Review of «Calibration of a Water Vapour Lidar using a Radiosonde Trajectory Method » Shannon Hicks-Jalali et al.

The paper by Hicks-Jalali et al. presents a new version of Whiteman et al.'s (2006) Raman lidar calibration method with radiosondes. This new method takes into account horizontal air mass movement. Hicks-Jalali et al. present a detailed description of the uncertainty associated with calibration.

General comments:
Shannon Hicks-Jalali compares the calibration results obtained by the « traditional » and the « improved » methods. If we regret that the retrotrajectory work is only done in 2D and not in 3D, the idea is interesting to continue to overcome the problems related to the non-colocation of the reference measurement with that of the lidar. However, I remain unconvinced by the discussion and conclusions of the article which consider that the new method is significantly better (in case of an « heterogeneous » atmosphere) because it is poorly supported by the results of the article in terms of number of calibrated nights (no more) or uncertainties (not less) with the use of both methods, and does not benefit from any validation (comparison of calibrated profiles with a third instrument). It is imperative to rework the discussion and review the method's contributions in a more factual way.

I think the article is not publishable as it stands, the article needs to be reworked to answer the following major remarks (more specific comments follow):
1. The structure of the paper needs to be reworked to improve clarity and precision. I suggest to include section 3.2 into current section 2.2. I suggest to improve internal structure of Sect. 4 (see major comments n°4), to remove the Summary section (you should share its content into sections 4, 5, the conclusion and avoid repetitions) and to split the current section 7 into a Discussion section (that you need to develop) and a Conclusions section.

2. The « traditional » method is not described simply at the beginning of the article, there is just a list of bibliographical references that the reader must read without knowing which one is used precisely. Explanation loops are given as the article progresses but they arrive too late. A short or detailed description of the so-called « traditional » method should be added to the introduction and/or methodological part.

3. In the description of the lidar measurement and the description of the « improved » methodology, a lot of choice in filtering the data are made based on high or low SNR but it is never quantified. Please be more specific on this point.

4. Sect. 4 presents many issues:
    4.1 The structure needs to be reworked so that the reader can have the following elements:
      ○ Presentation of data
      ○ Methodology to differentiate between nights when water vapour is homogeneous and nights when water vapour is heterogeneous
      ○ Presentation of Table 1
      ○ General comment (the current last paragraph)
      ○ Illustration of the different characteristics with Fig. 5 and 7
      ○ Conclusion with Figure 8
    4.2 Figures:
      ○ they are under used, even not used for specific subplots. Maybe there is too many figures,

○　the legends need to be shortened (some analyses of the figures are made in the legend whereas it should be done in the text),

　　　4.3 The quantification of biases of « 0 % » in average. I suggest taking absolute values and indicating the sign of the bias.

5. It is repeated over and over in the summary and discussion/conclusions that the method is « more accurate », in other words but the results on uncertainty does not quantify this improvement. The discussion/conclusion about the advantages of the method, it must be thorough:

　　　5.1 Almost no discussion about the limitations of the method: 32 % of night are calibrated (what about the others?) on 6 years (the first 3 years not being calibratable also)

　　　5.2 The benefits presented are based on the theoretical expectations that motivated the implementation of this methodology. The uncertainty is presumably better but this is not reflected in the budget calculation.

5.3 Discussion should be pushed before opening the perspective of using this dataset to « UTLS climatology over 10 years »: how to do that with 24 nights above 2008-2016, how to calibrate night where this methodology could not be applied? It should have thought because the authors wants to use the whole dataset for trends study.

Specific comments:
Title
I strongly suggest to refer to RALMO in the title or to find a way to indicate that it is a test of this new methodology of calibration done on <u>one</u> lidar which <u>could potentially be</u> applied to others.

Abstract – page 1
Lines 1-4: I would keep theses sentences for an introduction because it is too general. At least, please shorten this part.
Lines 5 & 6: Reference citations should not be included in this section unless they are essential. Using radiosondes is the most used technique for calibrating so please select a maximum of one reference, I would suggest Whiteman et al. (2006) which is the closest from the one you will use?. Maybe, the method is even better summarized in Whiteman et al. (2012).
Line 7 « movement of radiosonde »: I suggest replacing it with « movement of air masses »
Line 12: Precise on which period the calibration has been performed (i.e. 2011-2016).
Lines 14-15: The authors use « more accurately » but there is no conclusion in the article that quantifies that the uncertainty associated with the new technique is better than the « traditional » one . I suggest replacing « reproduces more accurately » with « reproduces accurately ».
Lines 16-21: The summary associated with the uncertainty budget is too detailed. Please replace this part by one value (or range of value) quantifying the total uncertainty associated with the calibration.

Page 2
Line 2: Please replace « the primary contributor» with « one of the main contributors ».
Line 3: Add reference to « ...high temporal and spatial variability »
Line 4: I suggest deleting « uniquely ».
Line 6: I suggest deleting « more » or please add a reference.
Line 8: Replace « take » with « make ».
Line 9: Delete « also ».
Line 10 « Several Raman… external methods »: I would place this sentence in the following paragraph.

Lines 17 to 22: It's too detailed whereas it is not the main subject of the article. I suggest deleting this part.

Line 29: Please add the use of GNSS as an external instrument to calibrate Raman lidars and a reference. I suggest: David, L., Bock, O., Thom, C., Bosser, P., and Pelon, J.: Study and mitigation of calibration factor instabilities in a water vapor Raman lidar, Atmos. Meas. Tech., 10, 2745-2758, https://doi.org/10.5194/amt-10-2745-2017, 2017.

Page 3

Lines 1-2 « External...do not contribute »: I suggest deleting this sentence.

Line 7 « as RS92 radiosondes are the most frequently used calibration radiosondes Immler et al., 2010 ; Dirksen et al., 2014). »: I'm not convinced that the main objective of this article was to correct sondes for the calibration of Raman lidars. Please rephrase.

Line 8: Replace « Vaisala » with « not corrected »

Line 9: Replace « errors » with « uncertainties »

Lines 9-10 « A portion of… uncertainty »: I suggest moving this sentence in the next paragraph, line 17 before « This paper attempts to... ».

Lines 12-15: I suggest moving the whole paragraph line 27.

Line 19: Please explain in few sentences (2-3) what is the « traditional » method, references are not sufficient considering that the improvement of this technique is the main subject of this paper.

Line 20 «as the radiosonde takes approximately 30 min to reach the tropopause »: It depends on which latitude the sonde is launched, it will be larger near the equator. Please specify it or add a location.

Line 21: Your statement should be supported by a reference or some statistics from your database.

Line 28 « in order to ensure that the lidar and the radiosonde are measuring the same air »: It was also the goal of the « traditional » method. Please check.

Line 28: I would suggest deleting « improved ».

Line 29: Replace « of the radiosonde and the » with « , ».

Page 4

Line 7: Add « in Payerne » at the end of the sentence.

Line 9 « their respective uncertainties »: The full uncertainty budget for the calibration is not given for the « traditional » method. What about the « representation uncertainty » for example? I will come back to this in more detail in following comments.

Line 13: Does the lidar always start working at 0:00UTC? Does « bi-weekly » refer to twice a week or one every two weeks? Please specify it in the text.

Line 14: Why is their only « a subset » of these radiosonde processed by GRUAN? Please explain it in the text.

Line 23: Please specify that the analysis was conducted on an initial set of 76 flights but in the end only 24 of them were used.

Page 5

Section 2.2: Precise that the RALMO is operating day and night. Give the effective measurement time (« 50% » in the conclusion, it should be precised earlier) and what explain that 50% of the time is not exploitable.

Line 2: The authors explained that the instrument « is designed to be an operational lidar, and as such, <u>needs</u> to have high accuracy, temporal measurement stability, and minimal altitude-based corrections (Dinoev et al., 2013; brocard et al., 2013) ». The study of the instrument's performance made in the bibliographical references and years of operation should determine whether the instrument really has a high accuracy, temporal stability of measurements and profiles that start close to the ground. Please be more specific.

Lines 4-6 « RALMO operates...scattering channels »: Move this sentence to the previous paragraph.

Line 9 « a sufficiently high SNR »: This is a major issue in this paper. The SNR is often used to select some data and reject others, but no threshold based on bibliographic references or empirical tests specific to this study is defined. You need to clarify this.

Lines 12 & 13 « After the filtering process...small features »: I wonder if this sentence should not be found in the description of the calibration methodology. Or at least move the interpolation part of the radiosonde profiles to section 2.1 and specify here that there is no interpolation for lidar profiles.

Line 15 « 3.1 Tracking air parcels »: I would move this title at the end of the page. Paragraph line 16 to 20 is more an introduction to sect. 3.

Page 6 – Figure 1

I would use « homogeneous region » instead of « homogeneous cylinder » to be more consistent with the text. I would suggest adding letters or numbers to the different steps and refer to them in the text.

Page 7

Line 7: As simple that it could be, you should provide mathematical explanation/an equation to illustrate your calculation.

Line 8: This assumption should be discussed a little more. You might provide a physical discussion about when (or if) this assumption is (would be) realist and its limitation, please cite references that could support this hypothesis or this discussion.

Line 9: What do you mean by « We do not <u>explicitely</u> consider the vertical movement of the air parcel in this method »? From your description of the methodology, I understand that you do not consider at all the vertical movement. Please be more specific.

Line 14: For my point of view, the « lidar region » refers more to the ~1 m diameter at 5 km that the radius of 3 km choosen after for the sensitivity test. You might use a name more related with the assumed homogeneity of the water vapor like or its use for calibration, « homogeneous lidar region » or « calibration region » as you call them after in the paper.

Lines 14-15: What is the decisional parameter? Is this the SNR? Which threshold?

Lines 16/17/18: You characterized some « very low SNR », a « large enough SNRs » and « the highest SNR ». Please quantify or explain why you define it this way.

Page 8 – Figure 2

Is this a conceptual scheme or a real example? Please precise it in the legend. If it is an example, give the date of the measurement. The second sentence of the legend is also explained in the text. Please replace this sentence with something like « The purple circle corresponds to the lidar region ».

Page 8

Lines 6-7 « The standard thirty minutes...tropopause »: It sounds quite general in your text but it corresponds to mid-latitudes. It would be a different duration for polar regions or in the tropics. Please be more specific by adding « at mid-latitudes » for example.

Lines 7-8 « Integrating...by radiosonde »: I suggest deleting this sentence.

Lines 9-10: The analysis of Figure 3 consists in one sentence. Either your analysis is too short, either the figure is not necessary. Please reconcile.

Page 9 – Figure 3

Legend « The integration time…analysis. »: Please remove these sentence from the legend. It is already explained in the text or is part of the analysis.

Figure 2 – Figure 3 – Figure 4
In this section, Figure 2 seems to be a conceptual method, Figure 3 refers to July 21 and Figure 4 to July 22. It is not explained why these 3 dates are choosen for each example. Considering that Section 3 details the method of calibration by sonde, you might choose the same date to illustrate the different aspects of it. If not, please justify why.

Section 3.2 - Pages 9 and 10
I suggest moving this section in or after Section 2.2 because it is more about the « Lidar measurement » than about the « Radiosonde Trajectory Method ».

Page 9
Lines 10-11: I would rephrase this way: « The central wavelengths of the water vapour and nitrogen channels of the RALMO were choosen to minimize temperature dependence. »
Line 14: Please add a reference.

Page 10
Lines 4-5 « In RALMO's case, the ratio...2014). »: Please move this sentence to Sect. 2.2.
Line 8 « the corrected signal »: Which correction? Please be more specific.
Line 9: Does « the correlated and weighted least squares fitting » correspond to the « traditional » method? If so, why not call it this way in the abstract so that you don't list so many references? It could be included in the introduction in this way as well.
Line 14 « low SNRs »: Please specify.
Lines 15-17: It is interesting because there is a desire to minimize the bias associated with not taking into account the vertical displacement of air masses.
Line 20: I suggest replacing « lidar region » with « calibration region ».

Section 3.3 – Pages 10 and 11
It seems to be the traditional method that is mainly described here except that the lidar data are selected as described in Section 3.1. Be careful to distinguish in the text of this part (and even in the entire article) between what is specific to your study and what is traditional. Perhaps the structure you have chosen is a little confusing on this point.

Page 11
Lines 2-10: I suggest moving this part to Sect. 2.1 or at least be more concise.
Line 12 « for July 22, 2015 »: Why did you choose this date instead of July 21 as in Figure 3? That would seem more consistent.
Line 15: Please replace « error » with « uncertainty ».
Line 15: Please add a reference. What was the range of uncertainty found in the litterature for the « traditional » method?

Figure 4 – Page 11
The date is not specified, please do it. The last sentence is not necessary.

Page 11
Line 3: What do you mean by « repairs »?
Line 4 « abnormaly high »: Please quantify or explain.
Line 6: I suggest: « and presence of clouds ».
Lines 6-7 « The filtering process...radiosonde launch »: This sentence seems to say that there were clouds every night of measurement for all the duration of the measurement. Is that what you mean?

Does this mean that there are 3 years of measurement that cannot be used because of systematic presence of clouds?

Line 8: If I understood correctly, between 2008 and 2016, there are only 76 calibratable nights and they are all condensed over the period 2011 and 2016. What solution for 2008-2010? On the other hand, over the 76 nights, with the implementation of this methodology, only ~30% can be calibrated? What about the other nights of measurements? It is essential to discuss these aspects in Section 7.

Table 1 – Page 13
Column « Difference »: Sign of the difference is missing or it should be specified that it is absolute value.
Legend: I suggest deleting the word « calibration » in the second sentence. The text from « Two nights in the homogeneous » to « variability in the water vapour » should not appear in the legend.

Page 13
Line 6 « show a good agreement »: Please quantify or explain.
Line 7: I suggest adding « water vapor conditions <u>around the location of the lidar measurement</u> ».
Line 8 « as « homogeneous » or « stable » nights in Table 1 »: Only « homogeneous » is used in Table 1 in comparison with « heteregenous » , « stable » is used in the « Comments column » for heterogeneous nights. Please reconcile.
Line 9 « an average bias of 0 % »: Please used absolute value to calculte the average bias. Maybe you should use the standard deviation and specify if the bias is more positive or negative.

Page 14
Line 1 « The bias on that night is reduced when using the trajectory method »: What do you mean? Please rephrase.
Lines 2-3: Figure 5 is composed of 9 subplots and show results for 3 dates but there is only one sentence that refers to it in the text. You need to analyse your figures or do not put them in the paper.

Figure 5 – Page 14
Legend - 4th sentence: Why do you mention « White vertical regions » whereas there is none of them in the figure? Please delete this sentence.
Legend - 7th sentence: You should use « magenta » instead of « pink » or change the color of the corresponding line in Figure 5.
Legend - 8 and 9th sentences: It is part of the analysis and should not appear in the legend. Please delete them.

Figure 6 – Page 15
There is no 2012-07-27 measurement in Table 1. Is this the right date? Please check.
Legend – 1st sentence: I suggest deleting « launch at 0 min ».

Page 15
Lines 1-5: This part explains the method to characterize a night as homogeneous or heterogeneous, it should come earlier in this section; before discussing about the results. Please reconcile.
Line 3: Figure 7 is quoted before Figure 6 (p16 - l.11) in the text. The order of the figures is not respected.
Line 3: As for other figures in the article, Figure 7 is commented in only one sentence.
Line 8: See my previous comment on the average bias.
Line 9: Please refer to a figure or a table.

Line 14: Why « 11 » whereas there are 12 heterogeneous nights in Table 1?

Figure 7 – Page 16
Why did you choose these three nights? You must analyze all three of them, when you add a figure it should bring an supplementary information, otherwise it means that the figure is not necessary.
Legend – 2nd to 6th sentences: This is already explain in legend of Figure 5, please just leave the part on the white vertical regions and refer to legend of Figure 5.
Legend – 7 and 8th sentences: Again, this does not have to appear in the legend.

Page 16
Lines 1-2: Shouldn't these nights be described as homogeneous in this case? Does this not question the methodology of homogeneous versus heterogeneous characterization?
*Line 4: T*he average bias is around 1 % here whereas it was 0 % earlier (p15 - l. 8). Please reconcile.
Line 5: Please specify what threshold you chose to define that the variability increases above 5 km.
Line 6: Here you precise the sign of the bias whereas you have not done it earlier, same for the significant number of digits. Please reconcile.
Line 7: Please rephrase the sentence.
Line 8 « better fits »: Please quantify.
Line 11: Once again the commentary on the figure does not exceed one sentence. Please develop or remove the sentence.

Figure 8 – Page 17
I suggest choosing one reference (either the trajectory or the « traditional » method) to avoid the multiplication of the data and to put the average bias and the standard deviation on the same plot.

Page 17
Starting Page 16 -line 10 to Page 17 – line 7: This is a description of the methodology and should come earlier in the section. Please put this explanation after the introduction of Table 1.
Lines 3-4: What do you mean by « the majority »? How many nights? In Table 1, it seems that it is true for 6 nights on 12. This represents half of the data not the majority.
Lines 4-5: Please specify the dates .

Page 18:
Lines 2-7: You repeat yourself. This was already mentioned earlier in the article.
Lines 17-18: Please precise how the fitting uncertainty is calculated.

Page 19
Line 2: Which part is the systematic part versus the statistical part of the uncertainty? This is important to determine given that one of your objectives is to establish trends in the UT/LS.
Line 5: Please explain what do you mean by the « measurement vector » with regard to the lidar measurement.
Line 12: Please correct the indices of the third term of Eq. (5)
Lines 13-25: The comments of the equation is not clear enough. First you should describe each term, precise the source of each uncertainty (radiosonde measurements, lidar (including lidar's photon counting and deadtime and covariance term), quantify the uncertainties before (« 8 % » coming to late in the next section, ?, 5 %, calculation?) and after propagation (values of Table 2) and you should use Table 2 to support this section. Actually, Table 2 comes too late in the paper. You should also precise how is considered your uncertainty on the calibration constant due to the

radiosonde: statistical or systematic? Then you can conclude on the total uncertainty of the calibration constant for which you are found after propagation: an average value of 4%.

Line 16: Why do you mentioned that the fitting uncertainty is the same that lidar's photon counting uncertainty. As you do it l.24 to compare it to the deadtime uncertainty? Is there a physical or metrological explanation? The fitting uncertainty seems to be due to the methodology and is different from these two uncertainties.

Line 23: I suggest not making a line break here and including it in the previous paragraph.

Line 23 « we assume a deadtime uncertainty () of 5% or 2 ns »: How is estimated this uncertainty: literature? test?

Line 25: Please remove « including dead time effects ».

Page 20

Line 7 « The calibration time series »: Do you mean the 24 nights? Why don't you show them on a figure?

Line 8: What do you mean by « de-trended »? Please explain.

Line 8 « over ten years »: You work on the 2011-2016 (i.e. 6 years), why do you speak about 10 years?

Line 17 – Page 20 to Line 14 – Page 21: This « Summary » part should not be a part in itself. These information should be in sections 4 and 5 and in the conclusion. It is necessary to rethink the structure at the end of the article, which for the moment consists in two parts: 6 Summary and 7 Discussion and Conclusions (see comments below).

Line 23 « due to the lidar profiles »: Please rephrase.

Page 21

Line 4: Please change « error » into « uncertainty ».

Line 4 « is negligible »: Do you neglect it in your calculation? This uncertainty does not appear in Table 2. Please specify it in the text.

Line 5: This part is supposed to be a summary but this value « 8 % » appears for the first time. This should have been discussed earlier.

All Sect. 6: This section contains many repetitions and brings many repetitions in relation to the whole article. There are key elements (such as the quantification of the uncertainty on the radiosonde of 8 % or Table 2) that should appear earlier in the article.

Page 22

Line 1: This article lacks a real discussion part, so it is necessary to restructure it as follows:

      6 Discussion

      7 Conclusions

Line 2 « has several advantages »: What about the limitations? The following points should be discussed: no consideration of the vertical movement of air masses, only 6 years calibratable over 9 and only 32% of the 76 nights with exploitable radiosonde profiles according to the method's prerequisites. A real issue is: which method did you use for all other nights of measurements than the 24 selected? Because otherwise it means that over 9 years of data there are only 24 nights of calibrated and therefore usable data? We can't make trends over 24 nights.

Line 19 « an automatic […] scheme »: What do you mean? Does this mean that this method has been implemented in a production chain for water vapor profiles of the RALMO lidar?

Page 23

Line 1: If we read the reasons for the exclusion of measurement nights for the calculation of the calibration coefficient on page 12, there does not seem to be more or less nights used with the new method rather than with the « traditional » method. Please reconcile.

Line 5: The uncertainty has not been quantified for the « traditional » method (at least not indicated in the article) or in this article. This is something that is difficult to quantify, but as a result it is difficult to conclude that it has improved it. Indeed from a theoretical point of view it is, but in practice (and in this discussion) you do not relate it to your results. This uncertainty appears in your introduction and in the conclusion part but no word in the section focusing on the uncertainties.

Lines 8-9 « This method could be conceivably... »: Yes , it is a good idea for ozone measurements.

Line 10 « The method could... »: Yes but it should be presented as a limitation and not only as a perspective because it means that the « representation uncertainty » may not be as negligible as that.

Line 12: I suggest starting here Sect. 7 Conclusions.

Line 15: How do you assess that? Reference? Maybe Whiteman et al. (2011b)?

Line 15 « 50 % uptime over ten years »: Why is this only appearing now? 9 years of measurements -> 76 nights that can be calibrated -> 24 nights calibrated in practice, what about the other nights? Where is that 50%? Which calibration methodology for the entire database?

Line 24: To study trends in the UT/LS, please recall the total uncertainty associated with your profiles or refer to the article that assess the performances of the RALMO to measured water vapor in the UT/LS on a routine basis.