

## Response to reviewer #1:

The authors wish to thank the anonymous reviewer for his/her valuable comments. Corrections prompted by this referee appear in green in the revised manuscript.

*I don't see a real value or potential impact of this paper in its present form in the problem of assessing and mitigating the variability of the calibration stability of water vapor Raman lidar systems. The paper does not contain any new solution proposed by the authors to address the calibration stability issue in water vapor Raman lidars. Considered solutions, most of which already proposed by other authors in previous papers, are aimed to achieve a higher stability of the lidar performances, with a consequent improvement also of the calibration stability. The topic treated in the paper is quite marginal as it reports few technical solutions, already reported in previous papers, which have been implemented by the authors also in their Raman lidar system to address the stability issues related to their specific system setup and the related implications in terms of the calibration stability.*

>> We think that the presentation and potential impact of our work has been improved thanks to the numerous positive comments from both reviewers. The aim of this paper was indeed twofold: 1) understand the origin of instabilities of our system and 2) improve the stability using various technical solutions (instrumental modifications and numerical corrections). Two sources of optical instability were identified which have a dramatic impact on the calibration stability of our system: the PMT photocathode non-uniformity and the beam structure fluctuations at the exit of the optical fiber. Though this kind of instabilities were already discussed in previous papers, we think that the value of our work is to present laboratory tests which allowed isolating and quantifying both effects. These tests helped us to discover that we also had vignetting problems which magnified these effects. In a second part, we show that N2 calibrations can get rid of the long-term drifts (over several months) in the calibration coefficients due to cumulated fluctuations of beam wandering and interventions on the system. Though the N2 calibration has been proposed some time ago by Vaughan et al., 1988 and Whiteman et al., 1992, it seems that it is not common nowadays as most lidars use now calibration lamps. We believe and show here that this technique is indeed effective in detecting changes in the detection system. Our results suggest that it may be useful to include this technique as an additional calibration means to operational systems. Further redundancy can be achieved from PTU and ZWD calibrations as also shown in this study (which moreover are two independent means of absolute WVMR calibration).

*This paper is written in an ineffective way. The real benefit for the lidar community of the technical results contained in it are poorly emphasized and consequently do not appear clear to the reader (see more comments below). The authors need to reformulate the abstract, the introduction and reshuffle the text of the paper in a way to more strikingly express the value and the potential impact of their scientific effort.*

>> We think that this objective is achieved in the revised version of the manuscript.

*Authors describe a system (figure 2) characterized by a 10-15 % variability of the calibration factor within a time interval of 5-10 min in signal break conditions and by a 30 % variability in less than one hour in signal fading conditions. Besides the described effort and suggested solutions in terms of small changes in the optical layout of the receiver, which are aimed to mitigate the system instabilities and partially reduce the variability of the calibration factor, my personal impression is that the system requires a substantial improvement in the receiver optical layout, making it more compact and possibly relying on a different coupling between the telescope and the spectral separation section of the receiver.*

>> Figure 2 shows measurements from the DEMEVAP campaign before the system was modified. Indeed, the break and fading illustrated with this figure highlight the instability in the signals (not in the calibration factor) which are explained by the two errors sources mentioned above and the vignetting problem at that time. The solutions suggested by the reviewer to make the system substantially more stable are indeed those that we mention in the conclusions. In the meantime, we show in this manuscript that a stability at the 3% level can be achieved with the present (optimized) system which is already a nice improvement.

*I think that in the present form the manuscript should be rejected. However, I realize and appreciate the amount of dedicated work the authors carried out and included in the manuscript and I would like to give the authors the opportunity to sensitively improve the paper and encourage them to resubmit a revised version of the paper. I am available to review again the paper after major revisions. I suggest the authors to re-shuffle the paper in a way to have a paper with a different focus, i.e. the description of the recent upgrades of the IGN Raman lidar, illustrating all changes and improvements implemented in the receiver and finally demonstrating the achieved improved performances of the system in terms of precision, vertical extent and finally also calibration stability. Results in terms of testing different fibers and optical paths, assessing the impact of the spatial non-uniformity of the PMT photocathode and the spot movement on fiber input, etc., deserve to be published, but in a paper where it is clear that the implemented upgrades are aimed at improving the system performances and reducing the optical instability of the receiver, while keeping the original suboptimal configuration of the receiver, with the telescope and the spectral separation section of the receiver completely separated. I also suggest a careful and dedicated reading and improvement of the English language by one of the senior scientists co-authoring the paper to improve the readability and effectiveness of the paper.*

>> We reworked the text following the specific comments given by both reviewers and believe that the manuscript reflects now better the implemented upgrades and the achieved improvements of our system performances in terms of calibration stability.

### **Answers to “Additional major points” section**

*The title reads: “Study and mitigation of calibration error sources in a water vapor Raman lidar”; however, the paper more specifically deals with the assessment of calibration factor instability and its sources and not on the calibration error sources. Besides what already suggested above, which would also ultimately lead to a substantial change in the title, in the present stage a more appropriate title of the paper would have been: “Study and mitigation of calibration factor instabilities in a water vapor Raman lidar”.*

>> We agree with the reviewer and changed the title of the article as suggested.

*In order to illustrate the motivations behind this research effort, authors specify in the Abstract that: “. . . such drifts are incompatible with both the long term stability required for applications such as climatology and the absolute accuracy needed for wet path delay correction of GNSS signals.” However, nowhere in the text of the paper they explicitly face this important issue of specifying what are the stability requirements for both climate applications and to provide the absolute accuracy needed for wet path delay correction of GNSS signals.*

>> We added in the introduction the following sentence: “According to the GCOS-112 report, the accuracy requirements for WV monitoring in the troposphere are: 2% precision, 2 % absolute accuracy and 1% stability (GCOS-112, 2007).”

*Authors write in the Abstract that: “In order to validate... the new procedure, measurements...”. However, there is no new procedure here as in fact, as also specified by the authors, the procedure and solutions reported here had been already illustrated and discussed in previous literature papers. This should be more clearly emphasized.*

>> We agree with the reviewer and we deleted a part of the sentence.

*Introduction, line 41, authors write that high accuracy and stability data “. . . necessitate a careful and continuous calibration of the system.” While a careful calibration is certainly required, a continuous calibration is indeed a peculiarity and drawback of optically unstable systems as the one illustrated in this paper. As already specified above, the objective of the paper should be the description of the upgrades implemented in the receiver (optimization of fiber and optical path, reduction of the effects associated with the spatial non-uniformity of the PMT photocathode and the spot movement on fiber input, etc.) to make it more stable and consequently reduce the short- and long-term drifts of the calibration coefficient. A continuous calibration is a very demanding requirement which is certainly not required in other Raman lidars with a more compact and better designed receiver.*

>> We agree with the reviewer that the term “continuous” is misleading. We replaced it with “periodic”.

*Introduction, line 46 and following. Here authors refer to “an independent and a dependent approach”. Dependent or independent on what? This is not clear in the text. You probably refer to a “dependency” on a measurement from an external reference water vapor sensor, but this is certainly not clearly indicated in the text. Additionally the terms “independent approach” and “dependent approach” were not those used by authors who introduced these Raman lidar calibration approaches, even those explicitly cited by the authors (Ferrare et al. 1995; Leblanc et al., 2011; etc., by the way Leblanc et al. became a AMT paper in 2012). This part should be re-written more clearly.*

>> We agree with the reviewer and we completed the sentence as “independent or dependent on measurements from an external reference water vapor sensor”. We also removed the term “approaches” and replaced it by “methods” (consistent with Sherlock et al., 1999).

*Introduction, line 61 and following. Here the authors write: “The so-called dependent calibration method is thus performed by calculating a normalization factor from the comparison of the “lidar profile” with either a radiosounding (Ferrare et al., 1995; Leblanc et al., 2011), or integrated water vapor (IWV) measurements from GPS, or microwave radiometers (Turner and Goldsmith, 1999), or ground based humidity sensor data (Revercomb et al., 2003).” This sentence is very generic and unclear as in fact it does not allow the reader to understand what the authors mean for “lidar profile”: is this the power ratio of the H<sub>2</sub>O over the N<sub>2</sub> Raman signals ? Or does this include any normalization term ? Usually authors distinguish a height-dependent and a height-independent component of the signal power ratio for the purposes of the calibration. This should be clearly specified already at this stage of the paper.*

>> What we meant here is a height-independent component; we added it in the text.

*Introduction, line 68 and following. Here the authors write: “. . . that is 5% for radiosonde data, 2% for capacitive sensors, and 2-5% for GPS IWV data.” Radiosonde is not a humidity sensor itself. A variety of different humidity sensors are considered in the different radiosonde packages used by meteorological services around the world. The capacitive sensor is used in specific radiosonde packages. Please, re-write this more clearly so that the reader can understand what you want to mean when you distinguish between radiosonde and capacitive sensor.*

>> Sentence rewritten as: “... that is 5% for the best operationally used radiosondes, 1-2% for ground-based humidity sensors,...”

*Introduction, line 181 and following. Here authors specify that both collimating lens L1 and focusing lens L2, L3, and L4 have a focal length of 46 mm. However, unless a achromatic doublet is used, the focal length is defined at a specific wavelength. At what wavelength is the focal length of 46 mm defined ? What are the focal lengths of the used lenses at 354.7, 386.7 and 407.6 nm ? How is the difference between these focal lengths and the value of 46 mm affecting the collimating and focusing properties of these lenses ? Certainly, the light coming out of L1 at 354.7, 386.7 and 407.6 nm is not be collimated if the focal length is 46 mm at a different wavelength; additionally, the uncollimated light coming out of L1 at these three wavelengths will not be properly focused on the corresponding PMTs. In this respect, in line 237 and following, authors specify that: “For each lens and fiber combination, the distance between the source –i.e. fiber output– and L1 is optimized with the software (ZEMAX) to obtain a collimated beam throughout the optical detector system”. However, it is not clear at what wavelength this software optimization was performed. Furthermore, un-collimated light beams passing through the interference filters determines the incidence angles of these beams on the filters to be different from 0 (which is the incidence angle assumed in the text by the authors), with consequent drifts of the interference filters’ center wavelength and changes in signal strength. All the above aspects have to be addressed and properly described in the paper.*

>> The focal length of 46 mm is as specified by the manufacturer (in the visible spectrum). The focal lengths of these lenses at 354.7, 386.7, and 407.6 nm, are: 46.45, 46.92, 47.18 mm. We checked with Zemax simulations that the impact of the different wavelengths is negligible on the spot size on the PMTs. Moreover in practice we use a more empirical adjustment with a blue LED at 468 nm and adjust the spot size visually to spread on the photocathode with a diameter of 6 mm (for a PMT aperture of 8 mm). At the LED wavelength, the focal length of L1 is 47.73 mm. At the actual wavelengths (386.7 and 407.6 nm), the focal lengths are smaller, thus the spot size will be slightly smaller. This was checked with optical simulations: with a spot diameter on the PMT of 6.18 mm @468nm, we get 5.10 mm @407.6nm and 4.58 @386.7nm.

>> Regarding the interference filters, the maximal incidence angle on the filter is about 10 mm / 457 mm (radius of L1 / distance between L1-L2), that is about 1.4° resulting in a displacement of central wavelength of about 0.02 nm, which is negligible compared to the FWHM of 0.38 nm.

*Section 3 of the paper (Optical Optimization) is primarily dedicated to the optimization of the optical layout of the receiver to eliminate the “vignetting” effect, with no reference to any other aspect of the optical system layout requiring optimization. Please, specify this better. If you don't report any other aspect of the optical system layout requiring optimization, change the section title to make it more specific to its effecting content, i.e. the removal of the “vignetting” effect.*

>> We agree with the reviewer that the title is misleading to the reader. We changed it into “Correction of the vignetting problem in the detection system”.

*Section 3. At what wavelengths was the ray tracing analysis performed? I understand that this was performed only for the nitrogen channel optical path, so I imagine that the optimization was carried out at 386.7 nm. I would imagine that a different optimization would pertain to the other wavelengths involved in the system. Is this correct? How is this accounted for?*

>> The simulations have been realized at 386.7 nm, and as we explained above, the collimation differences are not critical.

*Section 3. The description of the simulation in the text refers to lenses L1 and L3, while the graphical representation in figure 4 refers to L1 and L2, so there is an inconsistency. What was the pair of lenses involved in the optimization computations: L1 and L2 or L1 and L3?*

>> Figure modified.

*Section 3. Close to the end of this section the authors state that: “The focal lengths of the lens suggested here are somewhat arbitrary, and other configurations that meet the non vignetting condition are possible”. Please, explain better what you mean here.*

>> Sentence changed to: “The lens parameters were taken from existing and available standard products.”

*Figure 9 shows the only example in the paper of the WVMR profile derived from the Raman lidar signals against a radiosonde profile. The text associated to the figure in the manuscript and the figure caption do not specify if the measurement was collected in daytime or nighttime and what are the integration time and vertical resolution of the lidar data. Is the time integration 5 min and the vertical resolution 7.5 m as specified above in lines 372 and 370, respectively?*

>> Date, hour and place of the measurements added in the caption. Details about the integration time and vertical resolution were added in the text associated with the caption.

Section 5.2, line 408. Authors specify that: “. . . these measurements will be called "N2 Calibrations" hereafter". However, the term N2 Calibration was already used to indicate the same approach by Whiteman et al. (1992). Probably the authors should cite or refer to this previous paper for what concerns the N2 calibration approach.

>> Reference added.

Concerning Figure 12 in Section 5.3. Figure legend specifies that the bottom panel represents the normalized calibration coefficients after the correction of the instrumental drift represented by the variations of the "N2 calibration" coefficients for the 350-450 m. This sentence is not clear: what do you mean for “. . . represented by the variations of the "N2 calibration" coefficients for the 350-450 m" ? Do you mean the calibration coefficients obtained after the correction of the instrumental drift based on the application of the N2 calibration approach ? Furthermore, in the text of the manuscript the bottom panel of figure 12 is indicated to represent the normalized calibration coefficient, relative to the mean over the period, which is a different quantity with respect to what indicated in the figure legend. If the correct meaning of the quantity is the one indicated in the figure legend (in the way I interpret it), I am not surprised that the application of the N2 calibration approach leads to a sensitive reduction (down to a value of 2-3% per month) of the drift in the H2O calibration coefficients. A similar result had been obtained by Whiteman et al. (1992), when these authors introduced the N2 calibration approach.

>> Sentence in the caption changed to “...the instrumental drift based on the application of the N2 approach for the 350-450 m layer.” Concerning the normalized coefficient, we agree its definition was not clear, and we reformulated and completed the sentence in: “These coefficients have been corrected by linear function of time fitted from the N2 calibration results. We did not use the night to night coefficient correction of Whiteman because our N2 calibration coefficients are noisy and we did not want to add dispersion to the H2O calibration coefficients.”

Section 5.3, line 446. Here you write: “The important result here is to notice that the drift in the H2O calibration coefficient is very consistent with the N2 calibration results”. I am not sure I understand what you mean here with this sentence.

>> We deleted the sentence as it was a repetition of the previous sentence (that both calibration methods achieved similar drifts of 2-3% per month).

Section 5.3, line 448. Here you write: “We chose the 350-450 m layer for the correction because of its better stability”. Why a better stability is achieved considering the 350-450 m layer? The lidar performances look quite stable up to much higher levels. What is the motivation behind?

>> We added: “Regarding the N2 calibration, we chose the 350-450 m layer for the correction because of its higher SNR and smaller RMSE”.

### **Answers to “Specific minor points” section**

All the specific minor points have been corrected in the manuscript.