REPLY TO REVIEWER #1

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 #1:

The reviewer's comments are written in italic while our replies are in standard font. Within the manuscript all changes from the submitted version are highlighted in red.

1. ***"The manuscript does not represent a relevant or novel advancement of the field."***

We disagree because to our knowledge for the first time in ground-based remote sensing the following analysis has been carried out:

- a **robust** method for the **optimal** combination between lidar and microwave radiometer is developed and implemented
- the relative information content of both instruments and the sensitivity to their individual uncertainties is assessed

This will be of high importance for operators of such instruments at ground-based remote sensing sites and it is also supported by reviewer #2, who says that: "*The approach is novel in two ways: To my knowledge this is the first study where MWR and RL data are combined in an OE framework treating both as measurements. Second, the idea to use the lidar retrievals, i.e. absolute humidity profiles, as input with a simple forward model that does only unit conversions and interpolation is an elegant approach, that avoids the costly development of a forward operator of RL raw data. But, as shown in the study, it is still possible to compute the uncertainty budget of the retrieval accounting for measurement uncertainties in both MWR and RL data. […]The study fits well in AMT.* "
In addition, the short comments of Dr. Nico Cimini and Dr. Felix Ament support our opinion. These comments are: "*I think this manuscript well addresses the problem above, and provide useful information on the potential synergy of two instruments commonly used for thermo-dynamical profiling, that are microwave radiometer (MWR) and Raman lidar (RL). In particular, the manuscript points out important features such as…*" (N.Cimini), and "*I consider the manuscript as very useful to the science community as it quantifies the added value of this instrument combination and clearly depicts the main caveats*" (F. Ament).

2. ***"The degrees of freedom in Tab. 2 clearly show that the output is dominated by the lidar return."***

We clearly agree and this is what you should expect. However, in the manuscript we show that in the regions of non-complete overlap and when the SNR of the lidar gets too large, the microwave radiometer provides a valuable addition. Here we show, that both instruments in synergy provide a significantly more accurate solution than each instrument alone. In order to highlight the mutual benefit of the instruments, we now also demonstrate that the total water column can only be determined by lidar in combination with the MWR (See new Fig. 4 in manuscript).
The method has been presented and applied only to clear sky cases. This application is important for a first validation and understanding of the algorithm. While naturally in clear sky conditions the lidar shows a dominant performance, we strongly believe the method will be an even more powerful tool when applied to cloudy cases when the lidar signal is extinguished within the clouds. In this case, one can apply the algorithm to study and describe water vapor in and around clouds. This is exactly the work we are currently pursuing and will be part of a follow up paper.

3. „To be suitable for publication, I would expect the addition of a forward model of the lidar return"

An exact forward operator taking into account accurately all uncertainties of Raman lidar theory and measurement would certainly provide a more accurate approach. However, from our point of view adding such a complex forward operator (yet to be developed) is out of the scope of our study. We are quite certain that such an approach would not lead to significantly different bottom line assertions. We think the lidar product we consider together with its uncertainty specification (which theoretically includes the propagated errors originating from the raw lidar data) should suffice for a first, general retrieval assessment. Note, this is not the case for the microwave radiometer. For this instrument, the brightness temperatures describe poorly (compared to the lidar) the atmospheric profile. And we are forced to use complex, as labeled by the reviewer, radiative transfer models. While we do not go into detailed lidar forward modeling, we include a sensitivity study to investigate the effect on lidar uncertainty (Section 4.2.5).
At this point, we would also like to cite the comments of referee #2 and Dr. Felix Ament. The first mentions that: "*the idea to use the lidar retrievals, i.e. absolute humidity profiles, as input with a simple forward model that does only unit conversions and interpolation is an* ***elegant*** *approach, that avoids the costly development of a forward operator of RL raw data*". The second agrees with the argument: "*Further enhancements of the algorithm like e.g. a refined forward operator may provide some additional improvement, but I am convinced that they will not change the overall assessment of this paper. Moreover, the results of this paper will be very useful as benchmark for future algorithm development.*"

4. „lidar data is introduced using a qualitative cut-off is a poor application of the OE technique"

We are not completely sure what the reviewer is aiming at.

a) The clipping of lidar water vapor profiles above an altitude where the relative uncertainty is larger than 100%: Here please see Fig. R-1, that clearly demonstrates that the signal to noise ratio becomes so poor at altitudes above ~4 km during daytime in this case, that even negative mixing ratios occur that cannot be handled by OE.
b) The cut-off of 2.5 km, which we apply to all profiles: here we want to refer the reviewer to our reply to reviewer #2, point number 3, who also criticized this point. In fact, there is a strong difference between daytime and nighttime profiles and we have taken this into account in the revised manuscript.

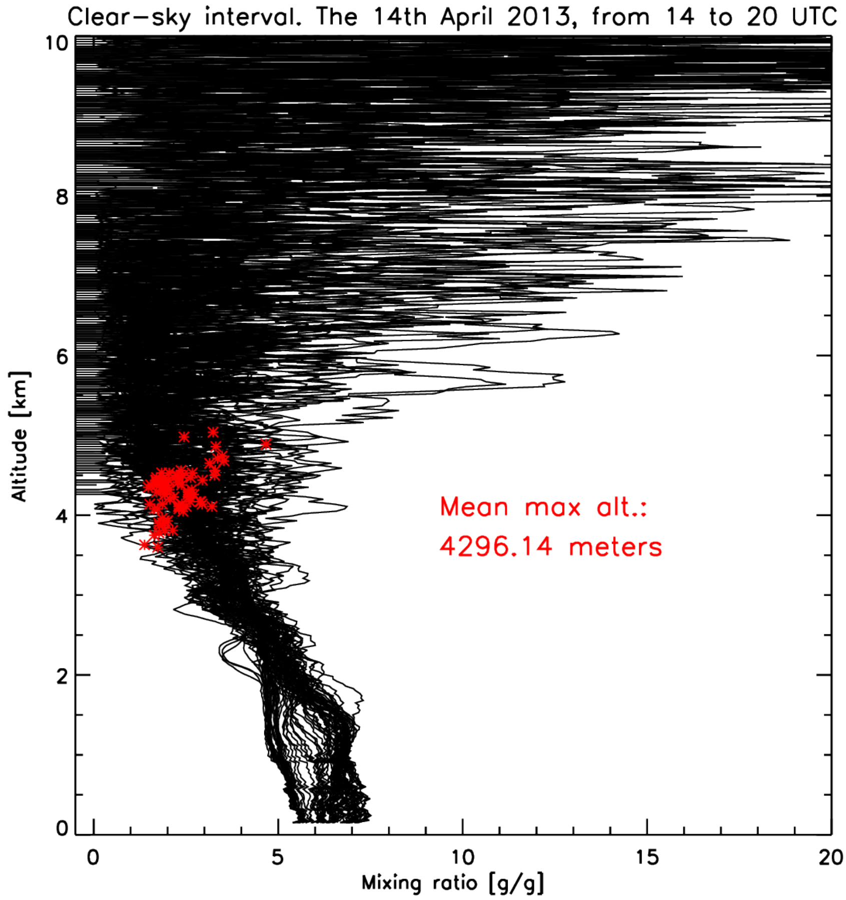

Figure R-1: Mixing ratio profiles in a clear sky interval, from 14 UTC to 20 UTC

For the interest of the reviewer we show the difference for an example profile, when we introduce the complete lidar profile in comparison to the cut profile. The following figure (Fig. R-2) analyzes this difference. We present the complete lidar profile, the retrieved profile from the cut RL data (at around 4 km) and the retrieved profile using the complete RL data. As expected, there is almost negligible difference in the lowest 4 km.

Above this altitude, the result is very sensitive to the lidar uncertainty definition. We are afraid that in this region of poor SNR the error could be underestimated. This can introduce artifacts on the retrieved profile and therefore these high altitudes are not taken into account.

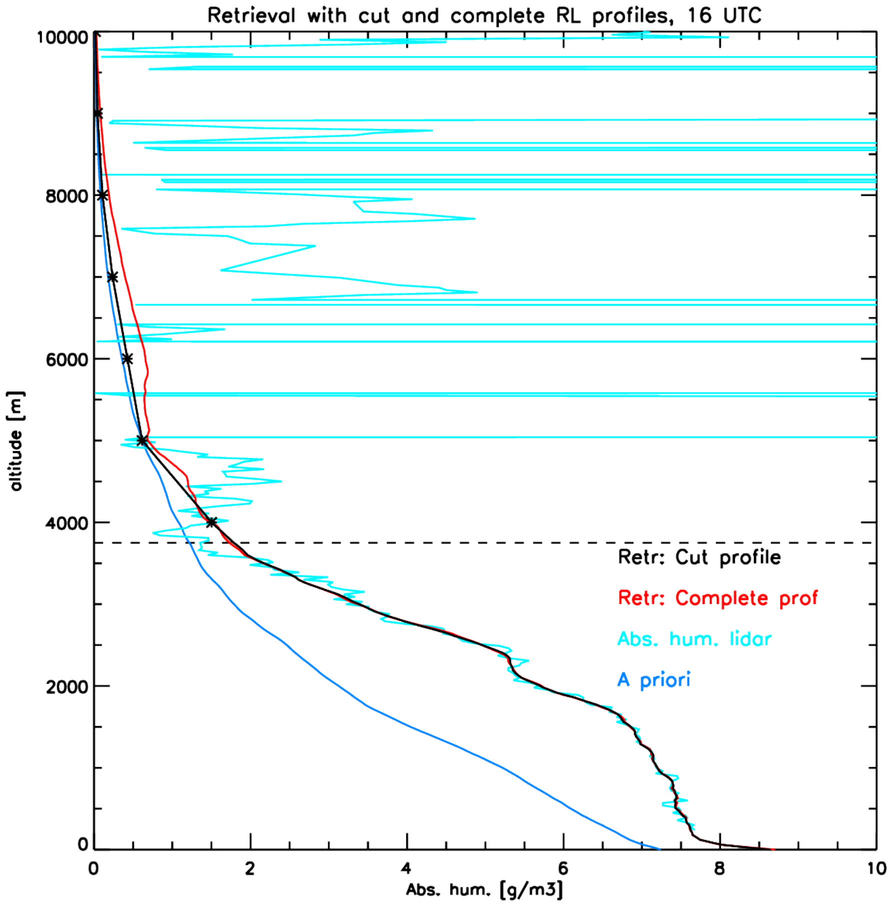

Figure R-2: Retrieved profiles when the complete lidar profile is used as input (in red) and when the cut RL data (at around 4 km) is used as input (in black). The initial RL water vapor profile (in cyan) and the a priori profile (blue) are also presented as references.

OTHER COMMENTS OF REVIEWER #1:

- ***Throughout: You frequently use 'error' when you mean 'uncertainty'.***

Accordingly corrected in the manuscript. The word uncertainty has been used when referring to instruments uncertainties. In contrast, the word error has been kept when referring to the theoretical error, that is, to the a posteriori error estimate calculated with equation 4 in the manuscript.

- ***5468, 24: These are excessively precise numbers considering they derive from the analysis of a single profile. One significant figure would more accurately represent your understanding of the change in error.***

The sections corresponding to the temperature and relative humidity have been omitted from the manuscript.

- ***Sec. 1: The referencing in this section is poor, especially for a student that is preparing a thesis, as they primarily draw from the previous work of the authors. Some reference to previous applications of optimal estimation to lidar and/or microwave radiometry should be included, in addition to the specifics of Raman lidar analysis. Some other papers from the HOPE project would also provide useful context.***

We introduced more literature on instrument combination, Raman lidar principle and retrieval in Section 1, and more details on Raman lidar are referred to Section 2. We also cite some papers on the HOPE campaign, which only have recently been accepted in the HOPE special issue of ACP. The changes have been highlighted in red.

- ***5469, 12: I thought we could resolve boundary layer turbulence with sodar and that the issue with cloud processes is the opacity of clouds, rather than the resolution of the observations?***

Unfortunately there is no instrument that can resolve boundary layer turbulence and the associated fluctuations of temperature and humidity under all conditions with sufficient resolution to test and further develop boundary layer parameterization within atmospheric models. In particular, sodar is limited by its relatively low maximum altitude range. During daytime in a developed convective boundary layer the typical maximum altitude ranges are 300-400m.

In respect to clouds, the reviewer is right that the strongest limitation of the lidar is its opacity, but for understanding cloud formation and development, humidity information close to the cloud boundaries is needed. In fact, in this paper we average over 5 minutes to adapt to the RL time-resolution and therefore eliminate most water vapor fluctuations due to turbulence (see Steinke et al. 2015). A future challenge will be to reduce the extension to finer resolution in order to provide modelers with the necessary information.

- ***5472, 19: Neither of these papers describe how the relative humidity is calculated. Considering the importance of the error covariance in OE retrievals (and your later conclusion that the uncertainty is underestimated), this technique should be briefly outlined alongside the procedure by which the uncertainty is estimated.***

The relative humidity (section 6) has been removed from the manuscript.

- ***5470, 3: From Fig. 9, I can see why you would blame the overlap function on failures in the lidar (as they are concentrated near the surface), but considering water- vapour measurements generally derive from a ratio of measurements, you will need to present more evidence for this (e.g. some manner of sensitivity study) to convince most readers.***

First, note that Figure 9 refers to temperature Raman lidar measurements, and has been eliminated. Second, there must have been some confusion between overlap and the blind region. The latter is associated with the slightly bistatic configuration of the lidar system, which is generated by the small, but not negligible physical distance between the lidar receiver and the transmitter systems. Due to this blind region, vertical profiles of water vapor mixing ratio typically start at 150 m and temperature profiles start at 300 m (see also in the paper at page 5472, line 27).

- ***5472, 9: Raman scattering is stimulated by all three wavelengths. We generally observe 355nm because the cross-section is greatest there, but the other wavelengths exist. For example Althausen et al. (2000) observes water vapour at 660 nm.***

We agree that the sentence was not clear enough. In the new version of the paper it has been modified as follows: "BASIL includes a Nd:YAG laser emitting pulses at its fundamental wavelength, its second and third harmonics: 1064, 532 and 355 nm, respectively, with a frequency of 20 Hz. The average power emitted at 355 nm is 10 W. For the purpose of these measurements Raman scattering is stimulated by the 355 nm wavelength, because of the higher cross-section with respect to the other wavelengths. However, water vapour measurements exploiting Raman scattering stimulated at other wavelengths have been reported by Althausen et al. (2000)".
We are well aware of the milestone paper by Althausen et al. (2000), which has been introduced as additional reference.

- ***5473, 7: For those unfamiliar with lidar, it may be useful to point out that the different ranges result from the relative strengths of the signals used.***

Included in the manuscript. The sentence has been reformulated as follows: "*Humidity profiles extend vertically up to different altitudes during daytime and night-time depending on the altitude where the signal gets completely extinguished. For water vapor this typically takes place around 5 km during daytime and around 12 km during the night. The different ranges result from the additional noise due to solar contamination during daytime*."

- ***Sec. 3.1: It is unusual to use the Gauss-Newton iteration rather than the Levenberg-Marquadt. The model you use is straightforward, but the additional computational expense of the L-M method is generally minor considering the improvement in convergence (G-N being more susceptible to local minima of the cost function). Regardless, the convergence criterion is poorly described in this section. You should be be consistent and use F (xi) rather than yi.***

To the knowledge of the authors the L-M method is useful and more trustful when the initial guess is far from the solution and when dealing with highly non-linear problems. None of the two conditions are true for the minimization problem we deal with. In fact, the lidar forward model is linear and the MWR is also close to linearity. The latter results from the weakness of the water vapor absorption line around 22 GHz. The authors preferred the G-N methods because is less computational expensive and converges quickly to the optimal solution. Many studies concerning MWR use the G-N (Cimini, 2010).

Note that in the manuscript, some comments around equation (3) have been included and y_i has been substituted by $F(x_i)$.

- ***Sec. 3.1: I initially thought you were using inefficient expressions, but later realised you are retrieving more states than data. It is advisable that the number of state vector elements n is less than the number of measurements m, as otherwise the solutions are underconstrained and can be sensitive to the initial conditions (or other retrieval parameters). I would be inclined to slightly reduce the resolution of the profile through the lidar region (or use a consistent grid throughout) such that n < m. The loss of resolution in the lidar region may be worth the increased reliability of solution.***

Underdetermined inversion problems are actually very common when dealing with passive remote sensing. The rank of the Jacobian is essential here. As reported also in the manuscript, the degrees of freedom for signal of a ground based K-band radiometer (typically using m=7 frequency channels) for the water vapor profile are maximum two, which poses only a weak constraint to the shape of the profile. If we use an a priori discretization of n>m vertical levels, the retrieval will adjust the a priori profile by only two pieces of independent information originating from the measurements; the retrieved product will be very smooth compared to the true profile. However, the resulting total uncertainty will not depend on the amount of vertical levels used, i.e. direct MWR IWV retrievals and the vertical integral of water vapor retrievals with n>m show the same (high) accuracy. This is due to the fact that the retrieval errors in the nxn a posteriori matrix are not independent. In the paper, we thus deliberately mostly write vertical discretization rather than resolution. As the actual vertical resolution of the retrieved profile is case dependent, we decided for a common vertical discretization for all of the retrieved profiles for sake of clarity and comparability.

- ***5475, 20: Influential though Rodgers (2000) is, you don't need to reference him three times in as many lines.***

Accordingly corrected in the manuscript.

- ***5475, 24: BT would be a vastly more common acronym for brightness temperature.***

The use of BT or TB is different in different communities. We have changed to the BT notation.

- ***Eq. (6): I thought that it was standard to place the diagonal elements of a matrix from top-left to bottom-right? You are trying to be consistent with Fig. 1, but you should check the style for this journal (and invert Fig. 1 if necessary).***

The suppression of sections 5 and 6, lead to a complete change in section "a priori".

- ***5481, 10: This paragraph, and many that follow discussing the lidar-only retrieval, are problematic. Since you have no forward model for the lidar data, the 'retrieval' without MWR data will be a weighted averaged of the profile and the a priori. That isn't a sensible product as the a priori uncertainty is large compared to the uncertainty on the lidar product. Either keep the current discussion but compare to the unprocessed lidar product, or concentrate more on the response of the retrieval where there is no lidar data (much as I would never trust retrieved values in those regions as there are zero degrees of freedom, the algorithm at least did something there).***

For consistency reasons it is very important to include a lidar retrieval including a priori information. In this way we can calculate a posteriori error and degrees of freedom for signal exactly in the same way as for MWR only or the MWR+RL combination. We need a comparable algorithm framework to compare the resulting numbers. Apart from this, in regions of non-complete overlap and when the SNR of the lidar gets too large, the a priori profile and its covariance do provide a certain constraint, which is better than no information at all. In this context please note that for many applications, continuous time series are needed. Therefore, we use the optimal estimation framework, which can provide uncertainty estimates, rather than filling gaps by simple interpolation. Please, also see our replies to your major comments number 2 and 3.

- ***5481, 27: Again, this isn't a meaningful statement. I would expect the error of the lidar data to closely follow the measurement covariance. It will be slightly reduced due to the a priori and the MWR data, but that isn't interesting.***

This is indeed what happens and what we expect to see; so this makes clear that the algorithm is working correctly and that MWR and a priori have no influence in this region: the RL measurements are the dominant factor in the region where RL is available. The uncertainties increase closer to the ground and above 2.5 km because the RL is not available. But the essential point is that the uncertainties outside of the lidar availability are smaller when using the MWR in the synergistic retrieval. To highlight this point, this last sentence has been included in the manuscript.

- ***5482, 3: I think statements such as this best demonstrate my problem with the algorithm's output. The results when using both inputs are only a 'best fit' by eye. There are substantially more data points in the lidar region than above, such that the improved fit in the free troposphere is mathematically irrelevant. In the lidar region, the algorithm basically doesn't change the input data (as the uncertainty on it is small relative to the microwave and the a priori). I can understand that, qualitatively, you were aiming to use the a priori covariance matrix to propagate the lidar information vertically, but I do not feel the algorithm you present is a sensible means of doing that (nor does the text communicate an emphasis on improvement above the lidar profile, if that is the case).***

We are not evaluating uncertainties "by eye". At this point in the text we have decided to not discuss the uncertainty quantitatively for the single profile because we discuss these uncertainties for the ensemble of retrieved case in Figs. 7 and 8 (new manuscript). We have now included in the manuscript in parentheses: "(please refer to Section 4.2 for detailed uncertainty statistics)".

The fact that the combined algorithm improves the humidity retrieval above and below the available lidar data region cannot be denied and is also visible (and discussed) in Figs. 6-8 (new manuscript).

Please note that propagating the lidar information vertically is only one aspect the retrieval successfully deals with. It's also the degree of freedom left from microwave signal that leads to an improvement.

It's also probably more important to ask whether or not the improved algorithm performance is physically relevant. In this case, 85% of the total water vapor is still above 2500 m. Meteorologically, we are here talking about the entrainment region of the boundary layer, where i.e. cloud formation and decay can depend on only slight changes in humidity. Thus, any improvement is this region should be of interest to weather forecasting applications.

- ***Fig. 4: Though this is a nice plot, how does the lidar product before your processing compare? Also, such plots are always more convincing when paired with a scatterplot.***

A comparison of the GPS derived integrated water vapor (IWV) with the original lidar data before processing in the OEM is not sensible, since the lidar does not give a complete profile. It lacks information in the lower part of atmosphere and also above a certain altitude, when the lidar SNR is too noisy. A more fair comparison is performed when working with a complete lidar profile, that is, after processing it in the OEM. The scatterplot for the three cases (combination, only-RL and only-MWR) against the GPS is presented in Fig. R-3, which has been included in the manuscript as new Fig. 4.

The bias and the standard deviation (in kg/m2) for all retrieval combinations to the GPS are presented, showing standard deviations that are in the order of uncertainty for the MWR and GPS (Steinke, 2014). The combination of the two instruments and the only-MWR case present similar standard deviations. Nevertheless the only-RL case presents a larger standard deviation (double the value). This numbers confirm that we do not need the RL to calculate the IWV, because the MWR is already providing the information on the integrated parameter.

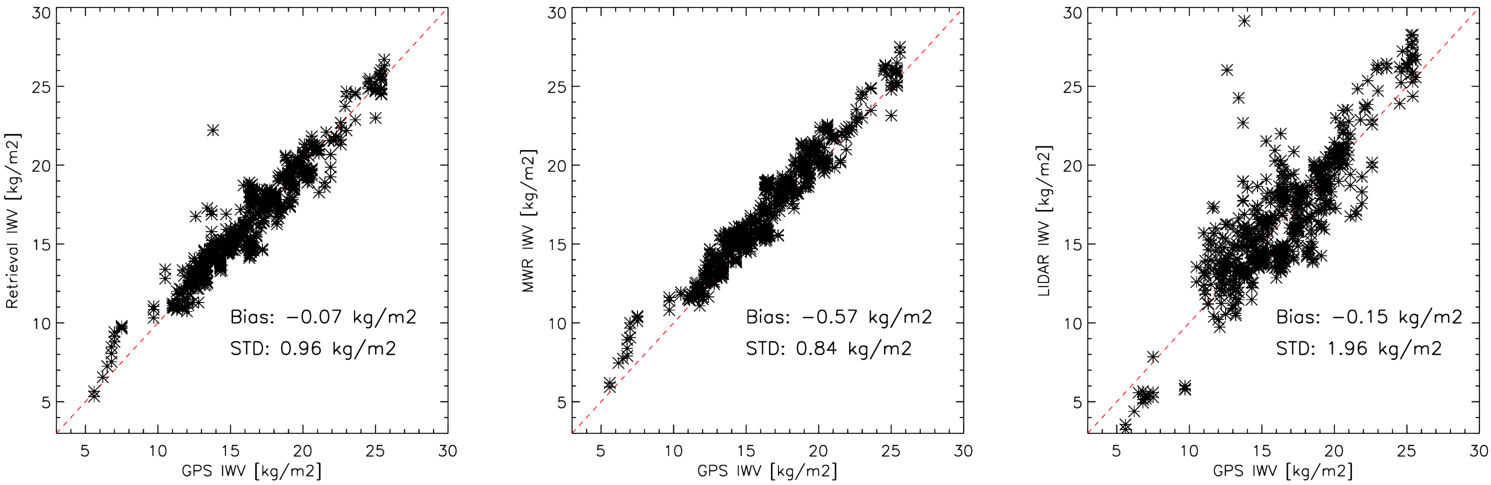

Figure R-3: Scatterplot for the three cases: the joint retrieval, only MWR and only Raman Lidar (from left to right), against the GPS.

- ***Sec. 4.2.2: To what extent are these radiosonde measurements independent? You imply they are used in the calibration of the lidar product.***

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. All radiosondes launched at times when BASIL was running were considered (approximately 60 comparisons). Considering only one single calibration value (any day at any altitude) throughout the

HOPE field campaign prevents and minimizes the dependence between a single lidar vertical profile and the simultaneous radiosonde profile used for the calibration.

- ***Fig. 7: You say the uncertainty is increased by a factor of four but the retrieved uncertainty on the product only increases by a factor of two. Though it doesn't need to be mentioned in the manuscript, it would be useful to confirm that you multiplied the covariance matrix by 16 (or multiplied the uncertainty by four before squaring).***

Yes, the initial lidar error is multiplied by a factor of 4 before introducing it in the retrieval routine. Then, when considering the variances in the covariance matrix, these are increased by a factor of 16.

- ***5487, 16: I have no problem with applying an empirical factor of four correction as a theoretical exercise, but there would need to be a more thorough error budgeting of the lidar product if you were to use this in practice.***

We agree. The sensitivity study where a factor of 4 was introduced is just intended for qualitative assessment. In this paper we only consider the random uncertainty and neglect other smaller uncertainty sources, like thermal effects on the optics or calibration constant uncertainty. No systematic errors can be handled by the OEM and need to be eliminated before lidar data is used in the retrieval. The sensitivity study where a factor of 4 was introduced is just intended for qualitative assessment.

- ***5488, 7: Considering you retrieve order 100 values in the lidar region, 12-25 degrees of freedom aren't 'large'. They're simply more numerous than those from the MWR. In fact, I'd take the relatively low number of degrees of freedom as a sign that you should retrieve on a coarser grid in the lidar region as each data point there is not conveying much independent information. The choice of resolution could be guided by the width of the averaging kernels.***

We retrieve 77 values in the lidar region but are only able to retrieve 15-25 independent pieces of information with the retrieval. This reduction is induced by the a priori covariance matrix: when one assumes that several altitudes are correlated, they are not independent any more and consequently, the number of degrees of freedom for signal (DOF) is reduced. The adjective "large" has been deleted from the manuscript, to avoid misunderstandings.
As a retrieval grid, we choose the resolution of the lidar for practical reasons. Note that, as already mentioned, the amount of DOF varies from profile to profile. But certainly future users of our data will be supplied with the averaging kernel to be able to carry out arbitrary comparisons with smoothed data (see new Appendix).

- ***Fig. 8: I know this plot was inserted in response to the Editor's comments, but it's difficult to interpret degrees of freedom without knowing the number of elements in both the measurement and state vectors (which are never mentioned).***

Introduced in the manuscript: caption of new Fig. 8: "*The number of elements in the measurement and state vectors are 79 (66 for the dashed case) and 91 respectively*".

- ***5490, 16: It improves, but I wouldn't say 'strongly'.***

The word strong has been suppressed.

- ***5491, 26: The two profiles shown are consistent with each other, such that I'm not certain the 'improvement' relative to the radiosonde isn't coincidental. You would need to demonstrate improvement over at least a few test cases.***

Yes, we do agree with your point. Nevertheless, we do not have more RL measurements in clear-sky periods when a radiosonde was launched. This does not allow us to study more cases. That is the reason why we have decided to remove this section from the manuscript.

LANGUAGE CORRECTIONS FROM REVIEWER #1:

1. ***5467: In the title, I believe 'high vertical resolution' would be better phrasing than 'high vertically resolved'. In fact, throughout the paper you use 'highly resolved' when I believe 'high resolution' is more appropriate.***
2. ***5470, 29: sensors compensates for***
3. ***5471, 21: of the clouds' lifetime***
4. ***5473, 1: due to an insufficient overlap***
5. ***5474, 27: of a horizontally***
6. ***5475, 14: An Optimal Estimation Method allows one to***
7. ***5481, 4: 'it allows one to work' or 'it can work'***
8. ***5482, 12: At ground level, the only two available***
9. ***5486, 19 As argued in Sect. 4.2.2***
10. ***5489, 10: the OEM allows one to work***
11. ***5490, 15: This sentence is repeated half-way through.***
12. ***5492, 25: The synergy presents its***
13. ***5493, 6: could be desirable.***
14. ***Fig. 1: (from top to bottom and left to right)***
15. ***Fig. 10: enclose the area where lidar***

Language corrections from 1 to 15 have been accordingly corrected in the manuscript.

We hope that the explanations given above could convince the reviewer that our manuscript is worth to publish.