonal matrix containing the variances of every altitude". The RL measurement is itself a retrieval and it needs to be specified which error sources have been taken into account for its uncertainty budget (calibration, FOV, saturation, background, . . .). Reference to corresponding publication is needed. The matrix is assumed diagonal. However, uncertainty due to calibration (5% as specified in 2.1), FOV, saturation or background, for example, introduce correlation between the errors at different altitudes which may have consequences on the retrieval. This issue must be properly addressed and discussed in the paper. In particular, Section 4.2.5 must be revised in this respect.

The RL measurement is - technically speaking - an indirect measurement, where the atmospheric product (water vapour mixing ratio or temperature profile) is obtained using an analytical expression relating to the measured lidar signals. Lidar signals are range resolved (i.e. expressed as a function of time and range, or altitude if vertically pointing as in our case). So the analytical expression returns atmospheric products defined at each altitude (vertical profiles) as the lidar signals. Lidar signals at different altitudes represent independent measurements, so one can say that there is no correlation between lidar-derived atmospheric products at different altitudes as long as no vertical smoothing is applied to the data.

As specified above in this reply to the reviewer, we are aware that in principle the specification of the RL measurement covariance matrix should also include the consideration of other error sources affecting RL measurements, besides the uncertainty derived from Poisson statistics. However, the overall error affecting water vapour RL measurements (typically lower than 5%) is usually much smaller than the random error, especially for daytime observations, and consequently in the present work other error sources have not been accounted for in the computation of the RL measurement covariance matrix.

Observation error covariance matrices for lidar systems including only the random error have been also considered by a variety of other authors (Wulfmeyer et al., 2006; Dunbar et al., 2014; Adams et al. 2015). Additionally, with respect to the systematic error sources mentioned by the reviewer in his/her comment: no systematic error associated with the FOV, saturation or background is present in the reported Raman lidar measurements. Consequently, no correlation with the errors associated to these sources has to be introduced.

Furthermore, the uncertainty associated with the calibration certainly does not introduce correlations between the random errors at different altitudes. Potentially, it would introduce correlation only between the systematic errors at different altitudes. However, to overcome this risk we wish to better specify what was the approach we considered to calibrate the RL measurements. During HOPE, BASIL has been calibrated based on the comparison with the radiosondes launched approximately 4 km away from the instrument. Mean calibration coefficients for both water vapour mixing ratio and temperature measurements were estimated by comparing BASIL and radiosonde data considering all radiosondes launched at times when BASIL was running (approximately 60 comparisons). Considering only one calibration single value (any day at any altitude) throughout the HOPE field campaign minimizes potential correlation between the errors at different altitudes associated with the calibration procedure.

Given all the limitations of the external calibration of the lidar as explained in Section 2.1, I would suggest to include the lidar calibration factor in the forward model and to treat it as a retrieval parameter. This would resolve to a good extent the issue of correlated errors of the RL input data and introduce the capability to improve the a priori RL calibration by combining it with MWR observations, another strength of the combination of RL and MWR.

While the proposed approach could be implemented, we believe this would represent a completely different methodology to face the problem. Here we are pursuing the approach of merging the information coming from two different sensors (MWR and RL) to determine atmospheric humidity profiles based on the application of an Optimal Estimation Method. In this respect, calibrated RL measurements are considered to be a stand-alone independent piece of information. For this purpose, we are assuming that the consideration of all available radiosondes (60) for the determination of a single calibration constant (considered any day at any altitude) prevents the dependence of a single lidar vertical profile from the simultaneous radiosonde profile used for the calibration. A different approach would be the one performed by Foth et al. 2015, where the authors perform a calibration of the RL by adding the MWR measurements.

On the other hand, an error on the calibration constant will induce a bias on the IWV calculated by the lidar. We calculate the bias in the IWV between the MWR and the RL, and find a bias of -0.4 kg/m2. This small bias induces to think that the calibration constant has been properly estimated.

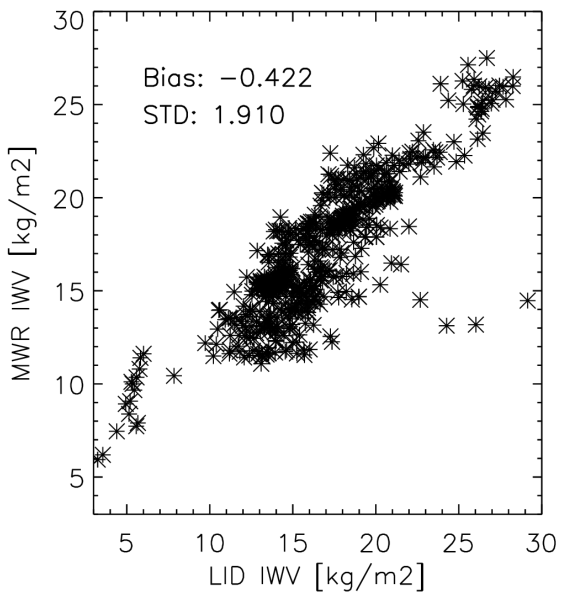


Figure 1: Scatterplot for the IWV from the RL and MWR

For all water vapor results presented, the RL data have been clipped at an arbitrary altitude for practical reasons. This is somehow in disagreement with an "optimal combination" of the two instruments.

This point has been defended already in the reply to reviewer#1, on the 4th of his major comments, where we explain our motivations to cut the RL profile. Please, vid supra.

MINOR REMARKS OF REVIEWER #2:

- ***P5469, l9: I think this statement is too strong. Many publications show that the lidar technique provides the high spatial and temporal resolution needed to study convection and turbulence.***

The manuscript has been modified as follows: "*Highly resolved, accurate and continuous measurement of these parameters, in particular water vapor, are required for a deeper understanding of many atmospheric phenomena (Stevens and Bony, 2013) Specifically processes on short time scales such as convection, cloud formation or boundary layer turbulence cannot be resolved completely by any single instrument alone. In order to overcome this limitation, the scientific community started merging different data from several instruments in the last 15 years.* "

- ***P5470, l1: replace "the quality" by "the range"***

Modified in the manuscript.

- ***P5472, l9: replace "frequency" by "repetition rate"***

Modified in the manuscript.

- ***P5472, l28: here and in the entire paper, specify if height is given above ground or sea level.***

Modified in the manuscript.

- ***P5473, l1: What kind of "problems" is referred here?***

We used an inappropriate wording. What is described here is not an OVF "problem", but an OVF "effect". The sentence was reformulated as follows: "This limitation is caused by the OVF effect, that is the absence of overlap between the laser beam and receiver field-of-view."

- ***P5473, l9: It is not clear how the lidar is calibrated. The calibration coefficient used in this study is the mean of all calibration coefficients derived from all individual BASIL-radiosonde intercomparison?***

This aspect has been clarified above. For the sake of clarity, we recall it here:

During HOPE, BASIL has been calibrated based on the comparison with the radiosondes launched approximately 4 km away from the instrument. Mean calibration coefficients for both water vapour mixing ratio and temperature measurements were estimated by comparing BASIL and radiosonde data considering all radiosondes launched at times when BASIL was running (approximately 60 comparisons). Only a single value for each calibration coefficient, which is considered valid any day and at any altitude, is then applied to all lidar profiles.

- ***P5375, l16: the forward model acts in the other direction: the measurement is calculated from the atmospheric state. To invert the forward model, OE is used.***

Modified in the manuscript.

- ***P5477, l1: the underlying assumption in OE is that the variables follow a Gaussian distribution. Has it been investigated if q is "sufficiently" Gaussian for the analyzed period? Have the authors tested also to retrieve log(q) instead? Please comment on this.***

Both q and log(q) are not perfectly Gaussian. This varies with the altitude. We had performed some sensitivity studies before implementing the algorithm and have decided for the linear form. In literature, other studies use also the linear form in their humidity retrievals (Ebell 2014, Turner 2014, Löhnert 2009).

- ***P5478, l18: replace "profile" by "data point", or so.***

Modified in the manuscript.

- ***P5479, l12: This phrase is not clear. Underline that level 1 data is used from MWR and level 2 data from RL.***

Modified in the manuscript.

- ***P5479, l14: The statement "a clear forward model cannot be defined" is certainly not true. The lidar equation is a form of the Beer-Lambert law, as is the radiative transfer equation for MWR, and is implemented in many languages.***

Modified in the manuscript.

- ***P4579, l16: In this section "correlation" and "covariance" are used as synonyms, please rectify.***

Modified in the manuscript.

- ***P5479, l20: using angular data, the dimension of the MWR covariance matrix would be 27x27, no? please correct and complete.***

This part has been eliminated from the manuscript.

- ***P5480, l4: what do you mean with "scaling the temperature grid"?***

We just meant adapting the original lidar temperature grid to the retrieval grid. But this section has been removed from the manuscript.

- ***P5480, l16: what is the motivation to have a dynamic retrieval grid? Wouldn't a MWR retrieval also run on a 30 m grid?***

There are two motivations. First, we want to keep the high vertical resolution from the lidar when this instrument is available. When no lidar data is available, the MWR is only able to contribute with a very limited resolution. The algorithm could be run using a constant 30 meters grid, but the amount of (un) necessary amount of computation time for doing the calculation of the Jacobians would make the algorithm very un-handy without providing any increase in accuracy.

- ***P5482, l17: Why is Figure3 only shown up to 5 km? All other Figures go up to 10 km, so please show Figure 3 also up to 10 km to illustrate the varying performance as a function of time and height. I suggest to add a dashed line to mark the boundaries of the RL data.***

On average, half of the water vapor is contained in the boundary layer, where also the strongest changes occur. In particular, figure 5 shows the effect of a cold front passage around 23 UTC leading to an interesting intrusion of dry air into the lower altitudes, which we want to highlight. Above 5 km, no significant changes occur. In addition, with this resolution, it becomes clear that the MWR can complete the blind region of the lidar below ~180 m.

- ***P5483, l17: see comment P5480, l16.***

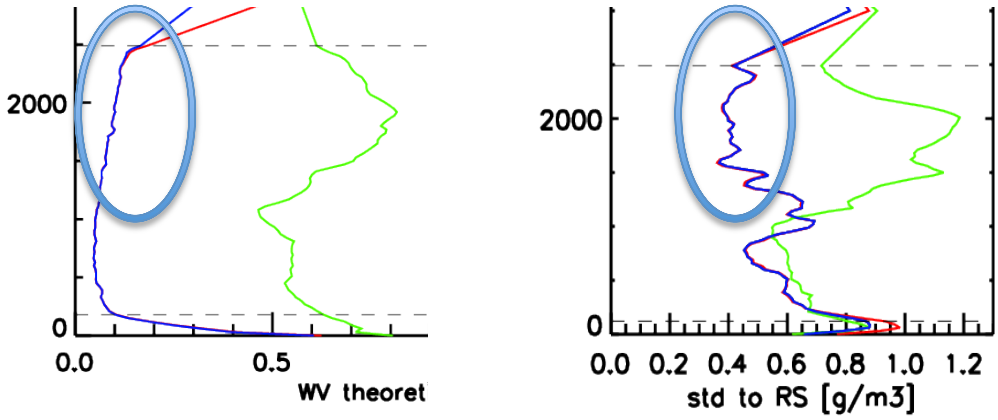
Vid supra.

- ***P5486, l13: The reduction of the error depends on altitude, however, here one "integral" value is presented. It is not clear from the text how this number has been obtained. Is this the reduction of the error averaged over all altitudes? Please clarify the text here and for all other occurrences of this error reduction estimation.***

A clarification has been introduced in the manuscript, at the end of section 4.2.3.

- ***P5487, l10: Sy has hopefully not been determined to the author's best knowledge. Instead, a very much simplified version of Sy has been assumed neglecting systematic error terms which are correlated. How is it justified to increase the measurement error by, as it seems, an arbitrary factor 4 and without introducing correlation? See my major comments.***

At some point, we feared that the first estimation of the lidar uncertainty was probably underestimated. In order to better understand the influence of the lidar uncertainties, we performed a sensitivity study: an increment in the lidar error. But we did not choose an arbitrary value for the increment. First in Figure 6, when we analyzed the theoretical error, we noticed that the error has an approximate value of 0.1 g/m3 in the region around 2 km altitude. Nevertheless, when we compare the random uncertainty to the radiosonde, the value is around 0.4 g/m3 in the same height region. This is four times larger than what the theoretical error is pointing out. That is the reason why we chose a factor of four.



- ***P5491, l5: It is not possible to retrieve temperature (T), absolute humidity (q) and relative humidity (RH) simultaneously, but either T and q or T and RH. Given Sa from Figure 1, I guess T and q are retrieved here, and RH is calculated afterwards. Please clarify the text. In my opinion retrieving T and RH would be an interesting approach. Have the authors investigated this, could you comment?***

Deleted from the manuscript.

- ***P5491, l8: Are the numbers in the right hand Panel of Figure 10 really the differences averaged vertically? I consider this bad practice, because positive and negative differences cancel each other. For example, is a value of 2.46% in any way representative for the differences between "combined (no RL temperature" and "RS", which are in the order of 10-20% at most altitudes?***

Deleted from the manuscript.

- ***P5492, l18: this is shown for absolute humidity, partially for temperature (DOF in Table 3) and not at all for simultaneous T, q retrievals.***

Deleted from the manuscript.

- ***P5493, l2: "The joint information . . ." there is not enough evidence for this conclusion.***

Deleted from the manuscript.

- ***P5493, l4: I guess RH has not been retrieved, but calculated from T and q. Please be precise and correct the text wherever necessary. Further, what is considered "suc- cessful" here? The fact that a T and q profile comes out of the retrieval even if no RL temperature data is used, is not surprising. More interesting would be to see what accuracy (a posteriori error!) is achieved with or without RL temperature data.***

Deleted from the manuscript.