<u>Response to Review #1:</u>

Major disadvantage for a climatology study is the limitation to the nighttime and cloud- less conditions. What I miss in the conclusions is who from the scientific community can use and benefit from such a climatology of nighttime water vapor.

At the request of Dr. David Whiteman's review, and based on your comment above, we have extended the conclusions and discussion to address the daytime and nighttime bias. We found that there was a difference of 0.2 mm/decade between the daytime and nighttime trends using the radiosondes. The difference between the two trends is much smaller than the uncertainty of each of the trends, therefore, it is not possible to say that there is a significant difference between the two regimes. We would argue that, in this case, the nighttime trend is representative of the average water vapour trend. If one had trend calculations based on a much longer time series and a smaller uncertainty, that assertion might not hold if the difference between the daytime and nighttime trend was detectable.

The PWV diurnal water vapour cycle for the Bern and Payerne was calculated in Morland et al. (2009) and most recently in Hocke et al. (2017) using microwave radiometer and GPS measurements. The average PWV diurnal amplitude is 0.5 mm, with a minimum of 0.1 mm and a maximum amplitude of 0.7 mm depending on the season (lower amplitudes in the winter and higher in the summer). Hocke et al. (2017) found that this translated to an average of 2% change in water vapour content over the course of the day with respect to the daily mean. More importantly, the peak of the diurnal cycle occurs around 19-20h local time and then reaches a minimum around 8-10h local time. This would suggest that the average nightly profile is actually a good indicator of the daily mean given the phase of the daily cycle of the region. Therefore, in this case, using the nighttime water vapour climatology can be assumed to be a good representation of the average water vapour. It is important to note that the same might not be true of other regions.

We have added a significant amount of text discussing the daytime and nighttime bias as well as the usefulness of the nighttime only climatology to the discussion section. We refer you to the new version of the paper for the new text.

P2 l17-18: *To my understanding, the variation could be a lot more than 100%, depending on meteorological conditions. I suggest to add a reference for this number, or restate the sentence in a more general way.*

You are right, a better phrasing would be that the "water vapour content can change by more than 100% over the course of the day".

The new sentence is as follows:

“Measuring an atmospheric water vapour trend … troposphere can change by more than 100% on a daily basis.”

P3 l8-9: *“published Raman wv lidars” - “publications on Raman wv lidars”... Also, Goldsmith et al., 1994 is not in the last decade.*

We agree that “publications on” would be a better phrasing. The sentence has been changed to:

“As far as we are aware there have been only four publications on operational Raman water vapour lidars in the last 2 decades (Goldsmith … etc)”

P3 l25: *I think it should be highlighted that this trend is inside the uncertainty range.*

The sentence will be changed to:
“They found a positive, but statistically insignificant, PWV trend at Cezeaux using …”

P4 l22-25: *It is important to make at least a short summary here (few sentences) for the method used for the retrieval, because it is crucial for understanding the rest of the manuscript.*

We would be happy to add a few sentences summarizing the optimal estimation method here. We will refer the reader to Sica and Haefele (2016) as well as Rogers (2000) for details.

The following sentences have been added to the text:
The OEM uses Bayes' theorem to constrain the solution space for the retrieval. It does this by adding in the use of an *a priori* state (**x_a)**. A probability of any given state of the system is assigned, assuming the errors of the system are Gaussian. The optimal solution for the system is then found by minimizing the cost of the solution, where the cost is defined as:

$$Cost = \left[\frac{1}{2}(\mathbf{y} - \mathbf{F}(\mathbf{x}, \mathbf{b}))^T \mathbf{S}_\epsilon{}^{-1}(\mathbf{y} - \mathbf{F}(\mathbf{x}, \mathbf{b}))\right] + \frac{1}{2}(\mathbf{x} - \mathbf{x_a})^T \mathbf{S_a}{}^{-1}(\mathbf{x} - \mathbf{x_a}).$$

The measurement vector is represented by **y**, **F** is the forward model for the lidar, **x** is the vector containing all retrieval parameters, **b** is the forward function parameter vector, **S_a** is the covariance matrix of the *a priori* values, and **S_e** is the measurement covariance matrix. The cost function consists of two terms. The first term is a weighted least-squares regression. The second term is a regularization term, which provides additional information to the solution through the specification of an *a priori* state. The *a priori* covariance matrix and the measurement covariance matrix define the solution space of the retrieval. Minimizing the cost function produces the retrieval solution **x**, where the solution is then the maximum *a posteriori* solution based on the probability distribution functions and is given by:

$$\hat{\mathbf{x}} = \mathbf{x_a} + (\mathbf{K}^T \boldsymbol{S_\epsilon}^{-1}\mathbf{K} + \mathbf{S_a}^{-1})^{-1}\mathbf{K}^T\mathbf{S}_\epsilon{}^{-1}(\mathbf{y} - \mathbf{F}(\mathbf{x_a})) = \mathbf{x_a} + \mathbf{G}(\mathbf{y} - \mathbf{F}(\mathbf{x_a})),$$

where **K** refers to the Jacobian matrix, **G** is the gain matrix. Where the Jacobian matrix is dy/dx and the Gain matrix is dx/dy. An in depth description of OEM theory applied to atmospheric physics can be found in Rogers (2000).

P4. L18: I think, it should be added a summary and discussion about the stability of calibration and any issues rising from it.

We'd be happy to add a few sentences here about the calibration. The following sentences have been added:

The continuous calibration is an internal calibration technique using the ratio of the solar background between the nitrogen and water vapour channels and was first introduced in Sherlock et al. 1999. The internal solar calibration method produces a relative calibration function which is scaled to the external calibration using the GRUAN-corrected radiosondes. The calibration function has an uncertainty of 5% of the calibration value and a corresponding uncertainty of 5% in the final water vapour mixing ratio. The uncertainty in the solar calibration method is the same as what would be introduced by using GRUAN-corrected radiosondes for external calibration (Hicks-Jalali et al. 2018). However, by using an internal calibration function, the lidar trends remain mostly independent of an external instrument.

P4 l 29: *The abbreviation OEM is nowhere defined and it is not well known. Also, it should be explained this approach , why it was selected and any drawbacks that it causes.*

Our apologies for not defining it previously. It should have been defined in the previous paragraph and that has now been fixed, in addition to a brief OEM description. There are multiple papers describing the background of OEM since it became an important technique for retrievals from satellite instruments in the 1970s. The third paragraph of section 2.1 has been rewritten as follows:

"One of the advantages of using an OEM retrieval over the traditional method (Whiteman et al. 1992; Whiteman, 2003) is the addition of the uncertainty budget and averaging kernels for each profile. The addition of the averaging kernels in particular are important because they may be used to more accurately compare results with other instruments which utilize OEM retrievals, such as satellite-based limb-sounding instruments, fourier transform infrared spectrometers, or microwave radiometers.

Another advantage of using OEM for lidar measurement analysis is that the measurements do not need to be corrected before being used for the retrievals. It can be more difficult to accurately propagate uncertainties through corrections to measurements which would prevent a complete uncertainty budget from being produced on a profile-by-profile basis. Leblanc et al. (2016) suggests a standardized method of calculating uncertainty budgets for the NDACC group, which is rigorous, but difficult to implement on a profile-by-profile basis. Additionally, corrections to the raw measurements can further induce uncertainties in the final product which may not be accounted for. Typical corrections for water vapour measurements include: accounting for photomultiplier paralysis (dead time), background noise, overlap, differential aerosol transmission, and sometimes merging multichannel measurements. The last of these can result in unknown uncertainties and biases in the water vapour, and is not necessary in OEM since the final retrieval is one profile which has been retrieved using all available measurements (Sica and Haefele, 2016). While the OEM has its advantages, some disadvantages of the method are that it is more computationally intensive than the traditional ratio method and that it is more difficult to implement. Nevertheless, a single retrieval does not take more than 30 seconds to run on an average personal laptop and the method is still quite practical for automatic and consistent processing of large datasets such as RALMO's. Longer run times occur when retrieving a large number of variables or when the bin size of the retrieved profiles are small.

We do not correct the RALMO measurements for the aforementioned possible signal effects; however, we have done some minor pre- processing before the measurements are entered into the OEM retrieval. We used nightly-integrated profiles …. "

P5 17: *This fact should be explained and provide some reference.*

This is explained in more detail in the thesis of the first author, as well as in Sica and Haefele (2016), both of which we will add as references to this sentence.

The sentence will be changed to:
"We found that these criteria effectively removed scans measured in the presence of optically thick clouds, and only left scans measured in the presence of optically thin/semi-transparent clouds such as cirrus clouds. Cirrus clouds are accounted for through the aerosol extinction retrieval in the OEM algorithm (Hicks-Jalali et al., 2019 and Sica and Haefele, 2016).

P5 l22: *Should the 30 min be continuous? If not, could natural variability add noise, when it could be hours apart?*

The 30 minutes is not required to be continuous. However, because we remove the presence of optically thick clouds we found that the signal level does not usually vary by more than a few percent over the course of the nights. The number of nights used in the study, and the monthly averaging for the final trend analysis, also helps account for occasional nights where the average nightly profile might have additional noise.

P5 l35: *It seems that the OEM output is one profile per night. But it is nowhere clearly stated. Is it possibly to discuss the variations expected if treated differently, on a clear winter night that could last 15 hours?*

Our apologies that this was not obvious. We had tried to state it on P5 l28, but the sentence is too vague. The sentence has been changed to:

"The final input profile to the OEM algorithm is a single "nightly-integrated" profile with an altitude bin size of 30 m."

The authors did not include a study on how the water vapour changes over the course of a night using OEM. However, if many points over the course of the night were included it would introduce a large autocorrelation into the trend calculation. In fact, we found that using daily values produced a large autocorrelation factor in the time series which led us to use monthly averages instead.
Additionally, the maximum retrieval height is determined by the amount of signal in the profile. Reaching the tropopause requires using nightly profiles for RALMO. We would have had to sacrifice several kilometers of altitude to do a study with higher temporal resolution.

P7 l 23 It is not clear what a "cost threshold of 3.5" is.

Our apologies that this wasn't clear. We have added the definition of the cost to the introduction so that this is now clearer to the reader. Costs close to 1 generally mean that the solution is well described by the model and measurements. High costs suggest, and the "high" is relative depending on the application, that the solution does not fit the model or measurements well. A cost of 3.5, for us, is on the border of when the solutions do not fit the model well and is a conservative cutoff for our retrievals.

P8 l5 *Are these the uncertainties discussed in the next page? If so, I suggest moving this plot and paragraph, after the definition of the uncertainties.*

The authors would disagree as this paragraph is about the statistical uncertainties whereas the following paragraph is about the parameter (systematic) uncertainties, that is uncertainties which are not reduced by averaging. The text and plots align with the appropriate paragraphs.

P8 l7 *The fact that the highest uncertainties are associated with the very low concen- trations in the upper troposphere should be discussed here and probably add some examples or a plot of absolute range of specific humidity uncertainty at different levels.*

We have added the following sentences discussing the uncertainties to this paragraph:

“The average statistical uncertainties are … profiles in each month. The strength of the Raman lidar signal and the statistical uncertainty is dependent on the amount of water vapour present in the atmosphere. Therefore, high statistical uncertainties are associated with low specific humidity levels. At high pressures where more water vapour is present, such as in the boundary layer, the statistical uncertainty is less than 1%. However, at lower pressures and near the tropopause where water vapour quantities are low, the statistical uncertainties reach an average of 14%. In the winter, when the air is drier, we see slightly higher uncertainties than in the summer at the same altitudes. At the surface, 82% of the profiles used in this study have statistical uncertainties of less than 5% and 96% of the profiles have statistical uncertainties less than 10%. At 250 hPa, 77% of the profiles have uncertainties lower than 20%, while 23% had uncertainties between 20 and 25%. “

The authors would prefer not to add another figure to the paper. However, we have placed the distribution of the uncertainties for the surface and at 250 hPa in the text above to give the reader a picture of the variability in the uncertainty.

P10. L9. I think it is more reasonable to integrate the radiosonde starting from 100m (or the height that lidar measurements are trustworthy) , and compare this modified PWV, if the full profiles of the radiosonde are available. By this approach the results are directly comparable and not affected by any bias for the near surface area.

We did not state this in the paper, but this has been taken care of. For the trends analysis (not the IWV climatology), the radiosondes were interpolated to the same altitude grid as the lidar and the IWV was calculated over the same altitude ranges. It is the authors’ opinion that the bias between the radiosonde trend on the same lidar dates is due to the uncertainty in the radiosonde measurements. We know that the radiosonde humidity measurements are not of GRUAN quality and have a larger uncertainty. When the dataset is limited to 25% of the available nights, this results in a higher uncertainty in the trend and probably the larger bias that we see between the two trends. When more nights are added, for example in the radiosonde trend which uses all available nighttime dates, the bias between the radiosonde and the lidar is decreased and the trend value converges closer to the true value.

Figure 5. *Only the days with lidar profiles are used. But as discussed earlier, it could be that the lidar profile is constructed from measurements hours away from the ra- diosonde, during cloudy nights. I suggest to investigate if using only profiles with data close to radiosonde timestamp could lower the biases.*

Unfortunately, we cannot do this suggestion for two reasons. Limiting the lidar measurements to coincident measurements with the radiosonde would severely reduce the number of nights (by roughly 50%) in our time series as nights that have clouds at that time would have to be thrown out. Additionally, we would not be able to get as high into the troposphere by limiting the lidar data to the time of the radiosonde for the reasons discussed previously.

Figure 6 It is not easy to claim that the natural variability at 230-250hpa is 80-90%, where the absolute values suggest is almost no humidity and the uncertainties are very high.

The uncertainty is accounted for by subtracting it from the variability. The large variability in July and August is actually between 70-80% and is real. We have verified that there are no unusual profiles inside those months that missed the filter. The large variability at those pressures is likely due to the dynamics around the tropopause.

P12 l3. *The high variability in the 600-400hpa region is one of the most interesting findings of the study. Figure 7 adds a lot of credibility to this pattern. I was wondering if a similar Temperature variation plot could provide more information for this behavior.*

We have calculated a temperature variability plot using the radiosondes as shown in Figure 1 below.

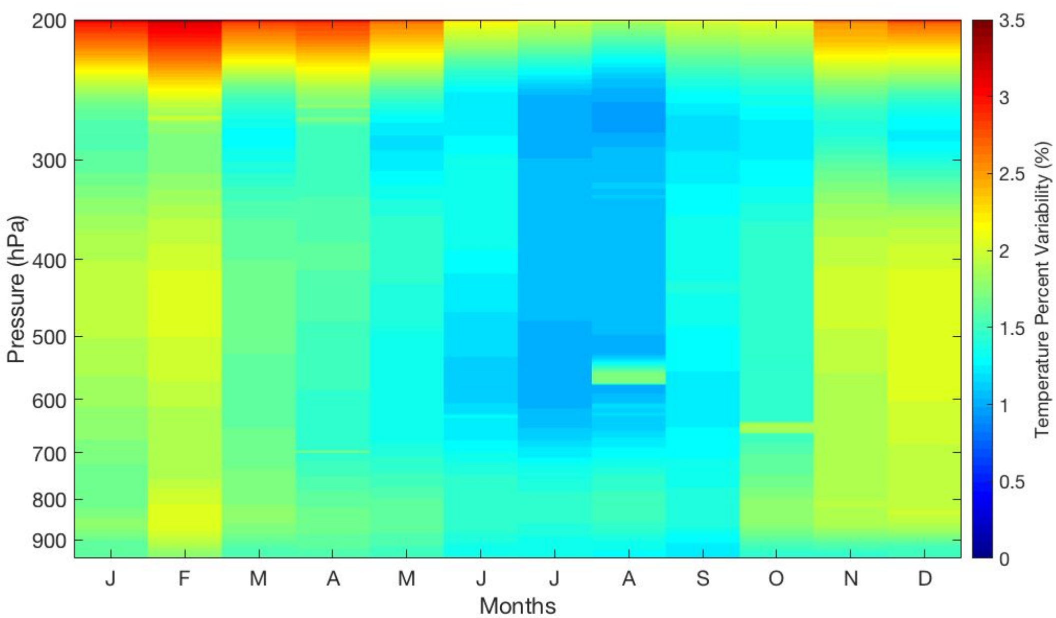

Figure 1: The temperature variability represented as a percentage of the mean temperature at each pressure level. The variability is calculated using daytime and nighttime radiosondes.

The temperature variability is calculated in the same way as the water vapour variability. The variability is defined as the rms of the temperature in one month, and then divided by the mean profile (in Kelvins) to produce the percent variability. Similarly to the water vapour, there is a decrease in variability during the summer months, and higher variability the rest of the year. However, there is no differentiation between the variability in the boundary layer vs the free troposphere as there is with water vapour. Additionally, the magnitudes of the variability are very different. If the water vapour climatology presented in the paper were in units of relative humidity, we think it would make sense to include the temperature variability in the paper. However, as the water vapour units we present are the mixing ratio, we do not think entering into a discussion of temperature variability adds much value to the paper.

Table 1: It is not clear how the trend %/C is calculated for RALMO.

You are correct, we did not explain how the %/C trend is calculated. The trend was calculated by dividing the water vapour trend in %q/decade (column 2) by the surface temperature trend (P16 l3). We have added the following sentence to the caption in Table 1 to state how the calculation is done:

“The trend values in column 3 were calculated by dividing the water vapour trend in column 2 by the surface temperature trend.”

Table 2: specify that these trends are derived from RALMO.

We have changed the first sentence in the caption to:

“Table of RALMO specific humidity trend calculations for each pressure layer.”

Summary and Conclusions sections are overlapping. I suggest to merge in one section.

We have merged the two sections as you suggested.

P20 l12, EOS climatology refers to what region?

According to the Hadad 2018 paper, the AIRS climatology was done for a 100 km radius around Cezeaux, France. We have added this information to the paper.

Congratulations on the very interesting work

Thank you very much for your time spent reviewing our paper and for providing helpful and thoughtful comments.