Reply to anonymous referee #2:

First of all, the authors thank referee#2 for his/her valuable comments and suggestions. A thorough revised writing has been conducted and further analysis conducted to end on this revised manuscript. Then, the manuscript is strongly different than the original version. The English has been revised by one of our collaborator who is a native English speaker and a specialist in hygrometry.
Please to find our response to your comments below :

GENERAL COMMENTS

Referee#2 suggested that temperature measurements from Pico-Light H_2O used in the comparison come from a radiosonde integrated in the instrument package. This comment has suggested that the instrument description was not as clear as it should have been. Therefore we have thoroughly improved the description in consequence. The temperature sondes are Sippican fast response thermistors. I believe this is what referee#2 meant by "radiosondes". We have to specify that water vapor mixing ratio measurements do not come from radiosonde measurements. The iMet-4 sonde onboard Pico-Light is only used for : 1- have backup localization and temporal information (e.g. balloon trajectory), 2- to compare if needed, humidity in the lower troposphere, in case of failure or questionable data.

"My first concern is that the temperature comparisons add little to the understanding of the performance of the Sippican VIZ sonde, certainly not without comparison against another in situ sonde temperature."

Reply: We have included temperature comparison since it is one of the input parameter for the spectra processing procedure. In spectroscopy, temperature plays a role in the line area (having therefore a direct impact on retrieved mixing ratio) and in the line width. Indeed, we have added comparison with in-situ RS 41 sonde launched by CNES 1h15 later after Pico-Light on February 19, 2019. Differences with RS41 temperature measurements are within the uncertainty of RS41 in the lower stratosphere. Considering temperature data from ground to 20 hPa, the mean bias is similar to biases between radiosondes which have been reported in the frame of WMO intercomparison campaigns (i.e.: 0.5K). Having similar results with WMO is one check point to ensure the quality of measurements.
The two temperature sondes used onboard Pico-Light, are located at each ends of the optical cell (top and bottom, see new figure 1), and are VIZ NTC thermistors from Sippican. Although supposed less accurate than platinium sondes used onboard Vaisala sondes, they are more sensitive to changes in ambient temperature and the bias with Vaisala RS41 sonde reported here allows to be confident in temperature measurements provided by the Sippican sondes used.

"With regard to the water vapor mixing ratio measurements, the MLS v4 water vapor comparisons suggest that the Pico-Light H2O hygrometer in these two flights is comparable to other in situ instruments. The MLS v5 comparison is of some interest, but the two flights present do not present the same story in the below-100 hPa region where v4 is understood to have an instrumental bias.
More significantly, it goes without saying that two flights – and two flights in significantly different meteorological settings - are a very slim basis upon which to make a judgement of the performance of an instrument measuring any atmospheric trace constituent, and water vapor with its strong vertical gradients particularly so. Thus I don’t see great value in the profiles presented; what would be of

considerably greater interest would be head-to-head intercomparison with a reference-quality in situ hygrometer."

Reply: Head to head comparison is scheduled in 2021-2022 from the CNES Aire-sur-l'Adour balloon facility. In this frame, we will compare in-situ measurements from FPH NOAA, FLASH-B, the micro-hygromètre (from LPC2E, CNRS, France) and our Pico-Light. In between, limiting the analysis to the cases where MLS and Pico-Light were sounding the same airmasses, the comparison with MLS v4.2 in the altitude range from 20 to 147 hPa allows an indirect connection between Pico-Light and other existing hygrometers, such as FPH NOAA or CFH. The manuscript has been revised in this direction. As an example, Yan et al., 2016, where NOAA FPH has been compared to MLS v4.2, has reported bias in in-situ water vapor measurements of 11% in average between 68 and 146 hPa. In our case, the bias in the same altitude range is of 10.8%, similar to Yan et al., 2016. The similarity in the bias obtained against the same version of MLS is an indication of the performances of Pico-Light and represents encouraging results. We agree that, in the troposphere, the strong variability in water vapor content impede any valuable conclusion. However, Pico-Light is primarily designed to probe the UTLS and lower stratosphere. For this reason, tropospheric measurements are not of primary interest.

SPECIFIC COMMENTS:

The manuscript has been re-organized and the description of the instrument itself has been expended and detailled compared to the original manuscript. The section Pico-Light H_2O has been revised in a cleaner manner. The data processing has been detailed whereas was absent in the former version of the document.

Details about how are used Sippican temperature measurements are given: only the coldest temperature is used since it suggests that it is less impacted by solar radiation. Additionally, temperature corrections are applied coming from the Sippican WMO testings.

As expressed earlier, the iMet-4 sonde is only used as a backup in case of failure in GPS measurements.

The pressure sensor is a Honeywell PPT1 and not PPT2 (mistake from our side). The accuracy is given at ± 0.05% for PPT1 and ± 0.0375% for PPT2. PPT1 absolute uncertainty is 0.5 hPa. A section dedicated to "uncertainty" (section 2.6) has been added in the manuscript which addresses the impact of physical parameters (environmental) : pressure and temperature, but also of other parameters to the measurement uncertainty. It has been demonstrated that the uncertainty due to combined pressure and temperature induces an error of 0.3% at maximum. The largest source of uncertainty being the spectra quality.

B. "Method for intercomparison with Aura-MLS retrievals"

The high resolution profile is linearly interpolated in the pressure log space to provide a low-resolution profile in the MLS pressure log space. Then, at each MLS pressure level, is associated one value of the mixing ratio from Pico-Light which comes from the linear interpolation and therefore an error bar is estimated based on the deviation of the 100 Hz values compared to the interpolated value summed in quadrature with other sources of errors. The low-resolution profile is the input to the averaging kernels.

C. “Results and discussion - Temperature” (lines 191-224)

Over 1 milliseconds, 20 measurements of temperature are performed and outliers removed to calculate the average temperature which is stored onboard. As for pressure, doing this way allows to improve the precision of the temperature measurements by a factor of $\sqrt{N}$, N being the number of data points used. This procedure is used for both temperature sondes. At the end, during the spectra processing, from the two averaged temperatures, the coldest one is chosen as physical input parameter to the fitting procedure.

D. “Results and discussion – Water vapor ” (lines 225-279)

This section has been thoroughly revised. Analysis using the GRUAN consistency definition is performed and MERRA 2 ozone 3-hourly reanalysis are used to help the interpretation of observed biases. This analysis helped in restricting comparison with MLS on pressure levels for which both instruments were sounding the same airmasses. This has allowed a new estimate of the biases and therefore modified our conclusions in a better perspective. Processing this way allows to demonstrate that the mean bias in the lower stratosphere is strongly similar to the one reported between the CFH (Cryogenic Frost Point hygrometer, Vömel et al., 2007) and MLS v4.2 which is even better that previously estimated. This is encouraging.

About MLS v5, we have observed that MLS v5 retrieval are systematically dryer than MLS v4.2. The results reported here are the first one for this version of MLS even though only 2 profiles are compared.

E. “February 19th flight ” (lines 280-308)

This section has been removed. In place, a discussion around the GRUAN consistency analysis supported by MERRA-2 ozone reanalysis has been added and allows to explain the observed discrepancies on some of the pressure levels.