Review of “Study and mitigation of calibration error sources in a water vapor Raman lidar” by David et al.

Submitted by David N. Whiteman, NASA/GSFC

The subject paper presents interesting work that addresses various sources of calibration uncertainty in a fiber-coupled Raman lidar. The calibration variations are mainly due to 1) the influence of variations in bore-site alignment on the performance of fiber optics, 2) inhomogeneous photomultiplier photo cathodes and 3) vignetting in the optical system. There is much work worthy of publication, but I recommend significant changes to the manuscript prior to it being suitable for publication.

General comments

The authors are comparing lidar measurements of mixing ratio with point measurements in the near field as well as integrated total column water versus GPS. Both of these methods are potentially influenced by the lidar overlap function, yet this issue is not addressed. Also, the lidar profile is extended upward beyond a certain altitude through the use of a radiosonde profile and model. Some quantitative discussion of how much the non-lidar portions of the profile influence the entire column is needed.

The authors make use of measurements using a nitrogen filter in both the water vapor and nitrogen channels to monitor drift in measurements. Exactly how this is implemented is not clear from the current discussion. Presumably the water vapor filter is removed to perform this test, in which case the drift measured does not include any changes in the water vapor filter. Also, since the PMT cathodes presumably have different quantum efficiency at the water vapor and nitrogen wavelengths, it seems that using the channel ratios to correct the drift must involve some calibration normalization procedure which is not explained.

There are some references that are not well chosen to support the statements that are being made.

Finally, there are numerous awkward uses of the English language which sometimes make for confusing sentences. I point some out in the specific comments. The authors are encouraged to address these and the other awkward uses of English for improved clarity of the discussion.

Specific Comments

1. Line 17. The term “a unique nitrogen filter” is used where something like “a single nitrogen filter” is likely intended.
2. Line 30. The statement made is that the ability of the Raman lidar to make water vapor measurements on small spatial and temporal scales “is of prime interest for weather forecasting and climate studies”. The citation given is Held and Soden, 2000. Do Held and Soden explicitly mention Raman lidar for providing such measurements? The way this sentence is structured, that would seem to be implied.
3. Paragraph starting with Line 46. A discussion of independent and dependent calibrations is given. But the authors neglect to discuss the work of Venable et al, which demonstrated a scanning lamp technique using, Avila, 2004 water vapor cross section values, that agreed with corrected radiosonde to better than 5%. The Avila 2004 cross section calculations were based on a full *ab initio* simulation of the water vapor molecule unlike the earlier model used in Sherlock et al, which only considered trace scattering as I recall.

   1. D. D. Venable, D. N. Whiteman, M. N. Calhoun, A. O. Dirisu, R. M. Connell, and E. Landulfo, A Lamp Mapping Technique for Independent Determination of the Water Vapor Mixing Ratio Calibration Factor for a Raman Lidar System, , Applied Optics Vol. 50, Iss. 23, pp. 4622–4632 (2011)
4. Line 58. Convoluted → Convolved
5. Line 59. Statement is made that the absolute accuracy of the independent method is not better than 10% and is not referenced. I suspect the authors are referring to the uncertainty of 10% given for the Raman cross section ratio in the Penny and Lapp 1976 paper. The Avila cross sections agree to within 8% of these numbers (and are thus within the P&L uncertainty estimate) but the Venable et al. calibration work shows agreement with carefully performed radiosonde-based calibration to better than 5%. These two results would tend to indicate that the absolute uncertainty of the integrated q1 cross section based on the Avila, 2004 work is better than 10%. In fact, Fernandez-Sanchez, the lead of the Spanish group that performed the Avila, 2004 cross section measurements, has stated (private communication) that the uncertainty of the integrated water vapor q1 Raman cross section from the Avila work is approximately 5%.
6. Lines 78-82. There are various ways that the technique described in Leblanc and McDermid, 2008 can fail to detect sources of calibration variation. Those failure modes are discussed in the reference below which should be cited. It is for the reasons in the cited paper that NDACC did not, in fact, adopt the technique described in Leblanc and McDermid, 2008 as the standard calibration technique within NDACC, despite what has been published. At the NDACC Raman water vapor lidar workshop held in Greenbelt, MD in 2010, which I chaired, we recommended the use of one or more lamps for assessing instrument performance but neither the hybrid technique, nor any other technique, was selected as a reference calibration technique for NDACC.
   1. Whiteman, D. N., D. Venable, E. Landulfo, Comments on “Accuracy of Raman lidar water vapor calibration and its applicability to long-term measurements”, Applied Optics Vol. 50, Iss. 15, pp. 2170–2176 (2011)
7. Page 5. The discussion here seems to imply that there was no lidar sub-system for automating the bore-site alignment. It would be helpful for the authors to address the question of how many of their troubles may have been reduced or eliminated if they had had an auto-alignment system that could hold the alignment to within, say, 10-20 microradians. This is the performance that we see in using a variation of the Licel bore site system that is commercially available. Such questions could perhaps be addressed through the authors’ use of the Zemax program.
8. Figure 3. These fiber optic mode patterns are consistent with those presented in the Whiteman et al, 2011 paper cited in #6 above. This reference would be helpful to include here.
9. Figure 4. This figure is quite small and there are various notations in the figure that are not described. The figure could be expanded and much greater detail used in explaining the figure and the optimization procedure. But the authors describe standard optical optimization considerations, the details of which I am not sure are of benefit to the reader. So I suggest that Figure 4 be eliminated and the optimization discussion reduced to speak in more general terms about what was tested and what was finally determined by the optimization.
10. Line 255. The term “adjustment tolerance” is used. Please be more specific in describing what adjustment you are referring to: x,y,x displacement? angular displacement?
11. Sections 4.1, 4.2 and 4.3 all show significant improvements in the calibration stability using the optimized optical configuration. It would be helpful for the authors to clearly state what aspects of the optimized optical design lead to this improvement. From my reading of the paper, this was not clear. Was it the removal of vignetting? Use of a larger portion of the photocathode? Reduced wandering of the spot on the photocathode due to displacements of the input beam?

12. Section 4.3. If I understand the experiment described, the authors translated the device shown in figure 5 along the optical axis to simulate the range dependence of the spot size that impinges on the fiber optic. In an actual lidar implementation, the angular information in the focusing lidar beam will also change as the laser beam propagates away from the telescope. It does not seem that this effect is simulated in the experiment that the authors performed. Please clarify this and comment on whether this has any effect on the results presented.
13. Line 363. The authors refer to “photon detection threshold”. I believe that this is the quantity that is usually expressed as “discrimation level” which should be provided with a unit, which I assume is mV in this case. Did the authors use the standard technique of determining the pulse height distributions of the PMTs to set these levels? If not, how did they do it?
14. Line 364. The authors state that after making the adjustment discussed in #13 that “the detected photon rate matched properly the expected Poisson Law”. More detail is needed here for the reader (at least this reader) to understand what is meant by this statement.
15. Line 366. This is another example of where the presence of an auto-alignment system could have benefited the measurements shown here. At some point in the manuscript, please comment on the feasibility of having such a device in the lidar system and which of the problems would be reduced or eliminated if such a system were in place.
16. Equations (1) and (2). The authors earlier describe the Sherlock technique of convolving the instrument transmission function with the Raman cross section. The authors also state that they are accounting for the temperature dependence of the Raman cross sections in their calculations, but the older versions of the equations given here do not capture those temperature dependent details. Please use equations that account for the variation of Raman cross sections with temperature to agree with your earlier discussions.
17. Discussion between lines 380 – 385. The authors refer to uncertainties of laboratory components (10%), differential transmission (2-5%) and Raman cross sections (10%). As mentioned earlier, Venable et al., achieved calibration agreement between their independent technique and that of corrected radiosondes to better than 5%. Thus, the numbers that the authors use here seem too large. Also, for the purposes of climate studies, it is very important to distinguish between random and systematic uncertainties. Let us assume that an independent calibration is performed using a Raman cross section ratio of X. If at a later date the cross section ratio is found to be Y, earlier data may be reprocessed simply by multiplying by the ratio of X and Y. In other words, systematic uncertainties, once they are discovered and quantified can (and need to) be corrected. Such correction is much more difficult to do when using a dependent calibration technique since issues of atmospheric variability may complicate comparisons. The ability to reprocess data when higher accuracy cross section information becomes available is a strong argument in favor of independent calibration techniques in developing time series of water vapor mixing ratio.
18. Line 394. Reference is made to an *a priori* calibration coefficient. What is this *a priori* calibration?
19. Lines 395 – 400. Discussion is given concerning the comparison of lidar measurements in two ways: 1) comparing a lidar range cell to a point sensor and 2) comparing IPW derived in part from the lidar profile with that of GPS. There are several issues that need discussion in these comparisons including that of the influence of the lidar overlap function on these calculations. The authors should not assume that the overlap functions for the water vapor and nitrogen channels are identical. The degree to which the channel overlap functions cancel in the ratio of H2O and N2 channels needs to be demonstrated within some level of uncertainty. See also #24.
20. Lines 400 – 405. Discussion is given of using the radiosonde profile to extend the lidar profile in order to calculate the total column water vapor. Additional discussion is needed that

quantifies the influence of errors in the radiosonde profile (due either to wet/dry bias, time lag, collocation issues, etc) on the derived total column water. See also #24.

21. Section 5.2. N2 calibration procedure. This procedure seems quite similar to what Vaughan et al implemented and what we discussed in our 1992 paper. If so, please provide those references here. Later in the paper (around line 512) the point is made that dispersion of PMT cathode quantum efficiency is a factor in applying this technique. Those points should be made here when the technique is being introduced. At some point, authors should consider how that dispersion might be affecting their results, such as whether some kind of normalization is needed to make the calibration corrections that are mentioned below.
22. Lines 420-421. Authors state that the slope obtained for the upper layer during Dimevap is not well determined due to low SNR. Then they speculate that “this drift” may be caused by aging optics or deposition of dust. Presumably the authors refer to some other slopes in the table, but not the ones for the upper layer during Dimevap, which are shown to be 0 within the uncertainty. Please reconcile this. Also, some of the slopes shown are as large as 8.7%/month which seems a very large drift. I would find it difficult to believe that aging of components could explain such a change unless, for example, some electronic component was in the process of failing. Was that the case? Regarding the possibility that dust deposition could cause such a drift, I would think that such dust would need to be a brown carbon-like substance that demonstrates considerable dispersion of UV absorption. Were the components subjected to outbreaks of combustion aerosols that could have caused brown carbon to be deposited on the optics? And was that deposition continuous so as to explain the drift observed? In summary regarding Table 2, the large slopes shown for some cases do not seem well explained in the current discussion. Are there other effects that could explain the drifts in addition to aging of components and deposition of dust that absorbs differentially in the UV?
23. Line 424. Column 5 is referred to where, I suspect, column 4 was intended.
24. Section 5.3.
    1. The authors discuss using a PTU sensor, that is earlier stated to be at a height of 15m above the ground, to compare with a range cell of the lidar that is centered at a height of 67.5m. It was no doubt a practical choice to not use lidar data at a height of 15 m above the ground since it is very difficult to get reliable lidar measurements so close to the ground. But this non-collocation is certainly an issue for calibration and needs to be discussed. The authors could present a high temporal resolution time series of comparisons to justify the technique to give the reader confidence that the lidar measurements are consistently reliable at such short range. Our experience is that there is considerable random variation in such near range measurements under the best of circumstances and, if the alignment is permitted to wander, significant systematic uncertainties can be introduced as well. So, I think the authors need to convince the reader that such point comparisons in the near field are valid to be making beyond what is shown in Figure 12. Then they should discuss what systematic differences that might be expected due to the fact that the PTU sensor and lidar are not sampling the same atmosphere. The profile comparison shown in Figure 9 is somewhat difficult to see but would seem to indicate a significant difference between the lidar and PTU values of mixing ratio near the surface. Please also clarify the range resolution of the lidar measurements in this discussion. During the Dimevap campaign, the authors mention using 2 point sensors at different ranges and compare with slant range lidar measurements. Even though it may be described in more detail in earlier publications, some additional description of this technique would be helpful here to contrast with what was done during the Saint-Mande campaign.
    2. Regarding the GPS comparisons, the authors use a composite water vapor profile derived from 3 sources of information (lidar, sonde, model) to compare with GPS. Please quantify,

on average, what fraction of the IPW is sampled by the lidar and what fraction is quantified by these other sources. Then discuss reasonable estimates of the uncertainties of the non-lidar data sources and, from this, provide an estimate of the total uncertainty in IPW due to the use of three data sources for the profile. Since most of the IPW is measured by lidar, it is possible that this technique can provide a useful comparison to GPS for the purposes of lidar calibration, but the burden is on the authors to convince the reader that the technique is a valid way to derive a lidar calibration. The current description is not sufficient to accomplish this task.

25. Line 446. Authors state that the drift in H2O calibration is consistent with the N2 calibration results and that the drift can thus be corrected. But the authors do not describe how they make the correction although they may be using the same technique described in Whiteman et al., 1992. But, since the N2 calibration technique involves measurements in the H2O channel at a different wavelength than for measurements of water vapor, the question of dispersion of quantum efficiency of the photocathode must be addressed. Also, as stated earlier, the authors did not really describe how this technique is employed and they should add such description. But, assuming that there is some QE difference to account for in the water vapor channel, how are the N2 calibration results actually used to adjust the drift in water vapor calibration? It would seem that some normalization of results is needed. If that is the case, how is the normalization done and what assumptions are made in the normalization technique?
26. Figure 12. Upper panel shows a significant difference between the calibration constants derived from GPS and PTU comparisons (perhaps as large as 10%) as well as a consistent drift of both calibrations. Lower panel shows the corrected times series where the drift has been essentially removed but also where now the relative calibration constants are both near 1.0. How has the N2 calibration been used to correct the difference in calibration values shown in the upper panels? This is not clear and would seem to require an assumption about what calibration source is "better" than the other or, at least, that some portion of one time series comparison is "better" than another. Please explain how this was done since it is really not clear at this point.
27. Line 499. Authors state that the main problem of the lidar calibration drift is the thermal expansion of the optical bench. This seems quite unusual an explanation to explain a long term drift and would require that the mean temperature of the lidar system is somehow changing consistently over extended periods. Was that the case? What was the range of temperatures experienced by the lidar system? I can understand how thermal expansion could perhaps explain drift within a single session but do not see how it could result in drift over extended periods of time since presumably the temperature of the lidar system will return to some mean value (determined by the lidar enclosure temperature) between experiments. Is the lidar system open to the atmosphere during measurement sessions and was the external temperature consistently changing in one direction during the experimental campaign? Still do the materials of the optical table have such a large coefficient of thermal expansion to explain the calibration drift seen? Such things can be modeled with the Zemax software. Did the authors perform such a modeling experiment with Zemax that showed the drift could be explained by the range of temperatures experienced by the lidar system?
28. Line 500. Thermalization → Athermalization
29. Lines 511-513. This is where the authors consider the issue of dispersion of PMT QE in the N2 calibration technique. These points need to be made earlier and discussed in much more detail as requested above.