Response to Anonymous Reviewer #1

Response to Reviewer's general comments

Raman water vapor lidar is being considered as a potential source of climate data records of water vapor within the Network for the Detection of Atmospheric Composition Change (NDACC) and is one of the main tier 2 instruments in the GCOS Reference Upper Air Network (GRUAN). This indicates that there is considerable international interest in the use of Raman lidar for climate quality measurements of water vapor. Any concerns about such measurements being suitable for trend detection purposes have to be carefully documented and discussed. In particular, any biases in the data need to be studied carefully and should be corrected. This follows from the advice of the Joint Committee for Guides in Metrology as expressed in their Guide to the Expression of Uncertainty in Measurements. So, this paper is unusual in that it needs to both develop and demonstrate the corrections that are the main result as well as to cover the material that is normally described in a calibration/validation paper. These are reasons why the paper has become quite long.

Furthermore, neither this reviewer nor the second reviewer appears aware of the situation within NDACC relating to the use of corrections to biased data. The concept of corrections for Raman water vapor lidar measurements has not been well received by a small number of individuals within the NDACC lidar community. There have been significant efforts to argue against the validity of the approach presented here and to even induce co-authors to remove their names from this manuscript. Co-authors did remove their names frm this paper due to these political efforts. These are additional reasons for the "narrative" style of the paper, the detailed discussion of the topics and the significant effort put into justifying the correction technique that is presented. But we agree with the reviewers that "This argument is sound and deserves publication" (reviewer 1) and "certainly deserves to be published" (reviewer 2). In light of the heightened politics surrounding the concept of corrections of Raman lidar data within NDACC, however, we feel the need to provide a thorough justification of the technique. Therefore, the remaining authors have attempted to provide careful discussions of:

1. The existence of various biases, in numerous lidar systems, that can contaminate UTLS Raman lidar measurements of water vapor.
2. The philosophical justification for correcting these biases by reference to and quotation from the Guide to the Expression of Uncertainty in Measurement.
3. The detailed equations pertaining to the signal-dependent correction including all assumptions in the equations.
4. Comparisons demonstrating the utility of all the corrections used in the analysis of the ALVICE MOHAVE-2009 data including those of overlap correction, temperature dependence correction, etc.

These are the reasons that the paper has become rather long. But, again, given the heightened politics and the efforts to discredit this approach to data analysis we believe that a very detailed treatment is needed.

That is not to say that the current manuscript cannot be tightened up and the reviewer's suggestions have been very helpful in this respect. In order to make the main point of the paper more succinctly, we suggest moving all of section 6 from the main body of the paper into an additional appendix. We will also work to shorten other text as we work through the revision.

We give responses below to the reviewer's detailed comments.

Response to detailed comments. Reviewer comments in italics

p. 7344 lines 2-5: 'The results discussed there indicate that the estimated total RH uncertainty for corrected RS92 measurements during the MOHAVE-2009 campaign were ±(5%+0.5% RH) for RH>10% and ±(7%+0.5% RH) for RH10 %, which corresponds to an uncertainty of ±6% at 50% RH, ±10% at 10% RH, and ±24% at 3% RH.' This sentence is very confusing because some of the numbers seem to be percentage RH and some seem to be percentage error in the percentage RH. Even so, they do not make sense. At 10% RH, the error is quoted as 7% + 0.5%. Is this not the same as 7.5%? And how does it then equal 10%? And are these RH or error percentages? This section needs a complete redrafting.

The numbers are correct but we will expand the text for clarity.

p.7344 l. 15 ALVICE

We will make this correction.

p. 7345 l. 12 have a strong

We will make this correction.

p. 7350 signal-induced noise in the photomultipliers can also produce the described effect, as is acknowledged on p. 7355; it would be helpful to refer to this briefly at the beginning of the section to emphasise that the current method doesn't apply to it. Section 5 is much too detailed: the corrections proposed are quite straightforward and do not merit over six pages of text. In the end the authors choose a very simple correction because the signal-to-noise limits application of the 'correct' solution. It would be sufficient to describe this correction and give a measure of its accuracy.

We will include a statement at this point in the manuscript indicating that signal-induced noise effects cannot be corrected with the approach outlined here.

This correction technique may be "straightforward" but it has never been published before and we believe that a detailed discussion is appropriate under such circumstances. We also believe that the detail given is necessary to justify the use of these equations to reluctant members of the NDACC lidar community.

p. 7358 l. 20. the final calibration value 20 used for the processing of the ALVICE Raman lidar data was determined by averaging the calibration constants determined from corrected RS92 and frostpoint hygrometer (FP). The two calibration constants differed by approximately 5 %. This averaging was done to compensate for the dry bias of the corrected RS92 data compared with frostpoint hygrometer shown in Fig. 12. I don't understand this. The FPH is generally regarded as the most accurate instrument for vertical profiling so why are you averaging the two calibration constants? Is this a trade-off of accuracy and precision? A proper explanation of this method is required.

The CFH has been shown to have an occasional moist bias for uncertain reasons in the lower troposphere (Miloshevich, 2009). The detailed study performed in Miloshevich et al., 2009 also indicates that the RS92 calibration accuracy is on the order of 4% in the lower troposphere although the results during MOHAVE-2009 may indicate a small bias in the corrected lower tropospheric data. Given that the two calibration values were within each other's respective uncertainty we chose to simply average the two results. We will introduce some additional text explaining this.

p. 7631. What is the point of including the format of the data files in a scientific publication? This belongs in a technical report not a published paper which should have some general relevance.

This material needs to be better motivated and we will do that and also move it to an appendix. But the reason that the elements of the datafile are important is that, to our knowledge, this is the first published attempt to quantify the full uncertainty budget of Raman water vapor lidar measurements. Such quantification is crucial to establishing climate data records from any measurement system and we are attempting to do that for Raman lidar here. There is an additional error source that for our case is insignificant but which we will mention in our revisions. That is the error in the photon pileup correction.

p. 7362 l. 9. 6 km not 7

In fact the height is slightly greater than 7 km, so we prefer to leave the text unchanged.

p.7363 l.23. Here a large discrepancy between lidar and radiosonde is attributed to smoothing of the lidar in a region of large gradient. Does this mean you are comparing profiles with different height resolution? This is an elementary error, easily rectified by smoothing the sonde profile in the same way as the lidar, which surely a group of authors of this experience will have done? Please clarify.

These are normalized difference plots performed at 1 km resolution. So the data are handled in 1 km layers for both instruments. Upon reflection, however, the explanation for the differences shown is more likely to be atmospheric variability that, for a few cases, causes the volumes measured by lidar and radiosonde to differ in their water vapor content. We will investigate this further and revise the explanation appropriately.

p.7363 l. 27. The reminder of this section could be omitted. Water vapour is known to be variable everywhere (not just over a mountain) and a comparison of a radiosonde or FP with an all-night average profile could not be expected to be as good as with a 1-hr average. Similarly, the differences here are much too large to detect any but the grossest variations in 1.

We find that the reminder helps to connect the results of figures 5 and 6. We have also found that water vapor variability can in fact be more marked in a mountain top environment than elsewhere. So we would prefer to leave these statements as they stand.

Section 7.1 If this has already been published it should not be included here: reference should be made in the introduction to the previous study. It is not simple comparing a total column measurement to an integrated lidar profile because so much of the water column is in the lowest layers of the atmosphere where the lidar does not measure. The paper does make this point and describes a correction procedure, but that procedure is not error-free and this section neither supports nor refutes the argument in the paper on correcting lidar water vapour profiles. If the authors want to include this comparison for completeness the section should be much shorter, refer primarily to the previous paper, and address honestly the uncertainty in determining an 'overlap correction' – this is not going to be the same each time a measurement is made.

These results are not shown elsewhere and so cannot be referenced. The point of including these comparisons is that GPS has been used as a calibration source for Raman water vapor lidar before and is one of the redundant sources of calibration for the Raman lidar within NDACC, GRUAN and for other applications. So it is useful to consider how the GPS may have served as a calibration source during MOHAVE. We will amplify on these points in the revised version to avoid the confusion expressed by the reviewer. We will also address the stability of the overlap correction, which is very high due to the use of an automated alignment system that holds the laser alignment to within ~10 – 20 microradians.

Section 8. What is the point of this section? It needs to be drastically shortened, much of the narrative condensed into a Table and draw a conclusion relevant to the narrative of the paper. All I have learnt from reading it is that there are a lot of malfunctioning Raman systems, some with wet biases and some without. Unless the results from the other lidars have been published and proper reference can be made to them I don't see any case for a section starting 'The data acquired by all three lidar systems during MOHAVE-2009 have undergone various versions of processing.'

One of the main ideas of the paper is that reference to lower stratospheric water vapor amounts can be used as a routine method for quality controlling upper level Raman water vapor lidar profiles. Such a technique could be useful at all stages during the various processings that go on during a field campaign. The fact that there were various biases found during the different processing stages implies that the techniques proposed here could have been useful at all steps along the way. We will make these points in the revised version.

Section 9 is an interesting idea which allows the results of MOHAVE to be generalised to periods when FP measurements are not available. While it certainly should be included, this section again is far too long, and labours what is an obvious (and potentially elegant) argument. Of course, if MLS and FP agree during MOHAVE the lidar profiles corrected with each should be similar.

The argument may be obvious and potentially elegant to this reviewer but, as we have said, this is definitely not the case with all members of the NDACC lidar community. It is for this reason that we believe that this very important section of the paper must be dealt with in the detail that is given. We prefer to leave this section unchanged.

The Appendix has four pages and four figures, and its material is tangential to the argument in the paper. This material belongs in a paper on the accuracy of radiosondes, which would make it easier to find and reference in future studies. A summary with one take-home figure is appropriate here, but no more.

We explored fleshing out the radiosonde accuracy study to become a separate paper on its own. However, we did not have sufficient support for such work. The results are important because they provide an accuracy assessment of current (as of 2009) Vaisala RS-92s corrected using the Miloshevich corrections where the results indicate a possible small bias in the corrected data. The results also demonstrate the utility of a surface reference station which is now an integral part of the ALVICE mobile laboratory for its field deployments. We would prefer to leave this as an appendix so that the calibration assessments performed as part of our MOHAVE work will be citable.

Fig.5 caption 'estimate'

We will make this correction.