Reply to Referee # 2

1. The statement of the problem looks strange: first, you use a priori in retrievals in the OEM method (while it is not necessary, the method by Hauchecorne and Chanin (1980) and your retrievals on coarse grid work without a priori), and then “remove” a priori information. While each step is understandable (note, they have different objectives: OEM tries to increase vertical range, while the removal of a priori detects the regions of useful information from measurements solely), their combination looks surprising. It took a while to understand that the OEM retrieval is performed only for determining the optimal coarse grid. The following should be addressed in the paper:

a. A clear motivation for each step and method with intended applications (climatology, trends, thermo-dynamical processes etc.) should be indicated. Limitations of each method should be indicated and discussed.

The Hauchecorne and Chanin (1980) (henceforth HC method) presented a robust method to calculate the Rayleigh temperature from lidar measurements using downward integration.
The OEM (on a fine grid using an *a priori* temperature profile) has several significant advantages over the HC method, which have been discussed in Sica and Haefele (2015) and Jalali et. al (2018). The advantages of the OEM overcome some of the drawbacks of the HC method and we will briefly summarize the advantages below.

1- The OEM directly provides a full uncertainty budget including the statistical and systematic uncertainties without any extra calculations afterwards. The HC method calculates the statistical uncertainty but it is incumbent upon the researcher to create an uncertainty budget through their own means, an example of which is the Monte Carlo technique (Leblanc et al. 2016) in order to calculate the systematic uncertainties.

2- The OEM calculates the vertical resolution from the averaging kernels at each altitude. The HC method has no such calculation, and it is left to the researcher to determine an appropriate vertical resolution. Leblanc et al. (2016a) have developed tools for calculating the vertical resolution for the HC method.

3- The maximum valid height can be chosen using the averaging kernel values mathematically and does not require the ad hoc removal of the top 10-15 km of the profile.

4- Many lidars systems require multiple detector channels to obtain the dynamic range of 10^6 or more necessary to measurements over a wide range of altitudes. The temperatures and their uncertainties should theoretically be identical in the overlapping regions; however, in reality, they are frequently different. In the HC method, in order to create a single temperature profile from different lidar channels' measurements, the profiles have to be merged. Merging induces another source of uncertainty which is difficult to quantify and has an unknown effect on the temperature profile. However, OEM retrieves one temperature profile from several lidar channels' measurements without the need for merging.

The OEM uses an *a priori* temperature or water vapour profile to constrain the solution space of the retrieval and the contribution of the *a priori* profile can be calculated using the averaging kernels at each altitude. The effect of an *a priori* profile (temperature or water vapour) on the retrieved profile may be of general concern for users, but is most important to consider for trend analyses or climatological calculations where all retrieved profiles should have relatively equal averaging kernel values at each altitude, and little or no *a priori* profile contributions.

Therefore, when conducting such an analysis, it is helpful to minimize the effect of the *a priori* temperature or water vapour profile from the OEM results, which is the purpose of the new technique presented in this manuscript. In order to do this calculation, the averaging kernels from OEM on the fine grid are used to calculate a coarser retrieval grid, in a way in which the contribution of the *a priori* profile is minimized. The price to pay for the removal of the *a priori* information is a slight deterioration of the vertical resolution, corresponding to a fraction of a degree of freedom, and larger uncertainties, which represent only the measurement information and no longer the total information available.

b. The results should be compared with the traditional method (Hauchecorne and Chanin, 1980), including H-C retrievals on your coarse grid (or/and averaging to a coarse grid). If the results are coinciding with the “removed a priori” retrievals, this should be indicated. Since lidar temperature profiles have been used in several gravity wave studies, please discuss the influence of a coarse grid on this application.

We can compare with the HC retrievals and interpolate them to the coarse grid. However, for the HC method will follow the “best practice” of beginning our integration at a signal-to-background ratio of 2, and the cutting off the top 10-15 km of every profile. Hence, it is not possible to compare directly above what was shown in the manuscript.

Figure 1 below, for 20120512 (the night in the paper) is one such case where the HC method is cut below the influence of the a priori temperature.

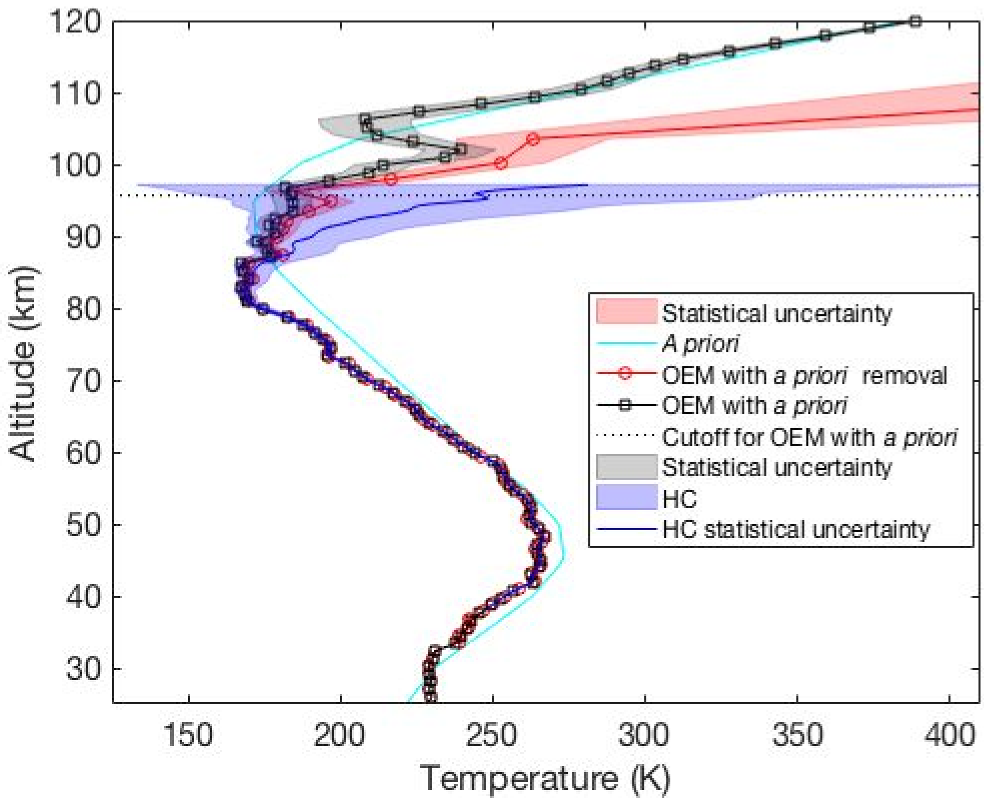

Figure 1: As mentioned above the top 10 km of HC temperatures (blue line) are removed due to the uncertainty caused by the seed pressure. In this case, it cannot be concluded that the HC result is closer to the fine or coarse grid result. In order to investigate this more, we used 9 nights randomly picked from PCL measurements and the percent difference between the fine and coarse grid retrieval with the HC method calculated.

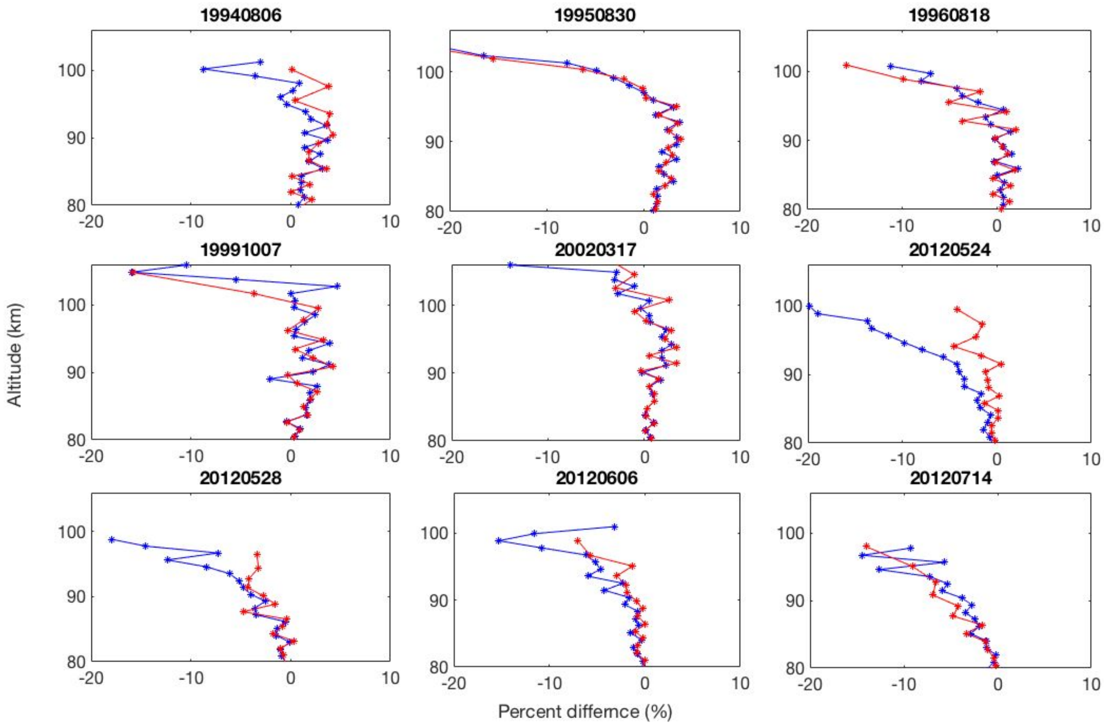

Figure 2: The percent difference between the fine grid retrieval with the HC method (blue line) and coarse grid (*a priori* removed) with the HC method (red line). Below 80 km the retrievals are identical, as the coarse and fine grid are identical.

Figure 2 shows that in general the method does just as well as the regular OEM, or better with respect to the HC method results. We may also conclude that, in general, the *a priori* temperatures do not have a large effect on retrieved profiles for most nights, however, for nights such as 20120524 and 20120528 the apriori seems to have had a larger effect which is removed by our technique.

The influence of gravity waves is beyond the scope of this paper but in general it could be mentioned that gravity waves can still be detected if the coarse grid is not coarser than the amplitude of the gravity waves. We would also refer you to some of the gravity wave studies of Sica and colleagues (e.g. Sica 1999, put in JAS reference), where we encourage the use of \Delta n / n as opposed to \Delta T / T for precisely this reason: temperature fluctuations will by definition have correlations built into them due to the integration required, whilst density fluctuations are uncorrelated (as the counters used for each altitude bin do not induce correlations into adjacent measurement bins).

c. Please consider changing/revising the title of the paper into e.g. “On optimization of retrieval grid for lidar measurements”, or “Information-based retrievals from lidar measurements” or “Retrievals from lidar measurements using information-centered vertical grid" or similar. From my point of view, this would reflect better the applied method.

We can change the title, but we would suggest instead: "A practical information-centered technique to remove a priori information from lidar OEM retrievals"

2. Using only one profile is far from being sufficient for any conclusion. At least, a representative dataset should be used. Variations in the optimal grid should be illustrated, and implications for applications should be discussed.

The purpose of this paper is to introduce a new method and we used two different state variables (temperature and water vapour) from two completely different lidars to show the results. We have tested many more nights, but as shown in our previous studies the OEM method, like the HC method, is extremely robust. More retrievals are shown in in Figures 2 - 6. *Figures 2, 4, and 6 will be added to the paper to discuss the additional days and nights which have been added to the study. The rest of the figures are shown here for clarification only.*

Temperature Result Expansion:
Figure 3 presents the cutoff height for the nightly temperature measurements shown in Figure 2 on the fine and coarse grid. In most cases the cutoff of both fine and coarse are close to each other. The cutoff for the coarse grid was chosen to be the last altitude at which the vertical resolution is smaller than 3 km. In some cases, this point is lower than the original 0.9 cutoff, such as on 20020317 and 20120714. It is up to the researcher to decide if the coarse grid heights are appropriate or not given the characteristics of their data set.

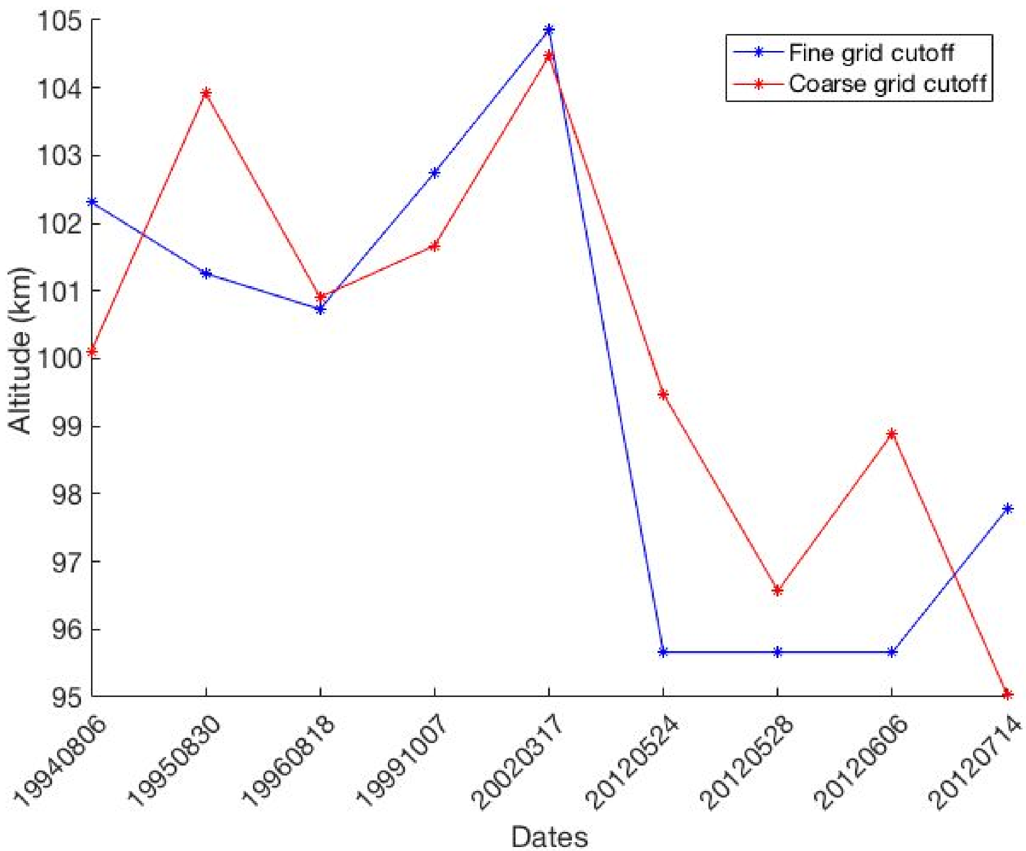

Figure 3: Cutoff heights for the 9 PCL sample nights on the fine (blue line) and coarse grid (red line).

Water Vapour Daytime

We have added 5 more days for the daytime analysis. The corresponding profiles are shown below in Figure 4 which will be added to the paper.

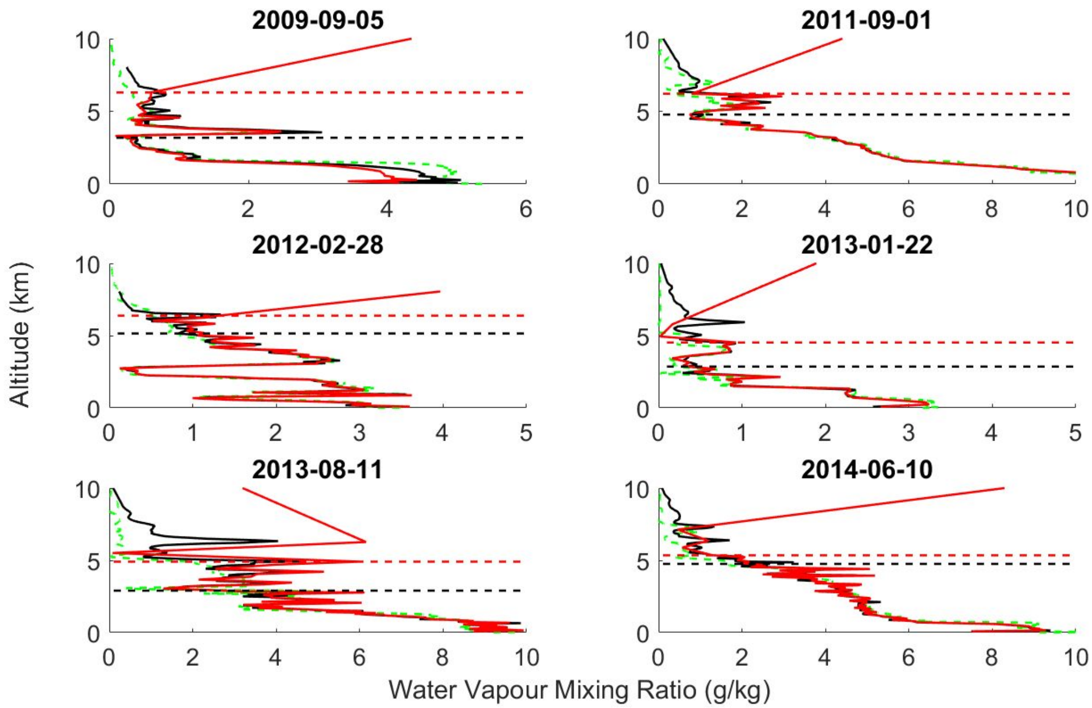

Figure 4:
Water vapour mixing ratio profiles for each daytime retrieval. The fine grid OEM retrievals are in black, coarse grid OEM retrievals are in red, and the radiosonde is the green dashed line. The black dashed line is the original 0.9 fine grid cutoff height and the red dashed line is the coarse grid cutoff height. The last point on the coarse grid is often skewed due to its high resolution.

The daytime water vapour OEM profiles typically reach up to around 3 - 5 km on the fine grid, and up to 6 km on the coarse grid (Figure 4). There is an average of 1 km difference between the 2 points. In some cases the differences are much larger, and this is usually due to the presence of dry layers causing the averaging kernel to decrease at a lower altitude. The large difference between the final altitudes on each grid is typically due to a slow decrease in averaging kernel values with height, as was shown originally. Additionally, in some cases, such as on 2012-02-28, the uncertainty never rose above 60%, in which case the second to last point on the coarse grid was chosen as the cutoff point.

The daytime water vapour OEM fine and coarse grid profiles produce similar results with respect to the radiosonde profiles (Figure 5). In order to compare fairly with the fine grid retrievals, the radiosondes were weighted using the averaging kernels of the fine grid retrievals and assuming that the radiosonde measurements each have unity degrees of freedom. The coarse grid comparisons do not require weighting as the averaging kernels each have 1 degree of freedom and are therefore directly comparable to the radiosonde. Prior to weighting, the radiosonde measurements were interpolated onto the fine and coarse grids. For each case, with the exception of 2009-09-05, there are very few differences between the fine and coarse grid retrievals from the radiosonde. On 2009-09-05, the coarse grid retrieval was shifted with respect to the fine grid OEM retrieval, possibly due to poor calibration on that day.

The daytime fine and coarse grid retrievals don't show very large differences with respect to the radiosonde, however, the coarse grid retrievals significantly increase the final meaningful retrieval altitude by an average of 1 km. Daytime water vapour retrievals are often limited in altitude due to the high solar background in both the water vapour and nitrogen channels. Increasing the final meaningful altitude by up to 2 km is highly valuable for forecasting and validation purposes.

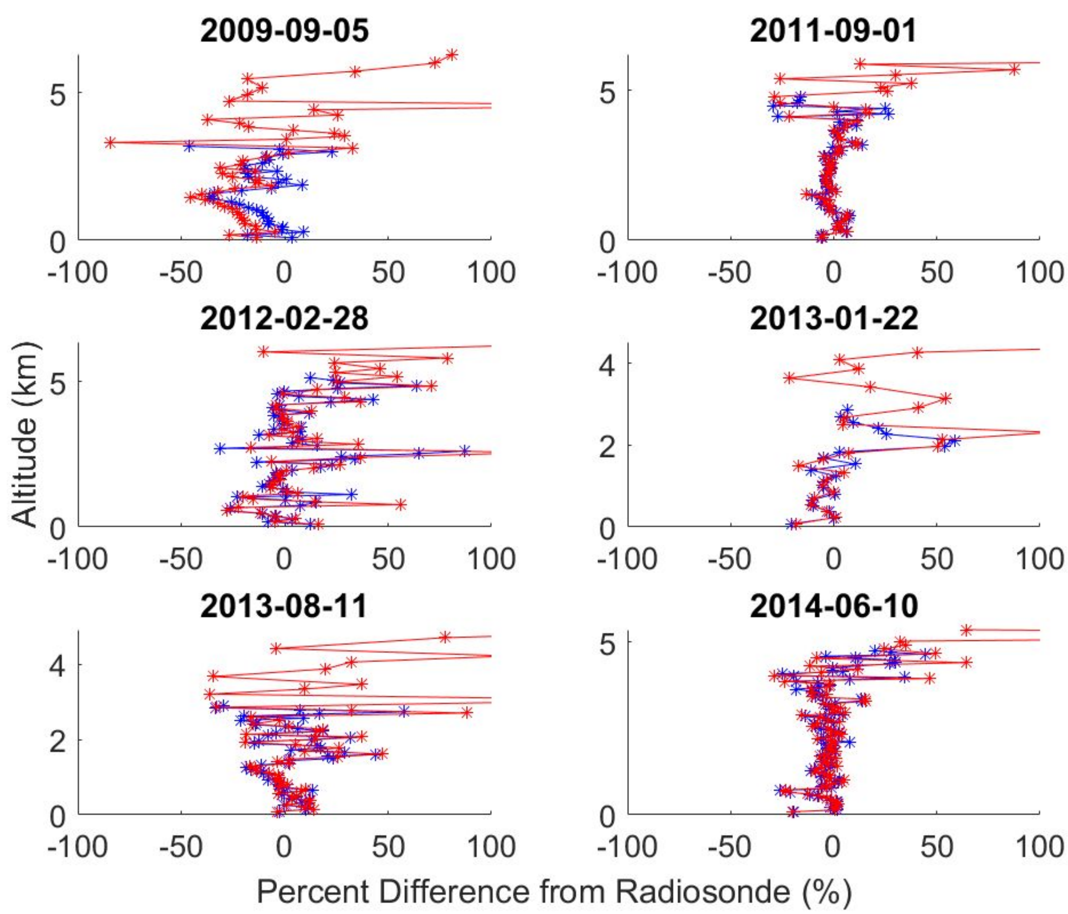

Figure 5: The fractional percent difference between the radiosonde and the lidar water vapour mixing ratio measurements for both the fine grid (blue) and coarse grid (red) OEM retrievals. The radiosonde measurements for the fine grid comparison were weighted using the fine grid OEM averaging kernels.

Water Vapour Nighttime

We have also added 8 more dates for the water vapour nighttime measurements and will be adding Figure 6 to the paper.

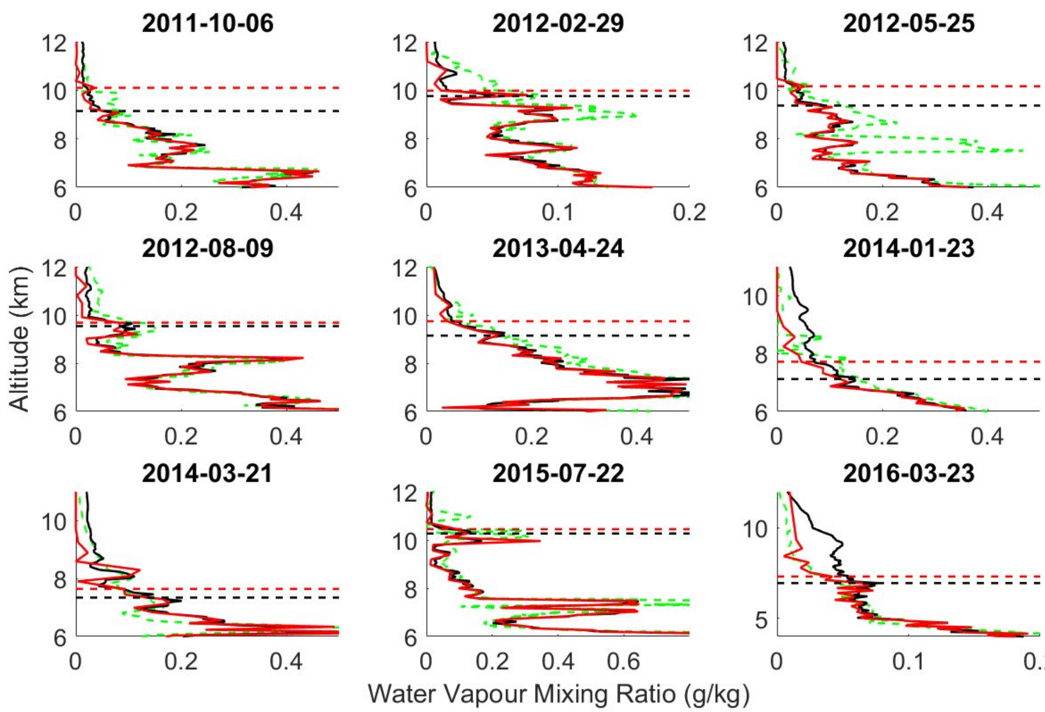

Figure 6:

All nighttime water vapour retrievals. The radiosonde is shown by the green dashes, the fine grid retrieval in black, and the coarse grid retrieval in red. The 0.9 cutoff height for the fine grid is shown by the black dashed line while the coarse grid cutoff height is the horizontal red dashed line.

In all cases, the water vapour nighttime coarse and fine grid retrievals produced results with similar percent differences with respect to the radiosonde. Fractional percent differences with respect to the radiosonde were calculated by interpolating the radiosonde onto both the fine and coarse grids (Figure 6). Similarly to the daytime procedure, the radiosonde measurements were weighted using the fine grid averaging kernels. No weighting was necessary for the coarse grid comparisons. The differences with respect to the radiosonde are similar for both the coarse and fine grid retrievals.

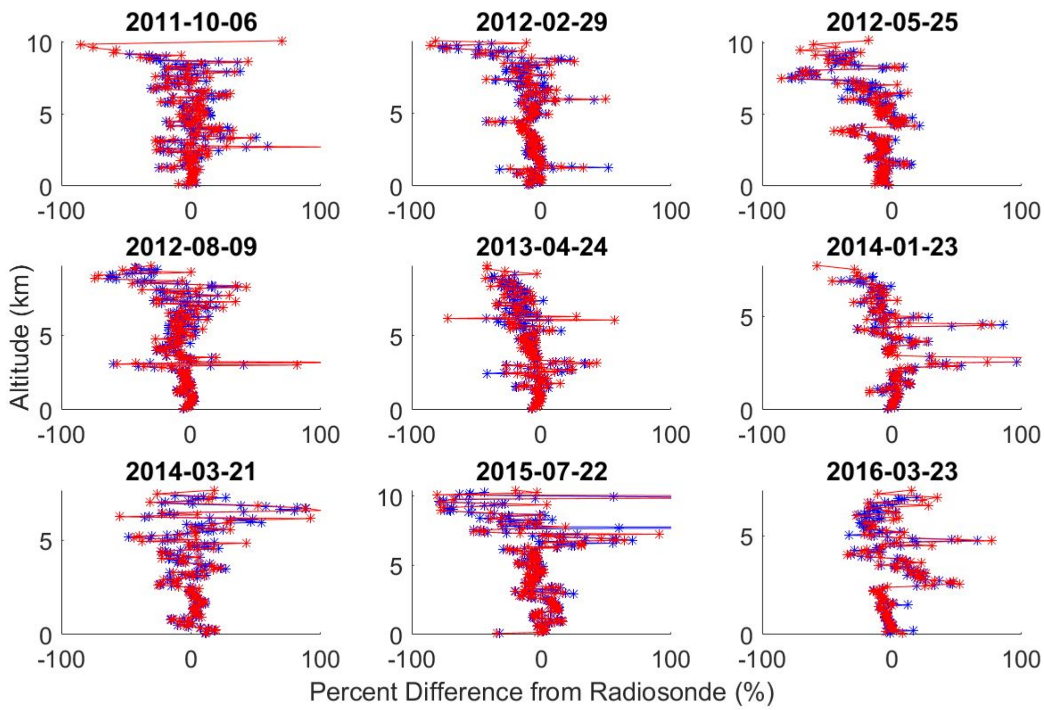

Figure 7: The fractional percent differences with respect to the radiosonde measurements. The differences with respect to the fine grid retrievals are in blue, and the differences with respect to the coarse grid retrievals are in red. The radiosonde measurements have been interpolated to both the fine and coarse grid retrievals in each case. However, the radiosonde has also been weighted with respect to the fine grid retrieval using the fine grid averaging kernels.

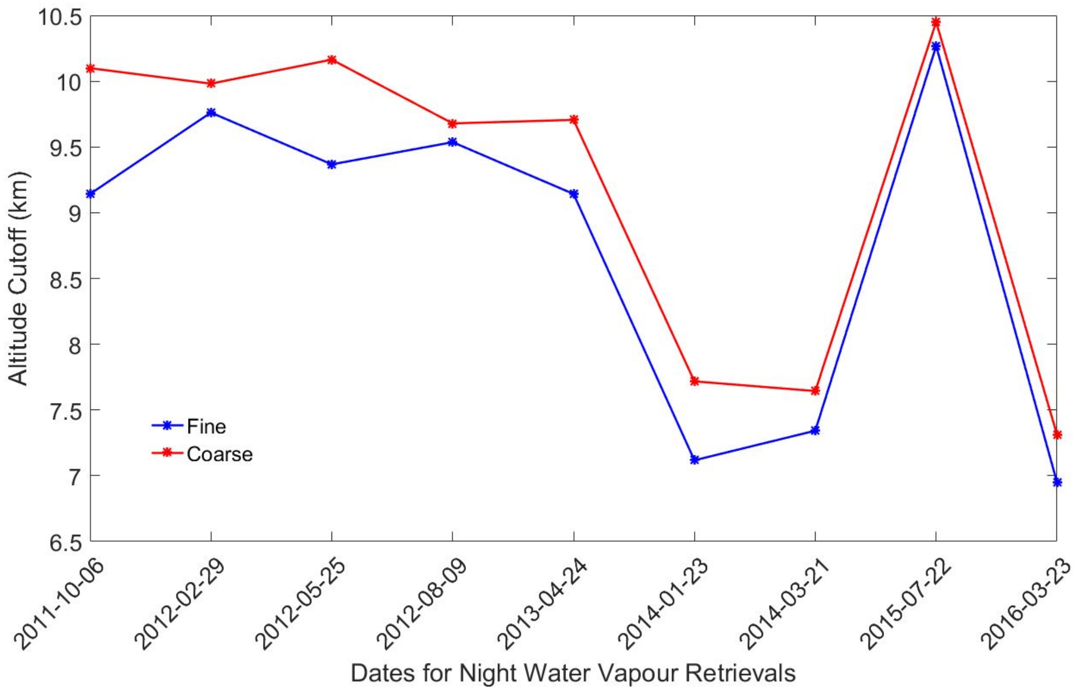

Figure 8: The cutoff altitudes for both the fine and coarse grid retrievals for each nighttime water vapour retrieval. The fine grid cutoff height is the last altitude at which the water vapour averaging kernels equal 0.9. The coarse grid cutoff height is the last height which has a total uncertainty less than 60%.

The nighttime cutoff heights shown in Figure 8 show a general increase in cutoff height when using the *a priori* removal method. The coarse grid cutoffs were chosen to be the last altitude below with a total uncertainty less than 60%. Choosing a maximum uncertainty of 40% would result in cutoff heights closer to the original fine grid's. In all cases the coarse grid increases the maximum acceptable altitude, however, in some cases by only a few hundred meters. On those nights, the averaging kernels decrease quickly after the original fine grid cutoff height, therefore there is very little information with which to create the coarse grid.

DETAILED COMMENTS

1) If the major comment 1c will be accepted, the whole paper needs restructuring and rephrasing

As discussed above, we are not comfortable with changing the paper title as suggested.

2) P.2, lines 14- 32 and illustration: the direct discussion of the figure in another paper (Jalali et al., 2018) and its further analysis complicates reading and is not needed in the introduction. It is sufficient simply mention that the influence of climatology at upper altitudes is large and indicate typical values.
We respectfully disagree with this statement. Figure 1 is provided to show the effect of *a priori* temperature profile on a large data set, such as the one that used to calculate the PCL temperature climatology in Jalali et. al. 2018. These lines are necessary as explanation for Figure 1. Figure 1 is also necessary to show the contribution of the apriori to a lidar data set, which has yet not been discussed in detail and is the motivation behind this paper.

3) P.3 L9 : "some of the method's advantages" Which method: yours or von Clarmann&Grabowski?

We are referring to our method in this case. To make it more clear the sentence will be changed to: "....some of the proposed method's advantages".

4) Section 3, "OEM theoretical background", should be radically shortened. It is not needed to rewrite many formulae from Rodgers book, please keep only essential ones.

This OEM has been newly applied to lidar data and is new for the lidar community. As such, we believe that adding in this background section is necessary and makes the paper more accessible. In section 3 only necessary equations that have been used in the paper are presented like equations 4, 5,6 and 7 and for the OEM introduction equations 1,2 and 3. It would be more difficult for readers to follow the paper without presenting these equations and we would like to avoid having readers need to refer to another text as much as we can. We believe keeping the equations present makes the paper more accessible for the reader and provides the necessary amount of introduction to the discussion.

5) X_a term is missing RHS of Eq.(6)
Thank you for catching this mistake. The term X_a has been added to the RHS of equation 6.

6) P.5 L 20 ". . . a full RANDOM uncertainty budget"
The term full uncertainty budget includes systematic and statistical (random) uncertainties. Therefore it is not necessary to add term "random" here. We will make sure to clarify this in the paper so the reader understands that both types of uncertainty are included.

7) P.6 L. 1 “ if the A is unity at each altitude” (and similar statements below in the paper). Be careful with phrasing: matrix cannot be unity.

We will clarify the language in the paper such that when we are talking about the entire matrix we will use “the identity matrix” or “unity matrix”.

8) Figure 2: It would be more clear if cumulative trace of A (y-axis) would be shown as fine grid levels (x-axis), and coarse grid points would be also indicated on this plot.

We could not think of a way to effectively show both the fine grid points with the cumulative trace in addition to the coarse grid points, which is why we did not originally do as you suggest. Our worry is that if we try to superimpose them, the reader will be confused and since the relationship of the coarse grid points to the fine grid points is the main point of interest, we believe it is better to leave it as is.

9) Figure 3 caption: it is sufficient to write only that the measurement response is indicated by red lines, since the definitions and explanations are given in the text.

We will remove the definitions from the caption and refer the reader to the explanation in the text.

10) Section 5.1 and 5.2 results: in retrievals, there is usually a trade-off between vertical resolution and random uncertainty. In your retrievals, both vertical resolution and retrieval uncertainty increase. Then an advantage of the method becomes doubtful. Please explain also reasoning behind using a climatology as a priori in RALMO retrievals. Evidently, that climatology cannot be a good a priori. Instead, ECMWF forecast/analysis would provide a more reliable a priori.

Increasing the uncertainty and vertical resolution is the trade-off of using this method. This doesn’t make the results doubtful. You should consider that the error bars of OEM and of ML are not comparable. The OEM uncertainty is combination of the a priori uncertainty and the ML uncertainties, and is smaller due to the extra information given by the *a priori*. Therefore, with removing the a priori from OEM and going to the ML causes estimate error increases. Mathematically, this can be shown using: 1/sigma_oem**2 = 1/(1/sigma_apriori**2 + 1/sigma_ML**2).

The goal of this method is remove the *a priori* influence from the final retrieval and it is successful. It also successfully provides an information-centered final altitude grid which can be used for trend analysis. However, researchers should be aware, that the cost of removing the *a priori* from the final retrieval is a lower resolution result with higher uncertainty. This lower resolution is negligible to non-existent when the averaging kernel is close to 1 and/or constant, but does increase when the the averaging kernel begins to decrease at higher altitudes. Researchers should decide whether or not removing the *a priori* is necessary for their work or not.

Regarding using climatological *a priori* profiles, the same question was asked by the other referee therefore we have copied our response here:

"We agree that the CIRA and US Standard model are not the most accurate models. However, it is necessary to use a consistent *a priori* throughout a climatology or trend analysis study to avoid inducing trends or bias into the results. To make it clear to the reader that this is a matter of choice, we can add a sentence to the conclusion stating that if one decides to use an *a priori* closer to the retrieval the effect may be smaller.

The lidar averaging kernels are so sharply peaked (that is they are at the resolution of the retrieval grid) for most of the profile, that in fact an *isothermal a priori* temperature profile could be used. Hence, for Rayleigh-scatter temperature using ECMWF would have little benefit (in addition to the fact it does not use measurements above 80 km altitude). For water vapour, the retrievals were designed with operational use in mind which requires a minimal number of dependencies in the code as possible, and preferably no need for internet. Additionally, it requires consistency over an entire data set. Therefore, we chose the US standard model over other reanalysis models like ECMWF which would require constant updates."

We have included a paragraph in the discussion section of the paper discussing our choice of *a priori*. So that the reader is clear that this was a matter of choice and that others may choose different priors depending on the goal of the research.

11) Illustrations based on one example in Section 5.1, 5.2 and 5.3 are not sufficient for any conclusion. The analysis of a representative dataset should be performed.

We have discussed this in our responses to Comments 1a-c. We will remove the discussion regarding dry layers as the new results show that using the fine grid averaging kernels to weight the radiosonde provides similar results to the coarse grid. Therefore, if a dry layer occurs at lower altitudes where there is little a priori influence, it may be more appropriate to weight the radiosondes and keep the fine grid on a daily basis than to use a coarse grid.

12) Summary should be updated after analysis of a representative dataset.

Agreed. We will update the paper's summary to reflect the changes we have made to your comments.

13) The text like in first two paragraphs of "Conclusions" should be avoided: the papers in AMT are for educated readers. The statements is the 3rd paragraph (p.21, lines 13-16) should be confirmed by the analysis.

Apologies, but we are not sure what is wrong with the first two paragraphs of the conclusions. We recognize that AMT papers are for educated readers, but the first two paragraphs clearly restate the motivation of the paper and the goals we set out to accomplish. We have changed a few sentences and hope that the paragraph below is suitable.

The statements in the 3rd paragraph are meant to show how the technique may be applied to a larger data set and we would argue that this is beyond the scope of this paper. In fact, this is ongoing work will be presented at a later date with regards to the water vapour profiles. However, we do not believe such an analysis would be appropriate for a methodology paper.

New Conclusion Paragraphs:
"When designing an OEM retrieval, it is often desirable to understand the effect of the chosen *a priori* parameters or profiles. This effect has been explored in detail for satellite-based and passive ground-based instruments, but not for the new area of applying OEM to active-sensing measurements such as lidar. Lidars are high resolution instruments with significant amounts of information available from their measurements as evidenced by the retrieval averaging kernels. The OEM helps to illustrate the robustness of the lidar data products with the advantage of providing diagnostic tools, such as the averaging kernel and a full uncertainty budget.

The *a priori* removal technique may be helpful for checking the *a priori's* influence on the
retrieval and in determining the appropriate *a priori*. It is most useful for lidar measurements with low signal to noise and a slow transition from regions of high signal to low signal as it can provide an additional 1 km of measurements. The method is less effective when signal strength changes rapidly, such as when the nighttime water vapour measurements quickly enter the dry upper troposphere or lower stratosphere."