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.

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.

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).

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.

Specific comments:

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

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.

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.

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.

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.

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.

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.

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.

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.
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.
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").
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 %."
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.
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.
Section 8. Needs to be shortened. A lot of information present in this section were already provided before in the paper.
Section 11. Summary and conclusions is far too long and need to be sensitively shortened, using a more concise style.
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.

Other points

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

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

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

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.

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

Page 7352, line 20. $\sigma_w=\sigma_{w*}+\sigma_{\zeta 1}$, 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.

Page 7353, line 6. It should read: "… the uncertainty on ζ_1 is …"

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

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

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 …"

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 ..."

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

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 ?

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

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.

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."

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.

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.

References

Mona, L., C. Cornacchia, G. D'Amico, P. Di Girolamo, G. Pappalardo,G. Pisani, D. Summa, X. Wang, V. Cuomo, Characterization of the heterogeneity of the humidity and cloud fields as observed from a cluster of ground-based lidar systems, *Quarterly Journal of the Royal Meteorological Society*, **133**: (S3), 257–271, (2007), DOI: 10.1002/qj.160, Ed. *Wiley*, December 2007, Hoboken, N.J. (U.S.A.).