Response to Anonymous Reviewer #2

Response to general comments (reviewer comments in italics)

The present paper is dedicated to the discussion of correction techniques to be considered in order to address and remove wet biases in the Raman water vapour lidar signal, primarily associated with fluorescence from different potential sources. The occurrence of fluorescence is an undesired technical problem, as in fact in the case of the authors this was associated with the unfortunate event of insect material that was deposited on the telescope early in the mission, while in the case of other systems mentioned here this was associated with the implementation of technical solutions in the receiver based on the use of fluorescent optical materials. In principle if the source of a systematic error is identified and can be removed, this should be the way to proceed instead of considering an algorithm to remove the effect of the systematic source. Especially because a correction for a systematic effect always relies on an estimate of the systematic error, which is itself affected by an uncertainty. I would imagine that after Mohave-2009, besides ALVICE for which this is clearly specified, all other lidars involved in this field deployment went through some system layout modification in order to remove this effect for future deployments. Additionally, the correction for the systematic error associated with fluorescence or other sources of wet bias (for example, interference filter bleed-through for the elastic signals) are variable with time (fluorescence can vary with time because of the long-term variable fluorescence behaviour of optical materials in the receiver; similarly interference filter bleed-through can improve with time because of the progressive degradation of the filters) and this produces a continuous dependence of the lidar measurements from the correcting sensors' data which makes the lidar measurements no longer independent (unless the systematic error is considered constant throughout the lifetime of the lidar system, which is very unlikely for most of these wet-bias sources). These considerations, to my opinion, will limit the potential future application of this correction technique, which is however well formalized and deserves to be published.

The reviewer makes the important point that biases due to fluorescence can be expected to change with time and thus can be time dependent. Biases can be due to degradation of optical materials which change with time. Or, more pertinent to our experience during MOHAVE-2009, they can be due to deposition of organic material on lidar receivers. Pollen and other organic materials such as bug wings are common sources of fluorescing agents and depositions of such material should be of concern as a potential source of time dependent fluorescence in Raman lidar receivers. This is particularly important for systems that operate open to the sky, as is the case for most NDACC lidar systems. This is the reason why we recommend an on-going quality control of the lidar data as a routine part of data processing for a Raman lidar system that is expected to provide climate quality measurements of water vapor. As we state in the paper, building or modifying a Raman lidar system to be free of significant fluorescence does not imply that measurements in the future will be free of fluorescence. This was our experience in MOHAVE where the ALVICE system arrived with a lidar shown to be fluorescence free in our previous field campaign but, due to deposition of fluorescent material on the receiver, it developed a fluorescence contamination. Therefore, there is a need to focus as much on on-going quality control measures and data correction as on the use of fluorescence-free optics. We make these points in the paper.

About the reviewer's comments regarding independence, we have to point out that any Raman lidar water vapor measurement that is calibrated with respect to an external measurement of water vapor, as is currently almost always the case, cannot be considered an independent measurement of water vapor. Any systematic error in the calibration source is carried into the error budget of the Raman lidar system and it is thus not independent of the water vapor calibration source. So we do not agree with the reviewer's concerns that application of a time-dependent correction would compromise the “independence” of the Raman water vapor measurements. Unless an absolute first principal's calibration procedure is pursued, the measurements are not independent. This statement is true with or without a correction for any biases.

I think that the ***real innovative aspect of this paper is represented by the possibility to use the water vapour measurements in the lower stratosphere from sensors different from lidar*** *(for example the Microwave Limb Souder), available in the last decade,* ***to correct biases in the lidar measurements and allow for trend analysis in the upper troposphere****. This important aspect of the paper,* ***which certainly deserves to be published****, becomes clear in sections 9 and 10 of the paper and I think should be made clear and moved to a much earlier stage of the text.*
The real focus of the paper is the correction procedure and not the results shown for the radiosondes, which do not really pertain to this paper. Probably, all the radiosonde calibration part (section 6.1), out of the main focus of the paper, could be moved to the appendix or to a separate paper, together with the present content of the appendix.

We agree that section 6 should be moved to an appendix. This will help to make the main points of the paper more quickly as the reviewer suggests. However, given that the paper is part of the special MOHAVE section of AMT, and MOHAVE was concerned with calibration and validation of water vapor instrumentation, this paper needs to cover more material than just the correction technique.

Concerning the presentation style of the paper, some portions of it are written ***in a narrative way and this penalize the paper****, which in some parts seems to be generic and in others not very clear.* ***The paper is by far too long and this makes not easy to capture the original and fundamental aspects of it.*** *Additionally, some large portions of the paper seem to pertain more to a technical report of a project than to a science paper.* ***Thus, an effort should be performed to shorten and re-write the paper in a more “science article” style.*** *Additional, an effort should be made to remove non-influent details. For example, I really don’t understand what is the benefit for a science paper to have such a detail description of the data format (section 6.2).*

We suggest to move all the material of section 6 to an appendix. Regarding the details of the items in the data format, we will add text to clarify that this is demonstrating, for what we believe to be the first time, the quantification of the full uncertainty budget of a Raman water vapor lidar. This is an important result from the standpoint of using these instruments for climate data records which require rigorously justified uncertainty estimates. By moving section 6 to an appendix, we believe that the main messages of the paper will be conveyed more tightly as desired by the reviewer.

Based on all the above considerations, I believe that this paper certainly contains new material and ideas that deserve to be published, but a certain amount of work is required in redrafting the paper with the goal of ***shortening it and putting in the proper light the very original and new aspects of it*** *before the paper can be published. In this regard, I would be happy and available to provide a final comment before the final acceptance of the paper.*
I wish to further specify that, do to the limited options of the evaluation form, I had to select “reconsidered after major revisions”, specifying that “I would like to review the revised paper”. However, I would more correctly define my evaluation as “accept after major revisions”, as the revisions are major but straightforward and I am sure they will lead to a final acceptance of the paper.

The reviewer believes similarly to Reviewer 1 that the paper is overly long. 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 first 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 mandated.

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, as we've said, we suggest moving all of section 6 from the main body of the paper into an additional appendix. We will also work to shorted 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

Introduction: I would partially modify it. In fact, besides two sentences on the importance of water vapour in the atmosphere, this duplicates the abstract in terms of information content. Almost no additional information is provided here with respect to the abstract. I would then modify the introduction, or at least widen it, with a more specific reference to the importance of having accurate water vapour measurements (probably mentioning the water vapour observational requirements for several scientific purposes associated with Climate and Meteorology) and highlighting the benefit of having reliable water vapour lidar measurements.

We don't really see a problem with repeating information in the introduction that is contained in the abstract but we will add some more introductory material to address the reviewer's concern.

The description of the field campaign (pages 7341-7344) is too long, filled with details that are out of the scope of the paper (for example, THPref), and could be shortened.

This paper serves multiple objectives given that it is part of the MOHAVE-2009 special section of AMT. The normal things that are contained in a paper describing such an intercomparison campaign (calibration techniques, comparisons versus other instruments, uncertainty budgets) need to be addressed. In addition we have to develop and justify the correction used. Both of these are accomplished in this single paper which certainly adds to the length but is the only practical solution given the unusual developments during the MOHAVE campaign (deposition of fluorescing material).

A couple of corrections where already indicated in the preliminary review, but they were note addressed by the authors. For example, as I indicated in my preliminary review, the following sentence is not clear (page 7344, lines 1-5). "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 RH≤10%, which corresponds to an uncertainty of ±6% at 50% RH, ±10% at 10% RH, and ±24% at 3% RH." The expressions ±(5% + 0.5% RH) and ±(7% + 0.5% RH) are not clear and some additional comments may be introduced to clarify them. ***Are the numbers percentage RH or uncertainty ?*** *For RH>10% the uncertainty is specified as ±(5% + 0.5% RH). Does this mean ±5.5 % ? These unclear expressions are used again in the Summary and conclusions (page 7375).*

The numbers and units/quantities are correct but we will add text to clarify their meaning.

Page 7349, lines 17-28. I think that the recommendations of the Joint Committee on Guides in Metrology in their Guide to the Expression of Uncertainty in Measurements (GUM) (JCGM/GUM, 2008) are obvious and really do not need to be included in quotation marks in a scientific paper addressed to experimental physicists. In line 17 you should remove everything after the reference to (JCGM/GUM, 2008). Some of the obvious sentences related to the JCGM/GUM (2008) are reproduced more than one times in the paper (for example, the sentence: "It is assumed that …identify such effects." is found both in page 7349 and in page 7378). Furthermore, the reference to JCGM/GUM (2008) is cited several times in the text, but is not listed in the reference list.

We are pleased that the reviewer considers the recommendations of the GUM to be obvious and appropriate to the current context. As stated above, however, there is considerable opposition to the concept of corrections within the NDACC lidar community. It is for that reason that the rationale and justification for the corrections is very important. For this paper to be successful and have influence within the Raman lidar community, we believe these points need to be made clearly and with some repetition where appropriate. These are the reasons for the direct quotations. It is strange that the GUM reference is not in the paper and we will certainly fix that.

In page 7355, lines 17-19, authors specify that: "The magnitude of the correction constants ζ_0 *and* ζ_1 *were chosen empirically to yield best mean agreement with the FP in the 10–20 km altitude range." This rigorous approach is characterized by this weak point here. A minimization of the root mean square deviation between the two sensors could be considered here instead of an empirical choice. Authors are invited to correct text and results accordingly.*

We considered various ways to determine this empirical value but there were no significant differences in the approaches so we chose just a simple one. No results of the paper are influenced by this decision.

Additionally, here you are commenting figure 2, where also equation (9) is tested. So why there is no mention also of ζ_2 *here ?*

We will refer to the correct variable in the revised manuscript.

In page 7356, line 5. Here for comparison I would also show how the correction from Eq. (5) and Eq. (7) would look when considering a night-time lidar run instead of 1 h.

We included this in an earlier version of the plot but felt it did not really add information since it made an already busy plot more so. The value of showing the curves that are present now in the plot is that the general magnitude of the correction can be seen despite the high noise in the results. Additional curves make this harder to see. We would prefer to leave this figure as it is.

Concerning Section 6.1, dedicated to the radiosonde based calibration technique, I believe that ***this is out of the scope of the paper and should be moved to a separate paper or in the appendix****.*

Viewed in the context of developing techniques for use of Raman lidar for generation of climate data records of water vapor, we believe the material relating to calibration is very pertinent. However, we feel that the main topic of the empirical correction would be more succinctly conveyed if we move this and all of section 6 to a separate appendix.

Additionally, authors mention that: "the algorithm used here for lidar calibration with respect to radiosonde profiles was developed as an outgrowth of the discussions at the workshop." If the approach was the outcome of the discussion at the workshop, I would imagine that the attendees to the workshop are either the co-authors of this paper or are acknowledged in the acknowledgments of the paper.

At the workshop there were discussions about the influence of atmospheric variability on calibrations of Raman lidar using radiosonde but there were no discussions of the ideas behind this algorithm. There are therefore no missing names on the paper.

A similar approach to minimize deviations between lidar and radiosonde data has been reported in literature by Mona et al. (2007, see p. 264), even if this was not ultimately used to calculate the Raman lidar calibration coefficient. Authors are invited to cite and acknowledge this previous work (see reference list below).

The procedure described in Mona et al, 2007 takes differing lidar profiles and composites those vertically to choose lidar data as a function of altitude that most closely match the radiosonde measurement time window. This is a technique that we have also studied previously in the context of the AWEX-G field campaign in 2003 where we referred to the technique as “TrackSonde”. We found that this approach to data analysis sometimes improves comparisons between lidar and radiosonde and sometimes degrades them. The reason that the comparisons are sometimes degraded is likely the failure of the assumption of a statically propagating atmosphere that this technique implicitly makes. We did not use the TrackSonde analysis procedure for the data shown in the paper. While atmospheric variability is the motivator for the Mona work and our new algorithm as well, there really does not seem to be much similarity between the time slice analysis of lidar and radiosonde presented in the Mona work and the adaptive calibration routine described in our paper. We therefore find it difficult to see how to reference the Mona work here.

Again concerning section 6.1, the description of the approach is quite qualitative here. Readers do not get sufficient information to be able to potentially reproduce it. This is especially true and needed for a technical journal. Thus, a more detail description of the algorithm is required, besides the flow-chart. Authors should introduce the expressions and formulas they use here, describing the algorithm in a comprehensive mathematical format. Additionally, all qualitative statements should be substituted by more quantitative ones. For example, authors mention that: “Ordered pairs are accepted as members of the final set of data used to determine the calibration value if, first, they were part of a regression with sufficiently high R_2 *and, second, if an individual ordered pair is within a certain percentage of a least median of squares fit line.” Provide information on what are reasonable values for both* R_2 *(authors specify “sufficiently high* R_2*”) and % deviation (authors specify “certain percentage”). Authors also specify “… insufficient number of points …”. Please, quantify what you mean for “insufficient”. There is also another point which is unclear: the text of page 7357 refers to* R_2*, which in most statistical textbooks is used to indicate the correlation coefficient; however, in the flow chart authors refers to the selection of a “minimum Rsq”, which is a different quantity, and no mention to* R_2 *is present in the flow chart. Please, clarify.*

This adaptive calibration routine relies on numerical algorithms for least squares and least median best fitting and various filtering techniques. While it would be possible to supply the equations for all of this it would be exceedingly tedious and add greatly to the length of the paper. Instead we suggest adding supplemental material that supplies the actual code developed for the procedure. Then the routine can be duplicated exactly if desired.

We will also clarify that Rsq stands for R^2 in the revised paper.

Page 7358, lines 14-15. Authors state that: “It was also found that restricting the RH values from radiosonde to values above 5% RH further decreased the standard deviation of the derived constants.” This is an important statement that needs to be supported through the provision of additional information. Authors should specify how much the standard deviation further decreased and in how many cases this was verified.

We will supply this information in the revised manuscript.

Page 7360, lines 6-25. Here authors should specify which of the error sources are systematic and which are random and specify which of the systematic error sources can potentially be corrected for (as only some of the systematic error sources can be corrected for). Additionally, it would be nice to *understand how the different error sources combine to finally assess an overall measurement error.*

We will add this information to the revised manuscript.

Page 7361, lines 3-4. Here authors mention that "The spatial resolution of all the profiles except the raw data profile is determined by the size of a moving window filter which varied from 30m in the lowest part of the atmosphere to a maximum of 1200m for ranges beyond 12 km." However, the size of the moving window filter is not the spatial resolution. Please, specify how these two quantities are related and how this translates in terms of spatial resolution values.

We actually provide that information but it occurs later in the paper. We will move that information to this point in the paper for added clarity.

Page 7361, lines 1-25. As already mentioned above, I don't see a need in a scientific paper to comment on the different data format.

Please refer to our comment above regarding specification of total uncertainty budget and its importance in developing climate quality data.

Page 7361, lines 12-15. Authors specify here that the: "corrections are for water vapor mixing ratio overlap dependence, temperature dependence of Raman scattering, atmospheric differential transmission and the signal dependent correction that is described in Sect. 5.". These, in fact, are among the systematic error sources which can be corrected for and this needs to be clearly specified for the benefit of the reader.

We will add this as mentioned above.

Page 7362, line 6. Authors here refer to the mean bias and RMS, but they don't give any information how this was computed (is this with respect to the mean of the sensors or with respect to one of the sensors ? what expression is used ?). Was "bias" and "RMS" computed point by point or considering a certain number of points. The use of the term "mean bias" suggests that a certain number of points are used here. Additionally, the presence in figure 5 of bias values not exceeding 30 % at 2.5 km for the "all-night" comparison, while values of "ALV-all" are 4 times (4000 ppm) larger than the values of "RS92" (1000 ppm) at this same height, suggests again that a mean on certain number of points is considered in the computation of the bias. Please, specify.

We will add information to this in the revised manuscript.

Section 6.2 and the first part of section 7 (until the beginning of section 7.1., i.e. the from line 26 of page 7362 to line 25 of page 7264) should be merged together. In fact this first part of section 7 does not at all refer to the "Comparisons of lidar profile and total column water vapor measurements", but only provides additional information in terms of profile-to-profile comparisons.

We suggest instead to move all of section 6 to an appendix. The problem cited by the reviewer is due to a poorly worded section title. The title is meant to refer both to profile-to-profile comparisons and lidar-total column comparisons. We will revise the section title to clarify this.

Page 7363, line 29. Authors write: "Below approximately 12 km, the lidar profiles are wetter than the FP profiles by approximately 10% discounting the regions of high atmospheric variability at *altitudes of approximately 4, 8 and 11km where deviations were higher." Why should the values at 4, 8 and 11km be discounted ? Then authors also say: "These differences are believed to be mainly due to a tendency for the atmospheric conditions over the mountain top site to moisten during the night and not due to real instrument measurement differences". Do you refer to the values at 4, 8 and 11km or to the wet values by 10 % ? This is not clear.*

The statements refer to the results at 4,8,11 km. The reason to discount them is due to the belief that the differences noted in the comparisons are caused by actual atmospheric variability and not differences in the measurement systems. This is a problem with comparisons based on a small number of samples. We will clarify these details.

Again in section 6.2 the reference to the document JCGM/GUM (2008) leads to the introduction of sentences that are somewhat obvious for a scientific paper addressed to experimental physicists.

We refer again to the desire of this paper to convince a skeptical audience of the need for and validity of these corrections. We believe that reference to the GUM and direct quotation motivates the corrections rather well and needs to be better understood within our community. We prefer to retain this information for those reasons.

Results of section 7.1 are out of the scope of this paper. This section should be shortened or moved to a separate paper as it contributes to limit the readability of the paper (far too long). Additionally, a lot of details introduced here are really not needed for a science paper, but are more for a technical note (for example: "The integrated precipitable water and pressure data from SA65 were combined with the temperature and RH data from the THPref instrument to provide a surface reference datafile containing RH, T , P , IPW and water vapor mixing ratio with a 5 min temporal resolution for the period 10–27 October as previously shown in Fig. 1").

Redundancy of calibration source is important for data quality control and development of climate data records. GPS has been shown capable of providing a useful calibration for Raman lidar if the overlap correction can be made reliably. In light of these considerations, it is useful to show how the calibration with respect to GPS compares with that of radiosonde. We will add text to clarify the value of having additional calibration sources when considering the task of developing climate quality Raman water vapor lidar data.

Page 7365, lines 7-10. Here, I would not speak about "agreement of calibration", but about "agreement between instruments", as the final outcome of that cited paper was the agreement between the Raman lidar, the frostpoint hygrometer and the airborne DIALs within 10 % after the Raman lidar calibration. The sentence should read: "During IHOP, comparisons made between lidar and frostpoint hygrometer (Whiteman et al., 2006a) and airborne water vapor lidars (Behrendt et al., 2007) showed agreement to generally better than 10 %."

We will revise this sentence.

In page 7365, lines 21-23, authors specify that: "By carefully selecting radiosonde profiles in a manner similar to that described in Sect. 6.1, an overlap correction was derived as the mean ratio of radiosonde and lidar data for the selected profiles." This is certainly true as long as you don't change the bore-sight alignment. Afterwards, this might be no longer true. Didn't the authors need to change (even slightly) the bore-sight alignment during the measurement campaign ? If yes, have *this change been accounted for in the overlap function ? What are the implications in terms of uncertainty on the overlap function ? Please, comment.*

We have an automated bore site system that maintains the alignment of the laser and telescope to within 10 – 20 microradians. We do not notice any significant change in the overlap function due to alignment. We account for the uncertainty in the overlap correction by including the standard deviation of overlap correction in the propagated uncertainty budget. We will add these points in the revision process.

Concerning the determination of the overlap function based on the use of the radiosonde data, authors should also specify how they deal with the fact that this estimate makes the lidar data no longer independent from the radiosonde data. Was the overlap determination based on a sub-set of radiosonde-lidar comparisons which were then excluded from the final comparisons illustrated in the paper ? Please, comment.

We must make the point again that the lidar data, being calibrated by either radiosondes or CFH or GPS or whatever external source of water vapor information, in no way offers an independent measurement of water vapor. This is a very important point. The application of the overlap correction permits the lidar data to best represent the profile of water vapor in the lower atmosphere with well-specified uncertainty budget (including the additional uncertainty due to the application of the overlap correction). The goal of a measurement is to supply a value of a quantity and a rigorously determined total uncertainty in that value. There is no added value by stating that a measurement is dependent or independent, black, green or blue.

Section 8. Needs to be shortened. A lot of information present in this section were already provided before in the paper.

Actually this is new information. The earlier discussion of biases was for systems not participating in MOHAVE. This material indicates that numerous biases were encountered along the way in the analysis of MOHAVE data. A point that we will add to the paper is that the correction developed here can be useful in that analysis.

Section 11. Summary and conclusions is far too long and need to be sensitively shortened, using a more concise style.

The summary and conclusions section is often one of the only parts of a paper that a reader spends time on. Many readers do not read the entire paper. Given the need to carefully justify the technique presented here, we would prefer to keep this section as it is.

As mentioned earlier, the appendix dedicated to RS92 RH accuracy and corrections is out of the goal of the paper and its presence here reduce the readability of the paper. This could be moved to a separate paper. However, this is not a mandatory request.

We do not have the resources to make this a separate paper. The results are significant in that they document the accuracy of one of the dominant calibration sources of Raman lidar. We wish for these results to be citable so would prefer to keep them in this paper.

Other points

Page 7342, line 3. Should be "mJ" and not "mj".

We will change.

Page 7342, line 10. Is 250 μradians HWHM or FWHM ?

FWHM as is customary with lidar.

Page 7342, lines 16 and 19. As far as I know the proper spelling should be "Di Girolamo" and not "DiGirolamo".

We will make this change.

Page 7343, line 25. I would correct into "SuomiNet total column water data, IPW, are reported here …" In fact the quantity IPW is used again in page 7365, line 15, without being previously defined.

We will define the term as suggested.

Page 7344, line 12. Should read "ALVICE".

We will make this change.

Page 7352, line 20. $\sigma_w = \sigma_{w^*} + \sigma_{\zeta_l}$*, Expression 8 is introduced, but there is no explanation of the expression and of the terms in the expression. These should be explained here (where the expression is introduced, and not at a later stage). Additionally, if these are standard deviations I don't understand why they simply sum instead of root sum square. Same is true (lack of definition and wrong expression) for expression 10 and expression 12. Authors are invited to correct text and results accordingly.*

The squares were mistakenly omitted. We will correct.

Page 7353, line 6. It should read: "… the uncertainty on ζ_l *is …"*

we will clarify which quantity is referred to here.

Page 7353, line 27. It should read: "… 20–30 ns, with maximum values …"

we will add the comma.

Page 7354, line 4. Substitute "6000 to 8000 ns" with "6 to 8 μm".

We prefer to leave the numbers as stated so that the comparison of magnitude is more immediately obvious.

Page 7357, line 26. Should be "… left panel shows …". Line 29. Should be "… right panel are …". Page 7358, line 2. Should be "… lower left panel …"

We feel that the current description is more clear and would prefer to make no changes.

Page 7361, line 8. Point 2. Authors specify that "all available data for a given night, independent of altitude." What does "independent of altitude" mean here ? Previously you had specified that "all the profiles except the raw data profile is determined by the size of a moving window filter ..."

We will add some text to clarify. What is meant here is that the temporal window is height independent.

Page 7361, line 11. The number "4" should be expressed as "four" here. Line 16. The number "3" should be expressed as "three" here.

We will make these changes.

Page 7362, line 6. Authors write "The mean bias and RMS for 33 Vaisala RS92 and ALVICE best estimate, 1 h and all night profile comparisons are shown in the middle and on the right of Fig. 5.". What does "mean bias and RMS for 33 Vaisala RS92 and ALVICE best estimate" mean ? What does "33" refer to here ?

We will expand this statement for clarity.

Page 7363, Figure 6. Values of bias are specified. Again, how are these computed ?

We will add text about the computation. We do not understand what the reference to "Figure 6" above is meant to indicate, however.

Page 7363, lines 19-20. Authors specify that: "The comparison of 1 h lidar profiles and the frostpoint measurements shows generally more scatter than the comparison versus the RS92 due to the reduced statistics." Reduced statistics of what ? Please, specify.

We will change "statistics" to perhaps "number of comparisons".

Page 7375. Summary and conclusions. The following sentences are redundant here and should be removed: "The time series of ancillary measurements is shown in Fig. 1. The measurements from the surface reference system called THPref and frostpoint hygrometer were used to characterize the accuracy of uncorrected and corrected Vaisala RS92 data acquired during MOHAVE-2009."

We see no redundant sentences here but we will look for such cases in the Summary and Conclusions.

Page 7375. Summary and conclusions. With reference to the sentences: "The comparison to FP shown in the appendix is consistent with this uncertainty estimate, but still there is evidence that the calibration correction documented in Miloshevich et al. (2009) is less accurate for 2009 radiosondes than for 2006–2007 radiosondes, the vintage of sensor used to develop the correction, due to expected changes in the RS92 mean bias with time, indicating that the uncertainty estimate is conservative." This aspect would need to be better understood and could be introduced and discussed in more detail in a separate paper including the comparison of ALVICE vs. radiosondes during Mohave 2009, the new approach for radiosonde-based Raman lidar calibration and the material presently in the appendix.

We do not have the support to write another paper on this topic. In order for these results to become published and preserved in the refereed literature, we would prefer to retain the material that is currently in the appendices of the paper

Throughout the paper authors refer to Whiteman et al., 2011a, and Whiteman et al., 2011b. However, in the reference list two papers from Whiteman et al. (2001) are present, but it is not specified which one is a and which one is b.

We will clarify this