

Reply to anonymous referee #3:

First of all, the authors thank referee#3 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 :

MAJOR COMMENTS:

Following reviewer's comments, discussion on error sources has been thoroughly revised and completed. The expended discussion can be found in the section **2.6. Uncertainties**. If the original discussion about the filament has been removed, a new discussion around the filament has been added based on GRUAN consistency which has been added to the analysis. In this frame, the contribution of the filament to the analysis has brought values to the understanding of the observed discrepancies.

DETAILED COMMENTS:

-**From lines 17 to 90:** the introduction has been completely revised.

-**Line 115:** the discussion about line shape effect can be found in the section **2.5 Spectroscopy**. It has been extended to the troposphere. It is shown that, due to the spectra signal to noise ratio and possibly to the weakness of line parameters, even at tropospheric pressures, no high order line shapes are observed. The first author has an extended experience in the study of line shape parameters at low temperatures for atmospheric applications (e.g. OCO-2, MERLIN space mission and Pico-SDLAs: Ghysels et al., 2011, 2013a, 2013b, 2017, Delahaye, 2019). An additional figure is used to illustrate the residuals from the fitting of tropospheric spectra supporting the affirmation (figure 6, see below).

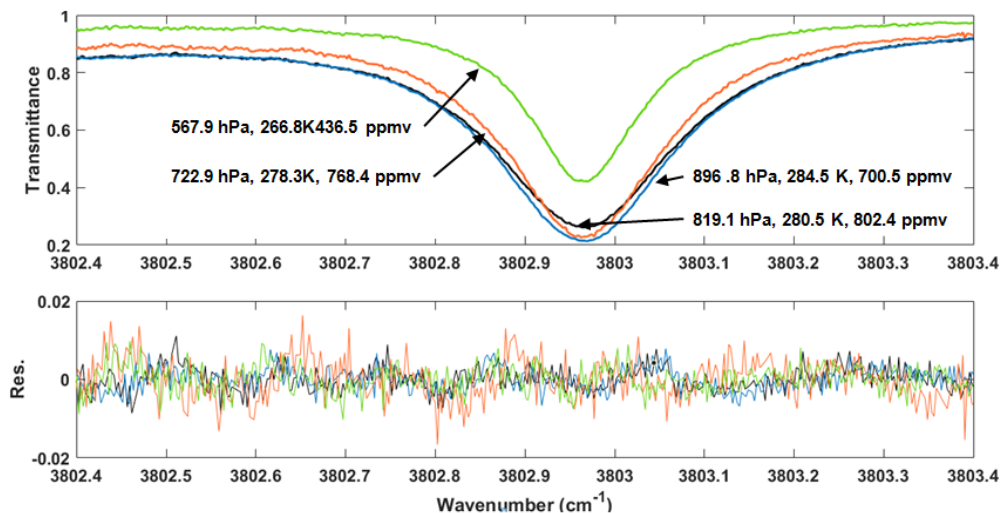
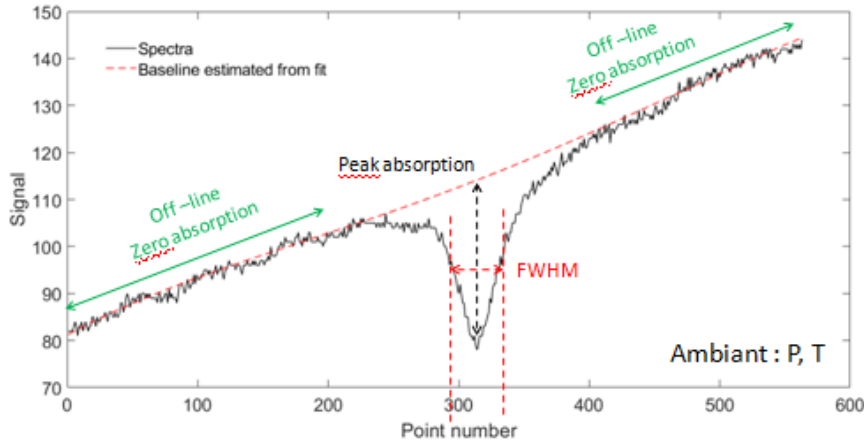


Figure 6 above shows examples of 4 atmospheric spectra of the the $4_{13} \leftarrow 4_{14}$ line of water vapor in the lower troposphere (highest pressure, where speed-dependence of line width becomes important) at 896.8, 819.1, 722.9 and 567.9 hPa. One can see that residuals using Voigt profile are flat (no sign of non-

Voigt effects). For these spectra, the signal-to-noise ratio scaled from 1200 to 1600. About speed dependence of line width, residuals can be seen while SNRs are close to 10000 or above, which is not the case here.

-Line 121: the balloon is bursting once the maximum altitude is reached. Authors have corrected sentences accordingly.

-Lines 140: an expanded description of the spectra fitting procedure is given in section **2.5 Spectroscopy** (paragraph from line 259 to 282) and figure 7 has been added to illustrate some of the main features (see figure below).



It has to be noted that the differential spectra is only used onboard as a tool to determine the peak position. This procedure is used to compensate for any spectral drift by adjusting the temperature of the laser semiconductor, therefore, it has not any involvement in the retrieval procedure. To calculate this differential spectrum, the electrical ramp is used balanced with specific gain.

- Line 149: After a careful checking, one spectrum acquisition is faster than expected, it is 8 ms instead of 20ms. In reality the measurement cycle (of a duration of 1 second) is such that: the first 200 ms are devoted to the acquisition of 5 elementary spectra and the other 800 ms are devoted to the acquisition of pressure, temperature and other technical parameters which are used for diagnostic. The pressure is averaged over half a second and the temperature is acquired within 1 millisecond during which 20 measurements are taken, filtered to remove outliers and averaged.

The pressure and temperature mean values are stored on the SSD disk and will be the input of the spectra processing.

As requested, we have added quantitative information on the impact of pressure and temperature measurements on the spectra processing in section **2.6. Uncertainty** (line 286-319). Table 2 has been revised so it is easy to find Pico-Light water vapor total uncertainty and its vertical resolution corresponding to MLS pressure level (see new table below).

Table 2. The relative uncertainties u of measurements of temperature T and mixing ratio X made by Aura MLS and Pico-Light H_2O . Also shown is the resolution $\delta(z)$ of the height z .

P [hPa]	AuraMLS			Pico-Light H_2O		
	$\delta(z)$ [km]	$u(T)$ [K]	$u(X)$ [%]	$\delta(z)$ [m]	$u(T)$ [K]	$u(X)$ [%]
22	3.2x265		5	25		8.5

31	3.6x165	0.6			0.2
46	3.2x230		5	20	6.4
68	3.1x190		5	15	5.2
83	3.1x190		7	13	4.3
100	3.0x198 4.7x165	0.7	15	12	0.2 3.8
121	2.6x193		20	8	3.3
147	2.3x188		20	8	2.6
178	1.7x183		25	7	1.2
215	1.6x178 4.2x165	0.8	40	7	0.2 1.1
261	1.4x173 4.2x167	0.7	35	7	1.6
316	1.3x168		65	7	1.2

The relative uncertainty induced by errors on pressure and temperature, as well as from other parameters like the baseline, frequency axis estimations, have been calculated using synthetic spectra reproducing real atmospheric spectra in terms of baseline variability and instrumental noise and with the introduction of virtual pressure and temperature errors. The uncertainty has been obtained from the deviation of the retrieval to the real value.

About the temporal resolution: Several aspects have to be considered:

1- We are limited at some point with the cycle duration. To have an autonomous instrument, the embedded software has to deal with different technical parameters and it takes some time to acquire all the necessary parameters for: 1- in real time, make sure the instrument will behave properly and 2- after flight, having essential diagnostic parameters saved so it is easy to figure out the reason in case of failure (minor or major). In fact, it takes 800 ms to manage everything in-flight.

2- For temperature, it is easy to acquire a sufficient amount of data within 8ms (here we use only 1 ms for 20 measurements) and having a temperature measurement which is improved in terms of precision. About the pressure sensor, Honeywell ppt1 has a resolution of at worst 0.1% FS (here 0.1 hPa at worst and 0.01 hPa at best). During the descent, the pressure variation over half a second is of 0.05 hPa near burst altitude up to 0.2 hPa at 800 hPa. But since the absolute pressure accuracy is of about 0.5 hPa, we consider that averaging pressure measurement over half a second won't introduce significant bias to the pressure measurement but will further improve the measurement precision.

However, if requested, it is easy to interpolate pressure and temperature at the same frequency rate as spectra.

-Line 151 : yes, measurements are taken on parachute descent. As requested we have added figure 4 illustrating the flight profile (altitude and fall speed).

-Line 162: the description about how air temperature measurements are obtained has been expended an improved (line 184-188). The temperature of concern on line 162 (old manuscript) is the temperature of the enclosure of laser diode. Indeed, the emission center frequency is sensitive to the surrounding environment temperature. To limit the emission frequency drift, the enclosure of the laser diode has to be maintained within an acceptable range of temperature and it is this correction which is mentioned here. However, for Pico-Light, the frequency detuning is set quite large to ensure that, in case of failure in this enclosure temperature regulation, the absorption line would remain within the scan. Anyhow, the

temperature correction does not have any impact on the retrieval since one spectra is recorded within 8 ms, which is too quick to integrate the frequency shift of the line (no line distortion). Additionally, the design of the thermal enclosure is such that only a small correction of temperature is needed.

-Line 178: no remote control means that the instrument has no TM/TC onboard. All data are stored onboard a SSD card and therefore, the instrument recovery is needed to obtain the data, unlike for former Pico-SDLAs.

-Line 192: Imet-4 is flown as piggy back only as a backup in case of failure of the GNSS system onboard Pico-Light. In this case, latitude/longitude/time data from Imet-4 are used. Pressure and temperature measurements would be only used if necessary in case of failure, at the cost of accuracy. RH measurements from Imet sonde can be used to compare humidity measurement in the lower troposphere if needed but are not used here anymore. Indeed, the comparison with imet RH in the lower troposphere has been removed since it was of less importance comparing to the GRUAN consistency discussion.

-Line 193: the absolute uncertainty has been added. In this case it is 0.5 hPa and its impact in the uncertainty budget has been discussed in section **2.6. Uncertainty** (line 302-306).

-Line 206: yes, the sampling rate of Honeywell PPT1 can be as high as 120 samples/sec. Pressure measurements are averaged by the sensor based on the desired averaging time (here half a second). The reasons for the selection of averaging time are detailed for the discussion about line 149 above.

-Line 206: right, I was considering the physical acquisition time.

"In the previous section you mentioned that the spectra are measured only for 200 ms, followed by temperature and pressure measurement and processing. If that is the case, then the uncertainty over 200 ms should be the same as over 1 s."

To clarify, in the text we have added the following sentences:

"The mixing ratio standard deviation in the stratosphere, and therefore, the precision, is of about 277 ppbv while using **unitary** spectra (no averaging) which corresponds to a precision of 130 ppbv for a 1 second integration time (co-addition of 5 spectra over 200 ms)."

In practice, due to the large amount of measurements, it is additionally possible to filter the high resolution profile through moving windows of 20 data points without altering vertical structures (to smooth the vertical profile). Doing this allows to further virtually improve the precision.

-Line 210: this has been done, as specified in response to the comments on line 151.

Line 262: in the old version of the manuscript, the MODIS water vapor maps were only used to check whether Pico-Light and MLS were sounding the same airmass (i.e. ozone-enriched for polar air mass). In the revised analysis and manuscript we have used MERRA 2 ozone fields instead to do the same job since ozone is a dynamic tracer. The MODIS contribution has subsequently been removed.

-Section 6.1 : as specified earlier, temperature is measured using Sippican NTCs onboard during the descent of the balloon. Indeed, in the comparison which is now included in the manuscript, is based on descent measurements from the Sippican NTCs onboard Pico-Light and ascent measurements of the platinum sondes onboard RS41. Of course, sensors are coated to minimize radiation effects and radiation correction is applied, leading to mean bias of 0.12 K in the stratosphere and 0.56 K over the full

altitude range. This is in agreement with comparisons between radiosondes in the frame of WMO campaigns.

-Line 296 : this paragraph has been deleted since it does not bring any value to the new results.

-Line 322 : additional details about MLS are now found in the section **4. Description of used datasets** and the selection criteria is defined in section **5. Method for inter-comparison with Aura-MLS retrievals and selection criteria** (lines 377-385). The corresponding paragraph is :

" The exclusion of improper MLS profiles was guided by two metrics output for each MLS profile: the "quality" and "status" criteria. The threshold and meaning of each criterion are given in the v4.2 and v5 data quality documents. The "status" criterion is a 32-bit integer containing several flag bits. The value of this criterion allows the user to know whether the profile is questionable and if so, the underlying reason. An odd value of this criterion means that the profile should not be used. The "quality" criterion acts as a threshold for scientific use. It is based on fits achieved by the Level 2 algorithms to the relevant radiances. Larger values of "quality" indicate better radiance fits and therefore more trustworthy data. The "quality" threshold for water vapor was set at 0.7. For both flights the "quality" criterion for each of the two MLS profiles used was above 1.87."

-Line 329-381 : the section 6 has been completely revised. The datasets have been reprocessed using new spectroscopy leading to updated comparison in section **6.2. Water vapor**. Additionally, the analysis of the bias between MLS v4.2 and v5 have been reinforced based on GRUAN consistency criteria and MERRA 2 ozone 3-hourly reanalysis. In this section, we have shown that MLS v5 retrievals are dryer than MLS v4.2 (which then are less consistent with Pico-Light in general). While MLS v4.2 absolute values of mixing ratios are more consistent with Pico-Light (following GRUAN approach), the discrepancies of MLS v5 with Pico-Light are found more logical when considering meteorology (in relation with ozone-enriched air from polar latitude observed from the MERRA 2 products). Details about this are found from line 467 to 506.

Corrections about "levels" have been applied.

-Line 347 : right, we have corrected the mistake.

-Line 354: see reply on line 329-381.

-Line 381 - 400: the section 6.3 has been removed accordingly since it did not bring any interest anymore.

- Line 430-433 : corrections have been applied. Done.

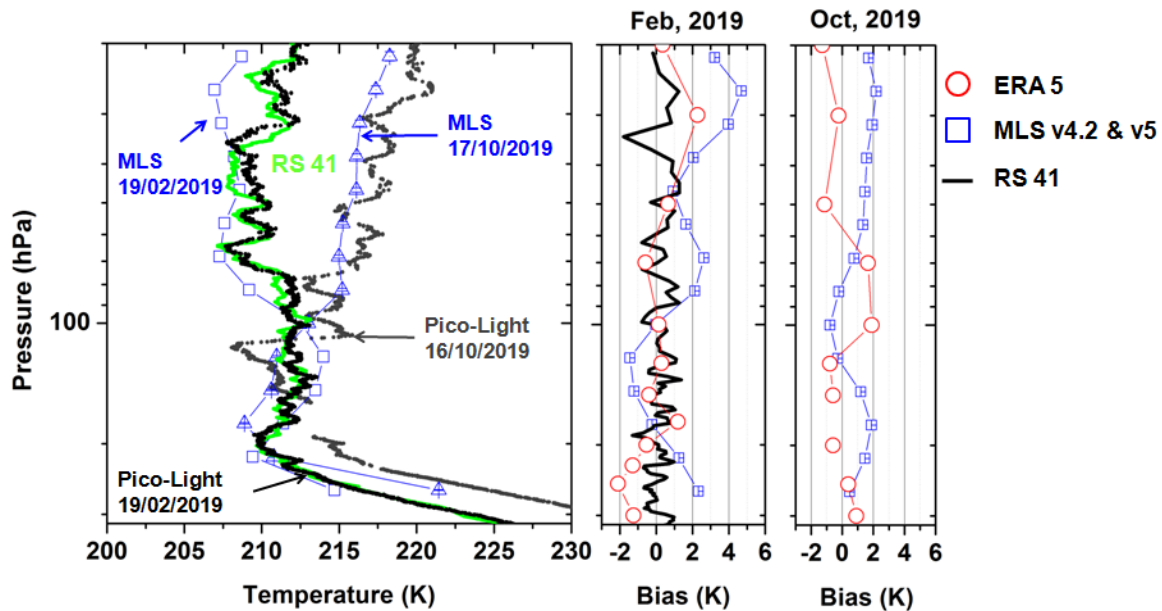
- Line 432 : Done.

- Line 430 : The conclusion has been completely revised based on new results.

In general, figures have been almost completely replaced. Anyway, on the old figure 3 and now, the new figure 5 and 6, the spectra are unitary, meaning that no averaging has been performed.

Figure 1 has been slightly improved so it is easier to locate each element. The laser diode is located at the bottom of the cell and photodiode at the top to minimize the impact of ambient infrared radiation onto the atmospheric signal

The figure showing the comparison with ambient temperature from VIZ Sippican sonde and other datasets (e.g. MLS, ERA and RS41) is now figure 9. In this figure, for clarity, only one profile of MLS (v5) is shown since temperature profiles from both versions are almost identical. ERA 5 profiles don't appear on the vertical profile panel but bias is still shown in the right panels. The bias original panel has been split into 2 panels, one for each flight. The bias with RS41 is shown.



The comparison of water vapor profiles is now shown in figures 11 and 12 (see below).

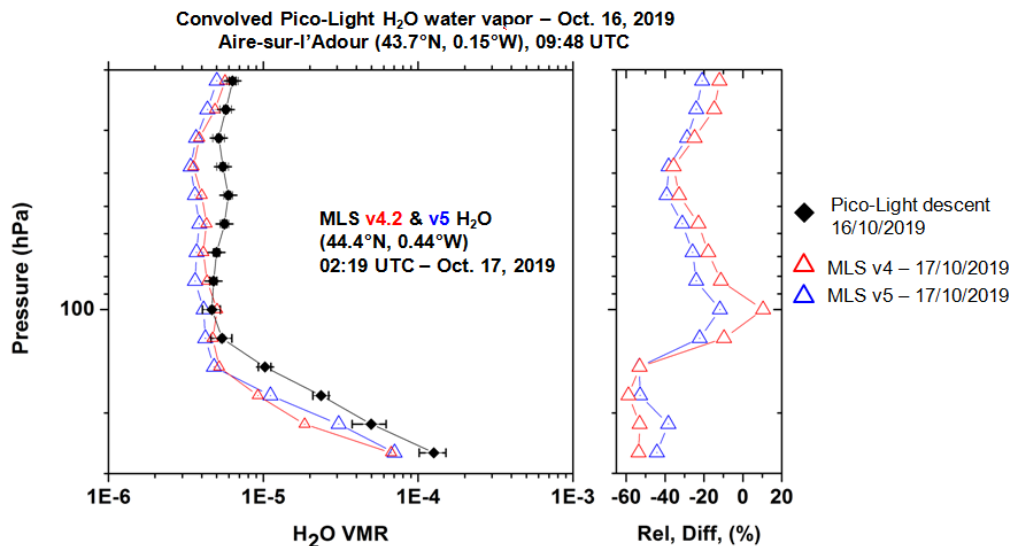


Figure 11: Convolved vertical profiles from the Pico-Light H₂O measurements (black diamonds) compared with MLS v4.2 (red open squares) and v5 (blue open squares) between 20 and 316 hPa on February 19, 2019. The right panel shows relative difference per pressure level between Pico-Light H₂O and MLS datasets.

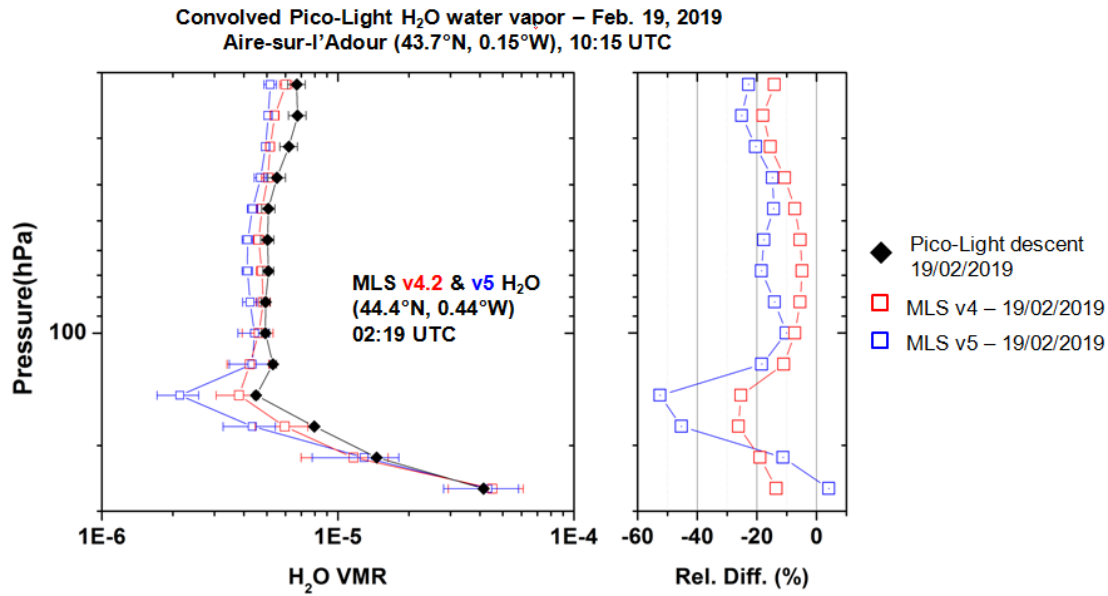


Figure 12:

Same as figure 11 but for October 16, 2019.

In these figures, only low resolution profiles of Pico-Light convoluted with MLS averaging kernels are shown for clarity.

High resolution vertical profiles are visible in figure 10 (see below).

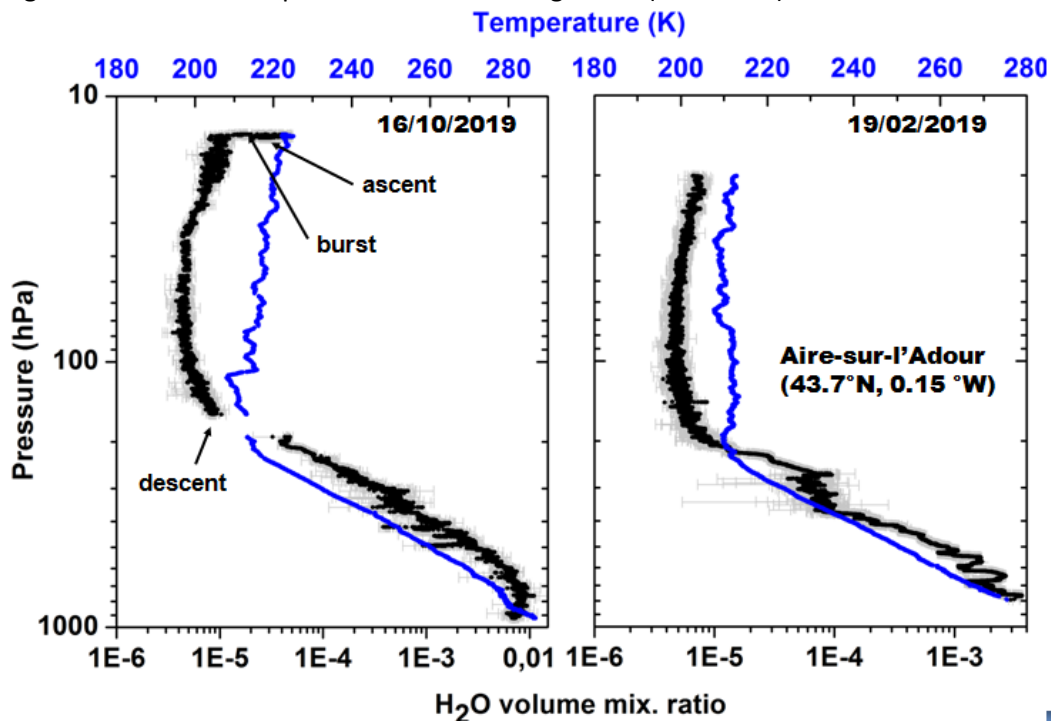


Figure 10: Vertical profiles of water vapor from the descent of Pico-Light H₂O on February 19, 2019 (right panel, black line) and October 16, 2019 (left panel, black line) together with error bars (grey shaded). The associated temperature profiles are shown as blue circles on each panel.

The gap around 200 hPa on October 16, 2019 comes from a bug in the electronics where data were not stored onboard. This problem has been solved afterward.

Old figures 10 and 11 have been removed. The comparison of Pico-Light with Imet sonde and RS 41 in the lower troposphere was of less importance compared to the discussion around GRUAN consistency (illustrated in figure 13 below) which has been added. Instead, we have included the following new figures :

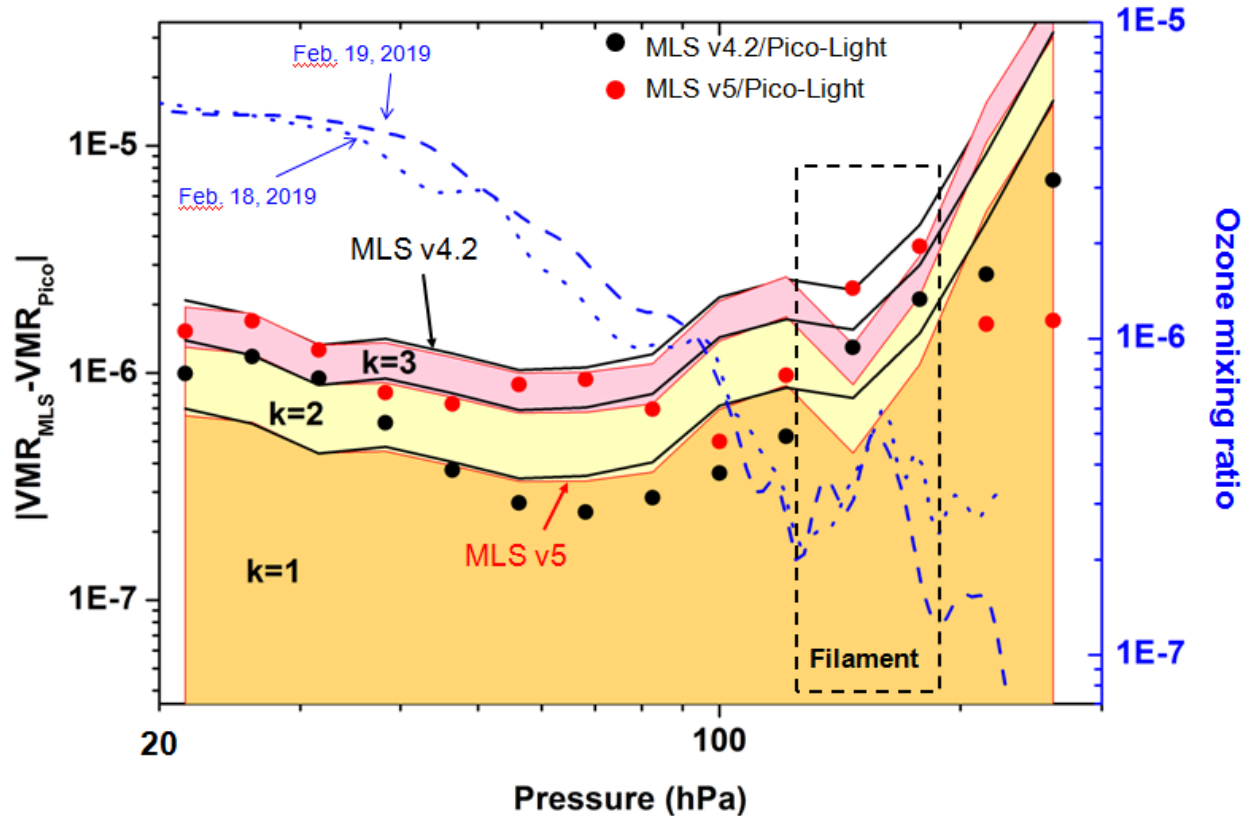


Figure 13: Consistency between Pico-Light and MLS v4.2 (black) and v5 (red) on February 19, 2019. The full circles illustrate the absolute difference in mixing ratio between Pico-Light and MLS datasets. The limits for $k=3$, $k=2$ and $k=1$ are represented as black full lines for MLS v4.2 and as red full lines for MLS v5. The area under the k lines for MLS v5 are filled with different colors for a better visualization. Ozone stratospheric profiles from the OHP (Observatoire de Haute Provence) LIDAR are shown in blue dash (February 19, 2019) and dot (February 18, 2019) lines.

Figure 14 and 15 provide a support for the analysis coming from figure 13.

MERRA 2 Ozone 19/02/2019 - instantaneous

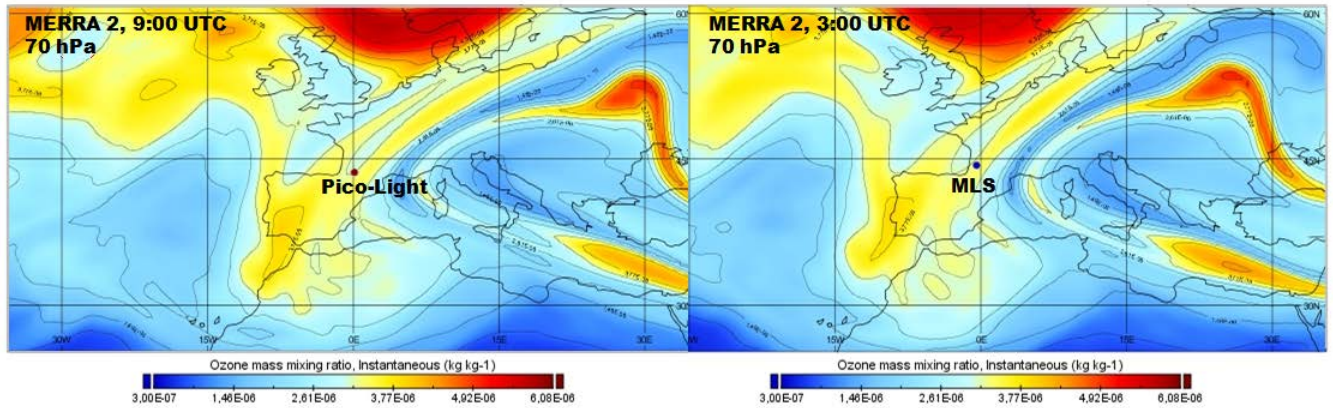


Figure 14: MERRA-2 ozone 3-hourly mixing ratio at 70 hPa on February 19, 2019, 3:00 UTC (left) and 9:00 UTC (right). The 9:00 UTC map represents dynamical conditions close to Pico-Light descent time and the 3:00 UTC map corresponds to the MLS case. Positions of Pico-Light and MLS mean position are shown in color circles.

MERRA 2 Ozone 19/02/2019 - instantaneous

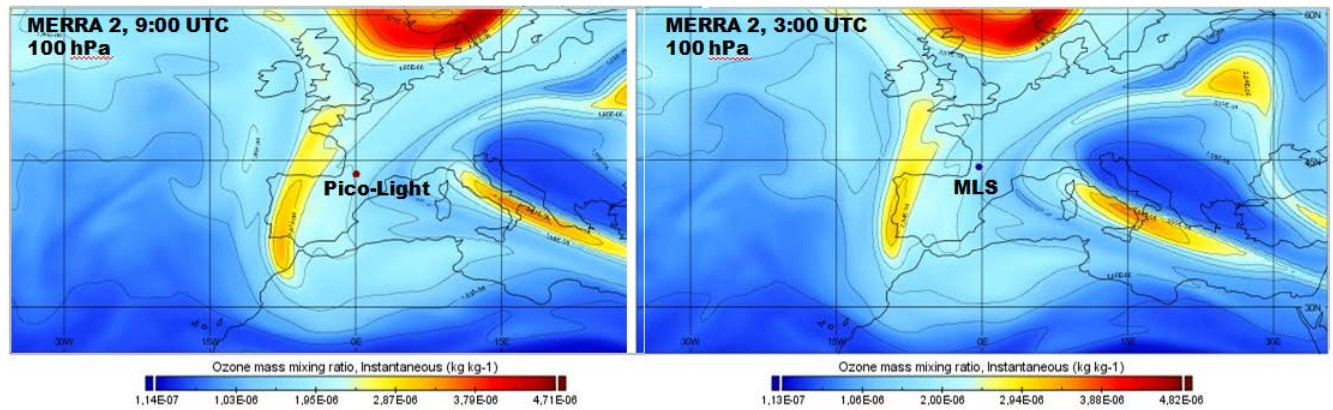


Figure 15: Same as figure14 but at 100 hPa.