<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 I17-18: *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:

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

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:

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

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.
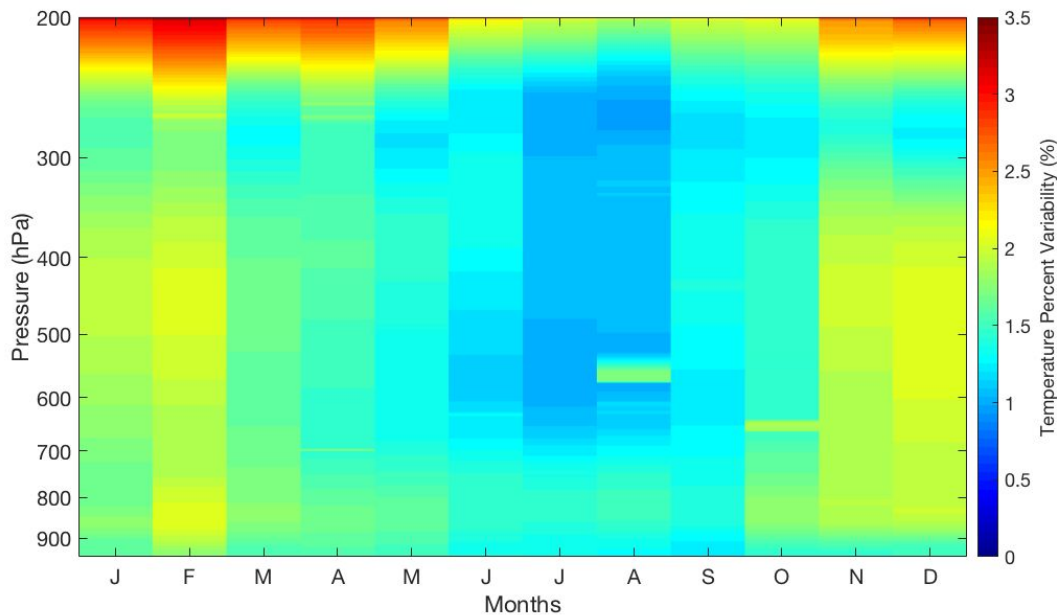

Figure 1: The temperature variability represented as a percentage of the mean temperature at each pressure level. The variability is calculated using daytime and nighttime radiosondes.

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

7

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.

<u>Response to Reviewer #2 - David Whiteman</u>

Response to Major Concerns:

1. *It is clear that to calculate PWV and trends as a function of altitude that nighttime clear weather profiles from the Raman lidar need to be used. What is the significance of limiting this trend study to nighttime only results? This question is not dealt with much in the text yet the radiosonde data are used to derive nighttime trends as well, if I understand correctly. It would be very instructive to use the radiosondes for trend calculations using daytime data only to be this nighttime limited lidar study in more context. Those results could be used to address the question of nighttime bias.*

At your suggestion, we have added the daytime PWV radiosonde trend to our results in Table 1 of this response. We have also included trends using both day and night radiosondes. It appears that there is very little difference between the daytime and nighttime PWV trends at Payerne. Additionally, the uncertainties in the trends are so large that even at 1 standard deviation we cannot accurately detect the daytime/nighttime bias. A much longer time series would be required to differentiate between the two regimes. The average PWV diurnal water vapour cycle amplitude at Payerne and over the region is about 2% (peak-to-peak 4%). The phase of the cycle is such that, on average, the peak water vapour amount occurs around 18-19hrs and decreases to reach a minimum around 6-8hrs the next morning (Hocke et al. 2017, Morland et al. 2006). This means, on average, the nighttime water vapour trends and climatology for Payerne are a good representation of the average water vapour content. Therefore, a nighttime only study is still quite useful to the water vapour community. The new radiosonde trends have been added to Table 1 in the revised manuscript.

Table 1: Precipitable water vapour (PWV) trends for RALMO and the radiosondes. The first column is the trend value and the second column is its corresponding 2-sigma uncertainty calculated using Eq. 5 in the new version of the paper. The first radiosonde trend is calculated using nighttime data coincident with the lidar measurements. The second radiosonde trend is calculated over all available nighttime dates. The third trend uses all available daytime only data and the last trend uses all available data over day and night.

	PWV Trend (mm/dec)	PWV 2*sigma (mm/dec)
RALMO	1.32	1.51
RS (lidar dates, night only)	2.31	1.53
RS (all dates,night only)	1.08	1.05
RS (all dates, day and night)	1.18	0.97
RS (all dates, day only)	1.19	0.99

2. *The authors also study geophysical variability including in the boundary layer in section 3.1. The value of such a nighttime only study of variability should be justified and, if possible, contrasted with the values in the daytime. The variability the authors have calculated, particularly in the BL, will likely be biased by the lack of daytime measurements. Why not include all day and night lidar data in a variability study which hopefully would at least be valid through the BL. Then the nighttime only results can be contrasted with that. Also, the authors use the radiosonde data to compare nighttime variability with the lidar. Why not include daytime radiosondes as well in the variability study to put the nighttime limited lidar-based study in more context?*

While the authors would have liked to have used the daytime RALMO measurements it was not possible for this trend analysis. RALMO has a known daytime bias in water vapour on the order of 10% (Dinoev et al. 2013) which is not yet fully characterized and understood. For this reason the authors were uncomfortable using the daytime data for a trend analysis. There is currently work under way to characterize and correct the bias and hopefully at a future date another daytime trends paper can be published.

However, we have now included the daytime radiosondes to the variability study (Figure 1). Adding the daytime radiosondes such that we now have the total variability (combined daytime and nighttime radiosondes) does not significantly change the average water vapour variability significantly. The largest changes occur on the edges of the boundary layer in the summer where the variability is increased by 10-15% when adding daytime radiosondes, and in the winter

between 350 - 200 hPa where the variability is increased by 10 - 20% (Figure 2). Everywhere else the variability remains the same. Adding the daytime radiosondes does not change the overall behavior of water vapour's variability and the radiosonde daytime and nighttime variability still mirrors that which is seen by the lidar. It is important to keep in mind the exact differences between adding daytime radiosondes to the variability cannot be taken as absolute given that there are many uncertainties associated with daytime radiosondes (Dirksen et al. 2014) which we have not taken into account and corrected for. The results here are meant only to show that the lidar captures the same behavior and to act as a sanity check.

Figure 1. Radiosonde-measured water vapour variability using both daytime and nighttime measurements (New manuscript Fig.7)
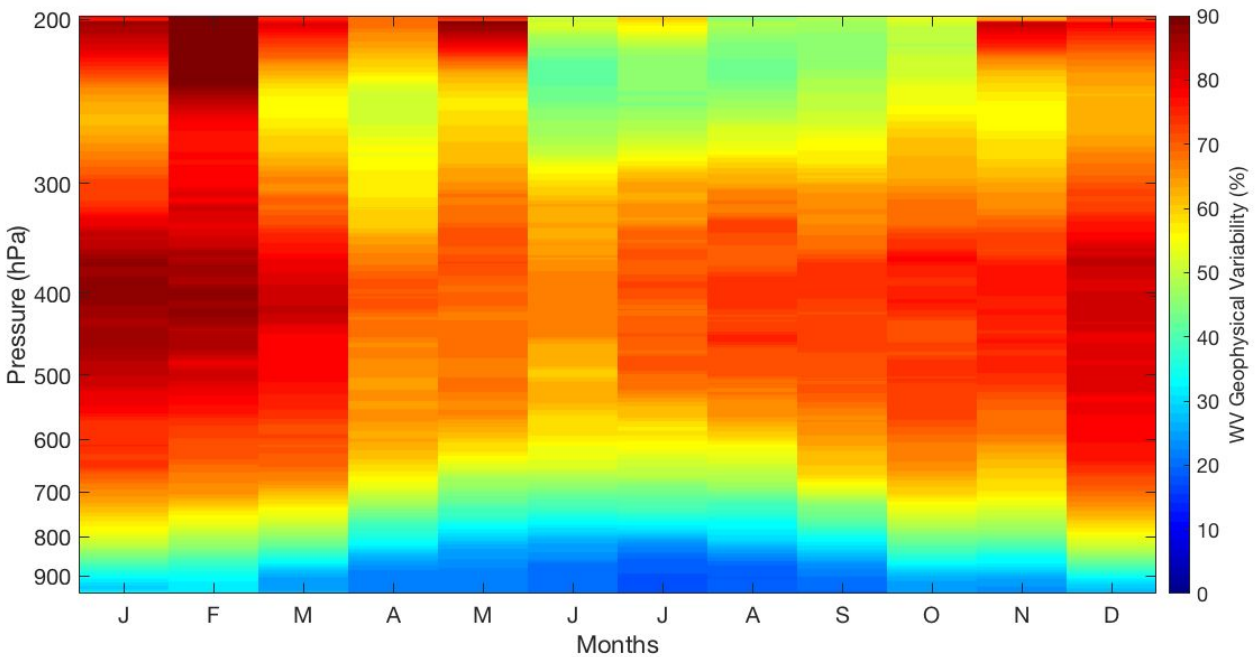
Figure 2: Percent difference between new Fig 7 (Day and Night measurements) and old Fig 7 (Night only). The largest differences are at the edges of the boundary layer where adding the daytime sondes increases the variability between 5 and 15% and in January and February between 300 and 200 hPa. The large difference in April between 300 and 200hPa and at 700 hPa was due to removing 2 nighttime profiles which should have been removed previously due to a faulty sensor and bad measurements.

Percent Difference (Night - Day&Night)
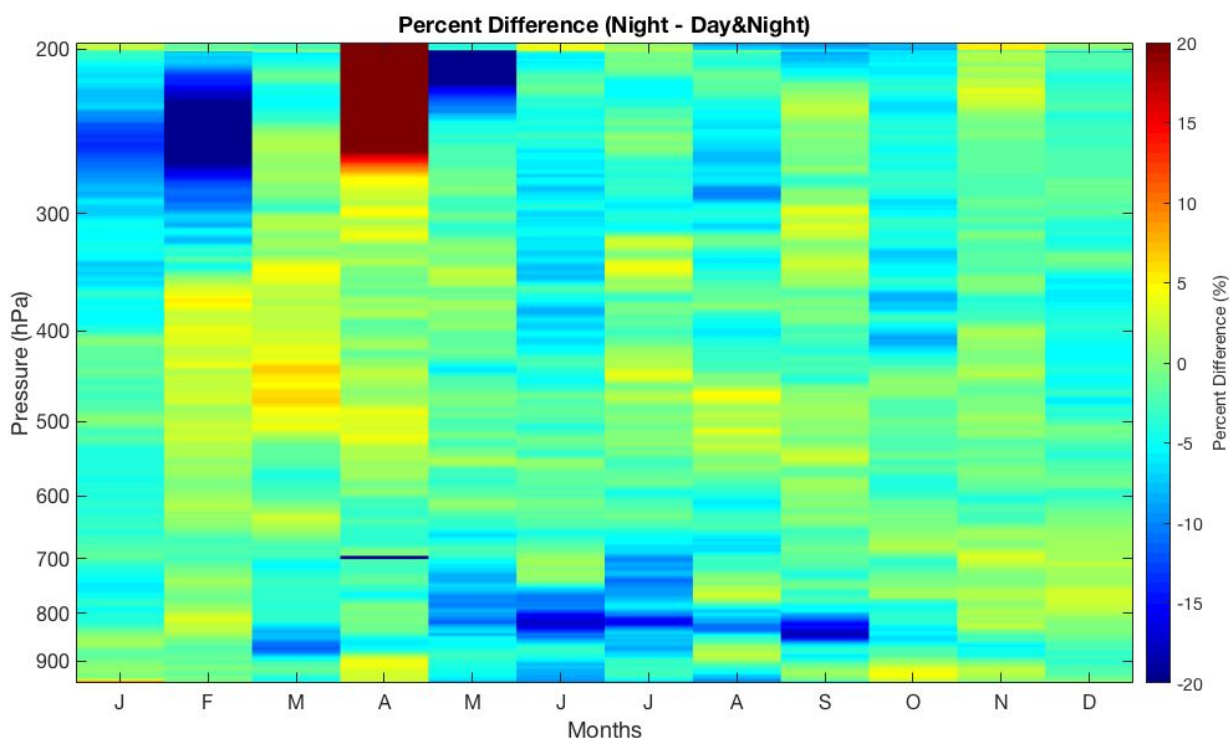
3. *The authors reference Whiteman et al., 2011 as indicating that their 11.5 year data set should be sufficient for revealing trends. However, the results in Whiteman et al. indicated that trends could be resolved at the 200 hPa level (approximately the most efficient level for trend detection according to the Whiteman study) if noise-free measurements were available 30x per month. The RALMO 30-minute profile measurements, according to figure 1, are available usually less than 15x per month. So the Whiteman results would seem to indicate that trend detection with the RALMO dataset would often fail and in fact from the authors results in Table 2 most trends presented are not revealed at the 95% confidence level, which seems consistent with the Whiteman et al. results.*

Yes, you are correct. We should have reiterated this fact in the final conclusions. We have added text to this fact in our discussion section. As the discussion section has been significantly re-written we will not paste the text here but refer you to the new version of the paper. When we mentioned your work at the beginning it was meant to serve as a hypothesis and supposition

only. We think that these sentences are confusing to the reader and have been rewritten as follows:

"We have 11.5 years of measurements and are therefore in the realm of possible detection at the 95% level; however, whether or not we can detect trends depends on the variability of the atmosphere at the levels we choose and the magnitude of the trend itself. As we will show, the variability of the atmosphere in the free troposphere will determine our ability to detect trends at the 95% level."

4. *The authors refer to Table 1 of Weatherhead et al., 1998 and seem to be referring to the σN term as the "uncertainty of the measurement". Instead this term represents the standard deviation of the noise of the time series, which is comprised both of natural atmospheric variability and measurement variability due both to random and systematic sources. This number can be calculated from the time series itself as it seems the authors do at other places in the manuscript. In Whiteman et al. 2011, using radiosonde data, values of σN that ranged between ~20-80% were found. I would expect values of σN much larger than the 6% used here.*

Thank you for catching this mistake. We had interpreted the value correctly in our results, but did not use the correct terminology in the text. The uncertainties for our trends were calculated correctly using the standard deviation of the trend residuals. We have fixed the text in the paper to say:

"According to Weatherhead et al. (1998), trends of at least 5% per decade should be calculable in 10 years provided the standard deviation of the noise in the trend residuals is less than 6% and the autocorrelation of the trend fit residuals is less than 0.4 (Table 1 of Weatherhead et al. (1998))."

Response to Minor Concerns:

1. *Occurs in several places. The terms water vapor mixing ratio and specific humidity seem at times to be used interchangeably. They are different quantities so please be clear on what quantity is calculated and maintain consistency.*

In all cases, the term specific humidity is used when we have actually calculated specific humidity and not the water vapor mixing ratio. While the OEM retrieval is in units of mixing ratio, we have converted them to specific humidity for the height-resolved trend analysis and the climatology to maintain consistency with the water vapour trend literature. We have added a sentence to this effect in the second paragraph of section 2.1 where the OEM retrieval is

introduced. We realize that the part where this gets confusing is when we do not convert the uncertainty discussion to units of specific humidity. We have now added a few sentences to explain the difference when the uncertainties are first introduced. The uncertainties are not converted to units of specific humidity. The differences between the uncertainties for mixing ratio and specific humidity are negligible.

2. *P 2, line 18. Suggest change to "Due to this huge variation, quantifying water vapor trends requires ..."*

We have changed this as suggested.

3. *P2, line 32. AIRS is referred to as a "satellite". It is an instrument on the Aqua satellite.*

Thank you for catching this mistake. We have corrected it.

4. *P2, lines 32-33. Authors state that AIRS can measure down to the surface. The reference provided discusses a vertical resolution of AIRS of approximately 3 km in the lower troposphere. I do not believe that a measurement with vertical resolution of ~3km can be considered to extend to the surface. Please revise this claim to be more consistent with your next sentence which seems to contradict this one by stating that the vertical resolution of satellite measurements is "typically on the order of kilometers".*

Thank you for correcting this. We have changed the sentence as follows: "... but a few like the AIRS instrument on the Aqua satellite can accurately measure through the troposphere ..."

5. *P3, line 5. Suggest change to "...trend measurements are ..."*

We have corrected this thank you.

6. *P3, line 9. Complete the thought with, e.g. " ...have been run operationally over the last decade   (refs) but none has been used to support a study of trends as done here." or something to that effect.*

Thank you, we have changed the text to: "As far as we are aware, there have been only four publications of operational Raman water vapour lidars in the last 2 decades \citep{Goldsmith1994, Dinoev2013, Hadad2018,Reichardt2012}. While these lidars have been run operational over the last 2 decades, RALMO is the only one which has presented a water

vapour trends study that we show here."

7. *P3, line 12. "...many of which were insignificant." implies that some were significant. Please expand briefly to discuss the details of the trends that were found to be significant.*

The authors don't think it would be a good idea to describe which trends were and were not insignificant here for the whole paper as Ross et al. (2001) is a global study and that would get quite long. However, we have added an extra sentence here so the new sentences read as follows:

"Ross et al. (2001) calculated PWV trends from 1958--1995 using radiosonde measurements from the surface to 500 hPa and found highly geographically variable trends around the globe. Their trends calculated across Europe were insignificant and differed in sign depending on the location."

8. *P3, line 20. Suggest change to "...found a positive but insignificant trend ..."*

We have changed it as suggested.

9. *P3 lines 25-26. The PWV trend values from Cezeaux have already been mentioned above. Delete this redundant information.*

Thank you, we have removed this redundant sentence.

10. *P3, line 29. A 20-year old reference (Weckwerth et al, 1999) is used to support claim about routinely available measurements with resolution better than 1 km. Can you find a current reference that supports that claim? The reference below may help*
    1. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014RG000476

Thank you for providing a better reference for this statement. This paper appears to corroborate our original statement, with the only instruments with higher than 1 km vertical resolution in the troposphere being lidars and microwave radiometers (in the boundary layer). Therefore, we think the original statement still stands but we will replace the Weckwerth et al (1999) reference with Wulfmeyer et al. (2015) instead.

11. *P3 line 34. This study is described here as being 10 years in length but elsewhere 11.5 years is used. Please reconcile.*

Thank you for catching this. The correct time is 11.5 years. We have changed this everywhere in the paper.

12. P3. Line 35. This is the place to first note that this study uses only nighttime measurements.

We have added the modifier nighttime to the RALMO measurements in this sentence.

13. P4 line 14. Suggest change to "...laser operating at 30 Hz..."

The sentence has been changed as suggested.

14. P4, line 33. Please also consider the fully propagated uncertainty estimates from the MOHAVE 2009 campaign contained in Whiteman et al., AMT, 2012, Appendix 3.

We have added this reference in addition to mentioning Leblanc et al. (2016).

15. P 5, line 15. unit is used of "counts/bin/m". I suspect this should be "counts/bin/min" instead. Right?

Yes, that is a typo thank you for catching it.

16. P5, line 21. Suggest change to "...water vapor profiles have ..."

We have changed this as suggested.

17. P5, line 26. Authors state that measurements are "naturally biased towards high pressure system conditions ...". I hope that the authors add an analysis of their daytime radiosonde data (major concern #1) and can then include a section that deals with the possibility of nighttime bias. That section could be referred to here.

We have added daytime trends using the radiosonde measurements to Section 4. We have added a sentence here referring to the section.

New Sentence: "We will compare the nighttime only PWV water vapour trends to daytime radiosonde trends in Sect. 4.2 and discuss the daytime/nighttime bias."

18. P6, line 12. Suggest change to "Measurements from GCOS ... highest quality radiosonde data product available".

We have changed this as suggested.

*19. P6, line 13. Suggest change to "Unique to GRUAN radiosonde data products is the
calculation of absolute uncertainty estimates for their measurements as a function ..."*

We have changed this as suggested.

*20. P7, line 5. Please include some more on these non-GRUAN sondes. What type? References of
previous use of this type of sonde? Efforts to get GRUAN certified?*

The Swiss C30 and C50 radiosondes are the swiss-made operational radiosondes that are not
GRUAN certified. There are no published validation studies describing the radiosondes, but they
have been used for comparisons with the lidar in previous papers. Alexander Haefele can be
contacted if the readers would like access to internal documents on the radiosondes. We have
added the following text to the paper describing the individual radiosondes used for the
operational radiosonde time series.

"The Meteoswiss operational radiosonde time series uses multiple radiosonde types: the SRS400
\citep{Martin2006, Morland2009, Brocard2013} from 2009--2010, SRS-C34 from 2010--Jan
2017, SRS-C50 from Feb 2017--March 2018 , and the Vaisala RS41 (Jensen et al. 2016) from
March 2018 - present day. The C34 and C50 radiosondes are manufactured by MeteoLabor. The
C34 uses a Sippican hygristor for humidity measurements which is quoted to have an accuracy
of 2% RH (Meteolabor 2010), however, no public validation studies have been published. The
C50 is the updated version of the C34, however, no public documentation exists on the C50
specifications … The scatter in the PWV measurements did not change appreciably between the
different types of radiosondes, therefore we have not attempted any homogenizing between the
different data sets."

*21. P7, line 22. Authors refer to "...which pass the cost threshold of 3.5 are ...". No cost function
has been introduced. Either change or add material prior to explain what this refers to.*

We have added new material to Section 2.1 discussing some OEM theory and have added the
cost function. The readers should now have enough information to understand. We refer you to
Section 2.1 for the new text and equations regarding OEM theory. We have not included a
detailed summary of OEM because we believe that there are enough references readily available
for the readers to study OEM in more detail.

22. P7, line 26. Suggest change to "This means that the minimum ..."

We have changed this as suggested.

*23. P9, Fig 3. Caption states fractional uncertainties but figure is labelled in %. Please
reconcile.*

Thank you this should say "percent uncertainties".

24. P9, line 1. Fig 4 is described as being in units of mixing ratio while Fig 2 was in units of specific humidity. Did you really change units to do these calculations?

We have now explained this in more detail in the text, but the OEM retrieval is in units of mass mixing ratio. However, to maintain consistency with other water vapour trend literature we have converted the final profiles to units of specific humidity. The uncertainties were not converted because these are calculated directly through OEM and the largest difference between the mass mixing ratio and specific humidity is less than 2% of mass mixing ratio.

25. P9, line 10. Change "ration" to "ratio".

Thank you for catching this typo.

26. P9, line 20. I am surprised that cloud retrievals contribute much to the statistics since the laser will be attenuated quickly in clouds. So do you mean to refer here to thin or partial clouds? Also, in your retrieval, if you detect a persistent cloud why not set the Angstrom coefficient=0 (clouds are white)?

The persistent cloud should actually be changed to say persistent *thin* cloud. These thin clouds go undetected by our cloud filter and therefore in the automated OEM retrieval process we don't know that they exist. If someone ran the OEM retrieval for each night individually, they could certainly set the Angstrom coefficient very close to zero if they detected a cloud. Setting it to actually zero would result in a singularity in the ARTS OEM algorithm and is therefore inadvisable. We decided to run the OEM with the same settings for the entire trends analysis to maintain consistency and the nights with persistent thin clouds did not affect the water vapour profile retrievals.

27. P 10, Fig 4. This may get too much into the details of OEM for a discussion here, but I don't understand how the calibration "constant" can have a height dependent systematic uncertainty.

The calibration uncertainty is technically "the uncertainty of the water vapour profile due to the uncertainty of the calibration constant". Therefore, the calibration parameter uncertainty is applied to the entire profile, however, the calibration constant uncertainty is in fact a constant value.

28. P 10 , Fig 4 caption. Authors state that "All other uncertainties contribute less than 0.1% on average. This does not appear to be the case for the NCEP air density. Please check.

The sentence is incorrect, thank you for catching the mistake. We have added this sentence to the caption:

"The average uncertainty contribution from the NCEP air density is .3%."

29. P 10, line 9. Is the difference in PWVs shown in Fig 5 consistent with the PWV contained in the bottom 100m? It might be a good idea to include an extrapolation of the lidar profile to the surface (using a measured surface value) to account for this missing part to try and resolve this difference.

We have double-checked these results by using the radiosondes. We could do this by interpolating the surface data, however, we felt uncomfortable assuming a straight linear relationship between the surface measurements and the lidar measurement. We compared the PWV calculated with and without the first 100 m up to the final lidar height and found that the difference is not completely explained by the 100 m difference in the altitudes used. The average water vapour content in the first 100 m is roughly 0.2 mm. There is a slight wet bias in the radiosondes compared to the lidar of around 0.4 mm. However, the difference is within the 1-sigma standard deviation of the measurements for both the lidar and the radiosondes. The difference between the two could be caused by the nightly averaging of the lidar compared to the radiosonde. Additionally, while the same nights are used for the radiosonde and lidar comparison, the lidar may not be measuring at the same time if clouds are present. If clouds are present at the time of the radiosonde launch, then the sonde could be wetter than the lidar which did not use the measurements during that time period. The average difference between the radiosondes and the lidar PWV is 5%, which is within the uncertainty of both the lidar and the radiosonde measurements. If you feel like interpolating from the surface station would be better, we can do so.

30. P11, line 15. "The see ..." ??

Thanks, we have removed this artifact from a previous edition.

31. P 12, lines 3-4. Authors state "The first, and most straightforward, explanation for the high variability at these levels is the presence of mid-level layers of clouds or aerosols". If this is the case, then you are not quantifying water vapor variability but rather something that is

contaminating those calculations. Please consider whether there are additional software filters you can put on the data to prevent this contamination.

That's true, we hadn't thought about it that way. We take care of any left-over mid-level layers of clouds and aerosols through the OEM's aerosol extinction retrieval. Therefore this cannot be the explanation. We will remove it from the text. This would leave the dynamics as the main contributor to the variability in the region.

32. P13, line 7. Suggest change to "...in the free troposphere could explain ..." since you are speculating here.

We agree and have changed this as suggested.

33. P13, line 11. Authors state "The smaller average concentrations of water vapour in the winter leads to a larger percent variability." Smaller average concentrations do not by themselves necessitate larger percent variability. You still need to invoke some dynamical argument here to explain it.

We agree that this argument is not correct. A more plausible argument is that the weather in central Europe during the winter and spring/fall is highly variable due to an increase in exchanges between colder and dryer Arctic air and warmer southern air. The dynamics of the polar jetstream heavily influence the weather during this season which could contribute to the increased variability in water vapour. We have removed those sentences from the text and added the new sentences below:

"European winter weather is highly influenced by the polar jetstream. An increase in the variability of water vapour during these months could be caused by an increase in the exchange of air between the Arctic and the mid-latitudes."

34. P 14. line 28. "We linearly interpolated ...", This seems a curious technique although any technique that you use to reconstruct data can be criticized. In any case, it does not make sense to me to perform a linear interpolation to re-construct missing data. Your fits clearly show that some sinusoidal behavior is more appropriate and that these linearly interpolated values then look like outliers. I suggest that you perform the seasonal fit (necessarily excluding the missing data) and then use the derived seasonal function to characterize the pdf of the noise around this fit function. The seasonal fit function and the calculated distribution function can then be used to create randomize fill values for the missing data. This bootstrap technique is preferable to what the authors have done and I believe a the standard technique for dealing with missing data in these type of trend calculations.

We have tested the technique you suggested on the trends which had missing data (all lidar trends and the radiosonde trend using only coincident dates with the lidar). We found this technique had very little effect on the trend value, however, in all cases it slightly lowered the trend magnitude. It appears that linearly interpolating between missing points slightly increases the trend value; however, it allows for the calculation of the autocorrelation of the time series. As we understood it, the suggested method does not produce a final profile with filled points. Therefore, the uncertainties in the newer bootstrapping method can only be calculated using the standard deviation of the trends. Trend uncertainties derived via the distribution of the trends result in much smaller values because the autocorrelation is not taken into account leading to a false sense of significance in the trend. As such, the authors would argue that it would be better to be more conservative and use the original trend calculations in the paper given the uncertainties calculated are probably more realistic and the difference between the two methods is roughly .25 mm/dec for the PWV trends and on average about 1%/decade for the height-resolved trends. The Tables 2 and 3 below show the differences in the trends from the two methods and their uncertainties. The Table 2 is for PWV trends and Table 3 is for the height-resolved specific humidity trends. Note that the suggested bootstrap method's uncertainties are smaller than those produced by the Gardiner bootstrap method. The suggested bootstrap method's uncertainties are smaller due to the fact that only the replaced points are shuffled and not all points therefore there is less scatter in the trend values than in the Gardiner bootstrap method.

Table 2: PWV trends using the original bootstrap method with linear interpolation and the suggested bootstrap method ("Whiteman PWV Trend")

	PWV Trend (mm/dec)	PWV 2*sigma (mm/dec) (w/autocorrelation)	PWV 2*sigma Gardiner Bootstrap (mm/dec)	Whiteman PWV Trend (mm/dec)	Whiteman PWV 2*sigma (mm/dec)
RALMO	1.32	1.51	1.34	1.17	0.32
RS (lidar dates, night only)	2.31	1.53	1.30	2.08	0.34

Table 3: Height-Resolved specific humidity trends using the original bootstrap method with linear interpolation and the suggested bootstrap method. Trend units are in percent per decade.

Pressure (hPa)	Trend (%/dec)	2sigma (%/dec) (w/autocorrelation)	2sigma Gardiner Bootstrap (%/dec)	Whiteman Trend (%/dec)	Whiteman 2*sigma (%/dec)
950	12.11	10.25	6.75	11.08	2.08
900	0.84	8.15	5.02	0.37	1.74
800	9.83	10.45	7.2	9.05	2.15
700	13.94	14.36	12.26	12.94	3.12
600	15.85	16.92	13.81	13.40	3.71
500	4.80	16.19	14.32	4.97	3.67
400	10.87	15.79	12.68	10.60	3.28
350	12.59	14.07	13.93	11.43	2.91
300	7.08	14.01	11.9	6.53	3.21
275	3.19	11.6	9.74	2.45	3.18
250	3.55	11.48	10.96	1.39	3.51

35. P 15, Fig 8. Suggest you add the equations of the linear trend fit to the figure for each of a, b, c.

Adding the entire equation would not fit in the figure, however we have done the same as in Figure 9 and added the trend values to the figure with their uncertainties.

36. P 15, line 9. "The difference between the two methods represents the bias from using only semi-clear nights during clement weather." The difference seems significant at the 90% level. Given that, can you conjecture as to why use of semi-clear nights during clement weather may yield different trend values? Or is the significance in the difference in the trends not large enough to draw such conclusions? Again, it would be nice to have some results that contrast day and night values as from your radiosonde dataset.

It does appear to be significant. However, after further discussion between the authors, and given the new results from the daytime radiosonde trends, we are now not sure that is the case. It is more likely that the large trend is artificial and caused by the fact that such a small portion of the radiosonde dataset is used (25%) in combination with the operational radiosondes' larger errors.

When using the entire dataset the trend converges to the real value as the random error is reduced by adding more radiosondes. We have added this discussion to the text and changed the sentence to "could represent the bias from using …".

New text is as follows:

"However, the bias between the two radiosonde nighttime trends is more likely caused by the number of points used in each trend and the uncertainty in the radiosonde measurements. Only roughly 25% of the possible nights are used in the trend analysis for the lidar dates, and as the radiosondes do have a larger and uncharacterized uncertainty it is more likely that the large trend value is due to a larger scatter in the radiosonde measurements. Given that the lidar trend is much closer to the radiosonde trends which use all available nights, including the radiosonde trends calculated using daytime and nighttime measurements, the large radiosonde trend using only nights consecutive with the lidar is more likely due to random error and not a difference in weather."

37. P15, line 13. The RALMO calculated PWV trend is not significantly different from 0 at the 95% confidence level. Right?

That is correct. It is statistically significant at 90%, which is stated in the paper.

38. P 16, line 14. Again linear interpolation is used here. I strongly suggest that you use the bootstrapping technique described in 34 above to fill these values. Linear interpolation will at least slightly skew the results as can be seen from your Fig 9.

We have answered this response above in Question 34. On average the difference between the trends from the two methods is about 1% per decade.

39. P19, line 23. Authors give calculated trend values of 1.3 and 2.3. Please add uncertainties to these values.

Thank you for pointing this out, we have added all missing uncertainties to the trends in the summary.

40. P20, line 5. "RALMO is the only lidar ..." Be careful ... the DOE ARM Raman Lidar certainly has produced such a dataset as well (over a longer time period, actually) with a higher percentage of up time.

Yes, the DOE ARM Raman Lidar has certainly produced a larger dataset than RALMO. However, the authors have not found a published climatology from that lidar. We have changed the sentence as follows:

"RALMO is one of a few lidars which has produced a published high vertical resolution water vapour climatology of the troposphere with 11.5 years of consistent measurements."

It's quite possible that we've missed a reference here. So if you are aware of the paper that has their climatology we'd happily include it in our references to make sure it is cited appropriately. We have looked in the ARM database as well as through other journals without any luck.

41. P20, line 23. "Most satellite climatologies of water vapour only extend down to 300 hPa ...". I would modify this statement since AIRS and the other hyperspectral sounders (CrIS, IASI) have some lower tropospheric sensitivity.

Thank you for this suggestion. The authors have decided to delete this paragraph as it does not seem to fit the discussion in the previous two paragraphs, nor is it really necessary. It seems odd to say that it is difficult to compare a climatology to anything but radiosondes when we compare it to the H2018 AIRS climatology in the previous paragraphs. Therefore, we have removed it.

42. P2, paragraph starting with line 30.

1. *A comparison of results of trend calculation is given here but there are no uncertainties given with any of the values. Many of the comparisons made may not be significant if you consider the uncertainties in the trend values stated. You many want to consult the 2011 Immler paper on GRUAN uncertainties for a description of language to use when describing the differences in numbers. See Table 1 for metrics to determine the use of terms such as "consistent", "in agreement", "significantly different", etc.*

   We agree that more detail regarding uncertainties and/or statistical significance should be included. Unfortunately, the Nyeki paper did not provide uncertainties in their trends, however, they did provide significance at the 90% level which we have now added to the discussion. We will refer to the language in the Immler paper in the updated discussion.

2. *Authors discuss both nighttime limited and day and night results here. As mentioned in major concern #1, I strongly suggest you consider expanding those daytime results to include trends calculated using your own radiosonde dataset to be able to expand this discussion. Such results seem conspicuously lacking here.*

We have added 2 new paragraphs to the discussion section regarding the day/night bias in the trends and in the climatology. As the Discussion section has changed appreciably, we would refer you to the updated section in the paper and will not paste it here.

3. *Authors state in conclusion at the end of this paragraph: "Therefore, while there certainly is a natural selection bias due to only using nighttime measurements in our study, the magnitude of the nighttime bias is not currently detectable." I would revisit this statement after considering the uncertainties as mentioned above and adding in results from your daytime radiosondes. It may be that the results still do not reveal a nighttime bias. However, it sounds too bold to claim that the "nighttime bias is not currently detectable." Instead a statement that seems defensible might be "Based on these results we do not detect a bias using only nighttime measurements."*

We have added the new daytime radiosonde trend calculations to the discussion and based on these results (presented in Table 1 after Major Concern #1) we have not detected a bias in using only nighttime measurements. We have changed our previous language to match your suggestion. We have made many changes to the Discussion section, therefore, we refer you to the updated section in the paper for the new text.

43. *P21, paragraph starting with line 30. Authors consider the magnitude of trends and the sensitivity of RH to changes in temperature and then compare those sensitivity numbers. Again, here, the uncertainties have not been considered. Please consider the uncertainty in both the trend of water vapor and the trend of temperature and propagate those uncertainties into your calculation of the sensitivity factors (%/C). I suspect that some of the differences are not statistically significant. And again when considering the differences you can use the language of Immler et al., 2011.*

We have now added uncertainties wherever possible to the discussion. Unfortunately, in the cases of Nyeki et al. (2019) and MS2017 we cannot put the exact numbers because they are not provided. We have added ours and others wherever possible. We have also added text regarding the significance of the trends in units of % per degree C. You are correct, when propagating the uncertainties they increase and the trends lose their significance.

44. *P22, line 11. Suggest change to "Interestingly, neither our trends nor ... conforms to ..."*

Thank you for catching this mistake.

45. *P23, line 25. Change to "ratios".*

We have changed this.

46. P24, line 6. I guess I don't understand what you're doing as it sounds like you are able to perform this check only 1 time per year (at highest SZA on winter solstice). Please expand to clarify this point.

We apologise that the text in this section wasn't clear. We can do this check every clear day. In order to minimize the seasonality component and the difference in the optical depth of the atmosphere at different SZA, we chose the highest common SZA. The highest SZA on the winter solstice is 20 degrees. Therefore, to be consistent across the entire year, the calibration is done at the time the sun reaches 20 degrees every morning. We have changed the text to make this clearer to the reader:

"One must be careful to consider both the diurnal and seasonal solar cycles when using this technique, therefore, it is important to use the same solar zenith angle each day at the highest possible angle. The highest solar zenith angle on the winter solstice is 20 degrees. Therefore, the calibration is conducted each cloudless morning when the sun is at a solar zenith angle of 20 degrees."

47. P24, line 10. Referring to the black points in Fig A1 (solar background time series) ... are these values corrected for differential aerosol transmission? If so please state, if not what is the magnitude of this effect?

Our apologies we thought we had added that detail to the text. The relative RALMO internal calculated calibration time series has not been directly corrected for the differential extinction. A sensitivity study was conducted assessing the value of $\Delta\alpha$ ($\Delta\alpha = (1 - _{N_2})/(1 - _{H_2O})$) over three years of RALMO data found that the overall impact of the τ correction amounts to a maximum of 5%, with an average of 2%. These findings are in agreement with what was found in the study by Whiteman and colleagues (1992). However, the GRUAN radiosonde calibrations were corrected for differential aerosol transmission (Hicks-Jalali et al. 2018), and the relative calibration time series is scaled by the GRUAN radiosonde external calibration.

A Raman Lidar Tropospheric Water Vapour Climatology and Height-Resolved Trend Analysis over Payerne Switzerland

Shannon Hicks-Jalali[1], Robert J. Sica[1,2], Giovanni Martucci[2], Eliane Maillard Barras[2], Jordan Voirin[3,2], and Alexander Haefele[2,1]

[1]Department of Physics and Astronomy, The University of Western Ontario, London, Canada
[2]Federal Office of Meteorology and Climatology MeteoSwiss, Payerne, Switzerland
[3]Triform SA, Fribourg, Switzerland

Correspondence to: Shannon Hicks-Jalali (shicks26@uwo.ca)

In this markup version new text is shown in blue, while removed text is shown in red.

Abstract.

Water vapour is the strongest greenhouse gas in our atmosphere and its strength and its dependence on temperature leads to a strong feedback mechanism in both the troposphere and the stratosphere. Raman water vapour lidars can be used to make high vertical resolution measurements, on the order of 10s of meters, making height-resolved trend analyses possible. Raman water vapour lidars have not typically been used for trends analyses, primarily due to the lack of long enough time series. However, the RAman Lidar for Meteorological Observations (RALMO), located in Payerne, Switzerland, is capable of making operational water vapour measurements and has one of the longest ground-based and well-characterized data sets available. We have calculated a 11.5-year water vapour climatology using RALMO measurements in the troposphere. Our study uses nighttime measurements during mostly clear conditions, which creates a natural selection bias. The climatology shows that the highest water vapour specific humidity concentrations are in the summer months, and the lowest in the winter months. We have also calculated the geophysical variability of water vapour. The percent variability of water vapour in the free troposphere is larger than in the boundary layer.

We have also determined water vapour trends from 2009 - 2019. We first calculate precipitable water vapour trends for comparison with the majority of water vapour trend studies. We detect a nighttime precipitable water vapour trend of 1.3 mm/decade using RALMO measurements, which is significant at the 90% level. The trend is consistent with a 1.38°C/decade surface temperature trend detected by coincident radiosonde measurements under the assumption that relative humidity remains constant;however, it is larger than previous water vapour trend values. We compare the nighttime RALMO PWV trend to daytime and nighttime PWV trends using operational radiosonde measurements and find them to agree with each other. We cannot detect a bias between the daytime and nighttime trends due to the large uncertainties in the trends. For the first time, we show height-resolved increases in water vapour through the troposphere. We detect positive tropospheric water vapour trends ranging from 5% change in specific humidity per decade to 15% specific humidity per decade depending on the altitude. The water vapour trends at 5 layers are statistically significant at or above the 90% level.

1 Introduction

Water vapour is the atmosphere's most important minor constituent. It plays a significant role in almost all aspects of the atmosphere, including: dynamics, circulation, radiative processes, and chemistry with other molecular and particulate species. Water vapour accounts for 60% of the greenhouse effect for clear skies (Kiehl and Trenberth, 1997) and has a significant feedback response to increases in temperature. It is, in this sense, Earth's "strongest" natural greenhouse gas. In fact, the temperature increase due to a doubling in CO_2 would be magnified to twice its predicted change from CO_2 alone if water vapour feedback is included in global climate (Held and Soden, 2000). While the majority of water vapour feedback is due to infrared absorption in the upper troposphere and lower stratosphere (UTLS), the change in lower and mid-tropospheric water vapour also has large implications for the hydrologic cycle and water vapour feedback (Dessler et al., 2013; Held and Soden, 2006). Tropospheric water vapour plays a key and complicated role in the hydrologic cycle, and it remains unclear how increases in water vapour affects precipitation rates, the chance of extreme precipitation events, or droughts. High vertical resolution water vapour profiles on the order of 100 m are extremely useful for modelers to accurately reproduce convection (Weckwerth et al., 1999). Tropospheric water vapour climatologies and trends are also critical to understanding and modeling water vapour's impact on global circulation systems. Tropospheric water vapour is predicted to increase by 7% for every 1°C increase in temperature in the lower troposphere and at even higher percentages (up to 15%) in the UTLS under the assumption that relative humidity is conserved (Held and Soden, 2000; Sherwood et al., 2010).

Measuring an atmospheric water vapour trend is difficult due to the fact that the magnitude of the water vapour concentration at a given altitude in the troposphere can change by almost more than 100% on a daily basis. Due to this huge variation, quantifying water vapour trends requires a level of precision and consistency in the measurements which is practically difficult to achieve. Early studies of water vapour trends in the troposphere were typically conducted with radiosondes (Hense et al., 1988; Ross and Elliott, 1996; Ross and Elliot, 2001). However, the trends calculated in those studies were most often not significant and their magnitudes varied greatly. Radiosonde biases have now been well documented and data sets require significant corrections and homogenization in order to be used (Elliott and Gaffen, 1991; McCarthy et al., 2009; Miloshevich et al., 2009). The GCOS Reference Upper Air Network (GRUAN) has recently made a significant step in homogenizing and correcting many modern radiosondes so that they can be more reliably integrated into trend analyses (Miloshevich et al., 2009; Immler et al., 2010; Dirksen et al., 2014).

Since radiosondes can require significant corrections, water vapour trends and climatology studies moved towards including other instruments such as microwave radiometers (Morland et al., 2009; Hocke et al., 2011), satellite-based instruments such as the microwave limb sounders (Hegglin et al., 2014; Khosrawi et al., 2018), global positioning systems (GPS, Jin et al. (2007); Wang et al. (2016)), and reanalyses models (Trenberth et al., 2005). However, the majority of these instruments can only measure columns of water vapour, or partial column profiles. Most satellite-based instruments are capable of measuring water vapour down to 300 hPa, or roughly 9 km altitude (Hegglin et al., 2013), but a few like the AIRS instrument on the

Aqua satellite can accurately measure through the troposphere down to the surface (Trent et al., 2019). However, while satellite measurements provide excellent global coverage, their vertical resolutions are typically on the order of kilometers which limits their ability to capture water vapour's large variability with altitude. Unlike these instruments, a Raman water vapour lidar has the ability to measure high resolution water vapour profiles in the troposphere (on the order of meters), as well as calculate column measurements (Melfi, 1972; Whiteman, 2003). Lidars' high vertical resolution compared to microwave limb sounders, or microwave radiometers, make them ideal instruments for studying the evolution of water vapour in the troposphere. The major drawbacks to using Raman lidars for trend measurements are that they are difficult to fully automate, and they can not measure during precipitation events or in thick cloudy conditions. As such, there have not been many lidars which have long enough data sets with enough stability to detect statistically significant trends for climatological studies. As far as we are aware, only four published Raman water vapour lidars have been run operationally over the last decade there have been only four publications on operational Raman water vapour lidars in the last 2 decades (Goldsmith et al., 1994; Dinoev et al., 2013; Hadad et al., 2018; Reichardt et al., 2012). While these lidars have been run operationally over the last 2 decades, RALMO is the only one which has presented a water vapour trends study that we show here.

Many studies have been made of precipitable water vapour (PWV) trends across Europe and Switzerland. Ross and Elliot (2001) calculated PWV trends from 1958–1995 using radiosonde measurements from the surface to 500 hPa and found highly geographically variable trends around the globe. many of which were insignificant. Their trends calculated across Europe were insignificant and differed in sign depending on the location. Trenberth et al. (2005) also calculated global PWV trends using radiosondes and models over the ocean and found mean ocean trends of 1.6 mm/decade using measurements from 1989 to 2004. Morland et al. (2009) calculated PWV trends using a radiometer, radiosonde, and model measurements in Bern, Switzerland from 1996 to 2007. Only the radiosonde measured a statistically significant trend at night of 0.49 mm/decade. Hocke et al. (2011) published an updated trend using homogenized measurements from 2 radiometers from 1994–2009. They did not detect an overall trend, but found seasonal trends of 10% change in PWV per decade in the summer and -15% per decade in the winter. Other studies around Europe using measurements from the 1980s up through 2008 report positive trends on the order of 0.3–0.5 mm/decade (Nilsson and Elgered, 2008; Ning and Elgered, 2018). Hadad et al. (2018), hereafter H2018, also measured PWV trends using ground-based GPS measurements and found a positive trend but insignificant trend of 0.42 ± 0.45 mm/decade. Recently, a water vapour lidar climatology using Raman lidar measurements was published by H2018 for measurements over France (45.75°N, 3.125°E) from 2010 to 2016. The measurements in our study have 5 more years of measurements compared to H2018. H2018 also conducted a water vapour trends analysis for their site using satellite, GPS, and ground-station measurements as well as ERA-Interim reanalyses; however, they did not include lidar measurements in their trend analysis. They found a positive, but statistically insignificant, PWV trend at Cezeaux using the GPS measurements of 0.42 ± 0.45 mm per decade, but negative trends at 925 hPa at the same location using ground-station measurements.

While there are many PWV trend analyses, there appears to be a large gap in the community in the area of height-resolved trends. The lack of height-resolved trends may be because there are very few instruments routinely operated with the capability to measure atmospheric profiles at resolutions under 1 km in the troposphere (Weckwerth et al., 1999)(Wulfmeyer et al., 2015). Some studies have measured trends at the surface as well as 850 hPa using radiosondes and models (Hadad et al., 2018; Serreze

et al., 2012). H2018 also uses measurements from the AIRS satellite to detect trends and calculates a water vapour mixing ratio trend of 0.13 g/kg per decade at 950 hPa and at 850 hPa. To our best knowledge, ours is the first study to use a Raman lidar to calculate water vapour trends at different layers in the troposphere above 850 hPa.

This study presents a tropospheric water vapour climatology and trend analysis using 11.5 years of nighttime measurements from the RAman Lidar for Meteorological Observation (RALMO) over Payerne, Switzerland. While not the first study to produce a tropospheric water vapour climatology from a Raman lidar, it is the first to use Raman lidar measurements to calculate precipitable water vapour (PWV) trends, as well as trends at different layers in the troposphere. Section 2 will discuss the measurements used in this study. Section 3 describes the methodology used to create the climatology and the calculation of geophysical variability over the ten years. Section 4 presents the trend results for both PWV and specific humidity at ten different pressure levels up to 250 hPa. Sections 5 and 6 present the Discussion, and Summary and Conclusions of our study, respectively.

2 Description of Instruments and Measurements

2.1 Raman Lidar

The lidar measurements were taken by RALMO in Payerne, Switzerland ($46.81°$N, $6.94°$E, 491 m a.s.l.). RALMO is a fully-automated operational water vapour lidar capable of reaching into the upper troposphere at nighttime and has been designed to run autonomously with minimal downtime, high accuracy, and temporal measurement-stability (Dinoev et al., 2013; Brocard et al., 2013b). Over the last ten years RALMO has been operating both day and night with an average of 50% uptime, with 40% of the downtime due to precipitation or the presence of clouds below 900 m. The other 10% of the downtime is due to routine maintenance and equipment malfunctions. RALMO uses a tripled ND:YaG laser operating at 30 Hz with a pulse power of 300 mJ. The laser has the ability to operate at 450 mJ, however, 300 mJ was chosen to maximize the lifetime of the flashlamps. RALMO includes 12 detection channels: 4 elastic or Rayleigh-scatter channels (near- and far-field), 4 pure rotational Raman, 2 Raman nitrogen (386.7 nm, digital and analog), and 2 water vapour (407.45 nm, digital and analog) channels. Raw measurements are recorded as 1800-shot profiles (approximately 1 minute) with an altitude resolution of 3.75 m for each channel from the surface up to 60 km. The raw 1 min profiles are called "scans". This work uses measurements from the pure rotational Raman channels and elastic channels to calculate nightly aerosol backscatter ratios, as well as measurements from the water vapour and nitrogen vibrational-rotational Raman channels to retrieve the water vapour mixing ratio profiles used in the climatology. The aerosol backscatter ratios are used to calculate transmissions and aerosol extinction for the water vapour retrieval.

The specific humidity profiles used for the climatology and trend analysis are determined from the RALMO measurements using the optimal estimation method (OEM) retrieval introduced in Sica and Haefele (2016) for water vapour lidars. While the OEM retrieval is in units of water vapour mixing ratio, it is finally converted to specific humidity for the purposes of this trend analysis to maintain consistency with the water vapour trend literature. 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:

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

The measurement vector is represented by $\mathbf{y}$, $\mathbf{F}$ is the forward model for the lidar, $\mathbf{x}$ is the vector containing all retrieval param-
eters, $\mathbf{b}$ is the forward function parameter vector, $\mathbf{S_a}$ is the covariance matrix of the *a priori* values, and $\mathbf{S_\epsilon}$ is the measurement covariance matrix. The cost function is consists of two terms. The first term is the weighted least-squares regression. The second is a regularization term, which provides additional information to the solution 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 ($\hat{\boldsymbol{x}}$), where the solution is then the maximum *a posteriori* solution based on the probability distribution functions and is given by:

$$\hat{\mathbf{x}} = \mathbf{x_a} + (\mathbf{K}^T \mathbf{S_\epsilon}^{-1} \mathbf{K} + \mathbf{S_a}^{-1})^{-1} \mathbf{K}^T \mathbf{S_\epsilon}^{-1}(\mathbf{y} - \mathbf{F}(\mathbf{x_a})) = \mathbf{x_a} + \mathbf{G}(\mathbf{y} - \mathbf{F}(\mathbf{x_a})), \tag{2}$$

where $\mathbf{K}$ refers to the Jacobian matrix, $\mathbf{G}$ is the gain matrix. The Jacobian matris is defined as $\frac{d\mathbf{y}}{d\mathbf{x}}$ and the Gain matrix is $\frac{d\mathbf{x}}{d\mathbf{y}}$. An in depth description of OEM theory applied to atmospheric physics can be found in Rodgers (2000). Sica and Haefele (2016) discussed the initial application of OEM to Raman lidar water vapour retrievals. Details regarding the changes to the original water vapour retrieval code can be found in Appendix 4B of Hicks-Jalali (2019). Other lidar OEM retrievals and their applications are discussed in Sica and Haefele (2015) and Jalali et al. (2018) for Rayleigh temperature profile retrievals, Mahagammulla Gamage et al. (2019) for Raman temperature profile retrievals, Farhani et al. (2019) for differential absorption lidar (DIAL) ozone profile retrievals, and Mahagamulla Gamage et al. (2020) for Raman relative humidity profile retrievals.

Processing the entire RALMO time series requires an ongoing calibration of the lidar. We have combined GRUAN radioson-
des and implemented a solar background calibration method to allow a consistent calibration across the 11.5 year time series. The continuous calibration is an internal calibration technique using 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., 2019). However, by using an internal calibration function, the lidar the trend of the measurements remain mostly independent of an external instrument. The details of this calibration are given in Appendix A. This new calibration is consistent with the GRUAN calibration of RALMO conducted in Hicks-Jalali et al. (2019).

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 is 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. Using

an OEM retrieval provided us with several advantages over the traditional ratio method (Whiteman et al., 1992; Whiteman, 2003), including a calculation of a complete uncertainty budget per profile, as well as averaging kernels. 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 pre-

5 vent a complete uncertainty budget from being produced on a profile-by-profile basis. Whiteman et al. (2012) provided robust estimates of a total uncertainty budget associated with the traditional water vapour method, however, the exact values would depend on each individual system.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

10 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

15 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 data sets such as RALMO's. Longer run times occur when more variables are retrieved or when the bin size of the retrieved profiles are small.

While we do not correct the measurements for these effects beforehand,We do not correct the RALMO measurements for the

20 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 in order to maximize our altitude coverage in the troposphere. It is not possible to simply integrate (sum) all profiles in one night due to the possible large variability in signal strength caused by clouds passing in and out of the lidar's field-of-view (FOV). Clouds can completely attenuate the nitrogen and water vapour signals depending on their composition and optical thickness. To ensure that the signal throughout the night is free of optically

25 thick clouds, we applied a cloud mask to all raw profiles between the start of astronomical twilight after sunset until the end of astronomical twilight the next morning. The cloud filter required the nitrogen signal-to-noise ratio to be at least 1 at 10 km altitude and also the background be no higher than 10 photon counts/bin/min. 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. which can be handled by our retrieval algorithm. Cirrus clouds and other aerosol

30 layers are accounted for through the aerosol extinction retrieval in the OEM algorithm (Sica and Haefele, 2016; Hicks-Jalali, 2019).

The raw nitrogen and water vapour profiles that remained after the cloud mask was implemented were summed to produce one profile for each digital and analog channel for a total of 4 profiles per night. Due to the fact that the effect of clouds will vary nightly, the measurements for each night do not contain the same number of raw profiles and thus, have varying integration

35 times. We required that all retrieved nightly water vapour profiles have at least 30 min of measurements over the course of a

night. The maximum integration time was 10 hours on a clear dry night during the winter. The minimum of 30 min was chosen to make sure we included low-signal nights in the climatology and did not bias the results towards cloud-free conditions. Thirty minutes is also the amount of time that a radiosonde typically needs to reach the tropopause at mid-latitudes. However, we do recognize that the lidar measurements in this study are naturally biased towards high pressure system conditions, since current Raman water vapour lidars require clear to semi-clear conditions in which to operate. We will compare the nighttime only PWV water vapour trends to daytime radiosonde trends in Sect. 4.2 and discuss the daytime/nighttime bias.

After applying the cloud mask and creating the temporally summed profile, the resulting profiles at 3.75 m resolution were again summed in altitude to produce 30 m altitude bins. The final input profiles to the OEM algorithm are is a single "nightly-integrated" profile with an altitude resolution bin size of 30 m. The altitude bin size of the OEM retrieval grid is 90 m, or the minimum resolution which allows the profiles to reach near the tropopause. The total distribution of lidar measurements over each year is shown in Fig. 1. The histogram shows the number of retrieved profiles per month for each year of measurements and is the maximum number of measurements available. The largest number of measurements occur in the summer, when there are more favorable meteorological conditions; however, the nights are shorter than in winter. The least number of measurements occur in December and January when there are less favorable conditions, such as a higher frequency of low clouds and precipitation. December and January are also missing a few years of measurements due to the replacement of the laser in late 2017 and early 2018. December of 2018 was extremely cloudy and there were no clear or semi-clear nights available for water vapour measurements. However, most months have at least 9 years of measurements.

It is important to note that the lidar measurements in this study are limited to only nighttime during periods of favourable weather. This selection introduces a bias in the lidar results, which we will discuss later. Despite the nighttime bias, Wang et al. (2016) noted that nighttime measurements are particularly important for understanding water vapour feedback contributions, as nighttime precipitable water vapour trends are better correlated with surface temperatures. The higher correlation at night is likely due to fewer shortwave radiation sources such as clouds, aerosols, surface evaporation, surface albedo, and a higher contribution from long wave radiation which is the dominant factor in the water vapour greenhouse effect (Dai et al., 1999; Wang et al., 2016).

2.2 Radiosondes

2.2.1 GRUAN Radiosondes

GCOS Reference Upper Air Network (GRUAN) certified radiosondes are currently the highest quality radiosonde data product available. These radiosondes have been well characterized and corrected for several biases (Dirksen et al., 2014).Unique to GRUAN radiosonde products is the calculation of of measurement uncertainties as a function of altitude Unlike many radiosonde products, GRUAN provides uncertainties for their measurements as a function of altitude which allows for better comparison between radiosondes and other instruments. The aerological station of MeteoSwiss in Payerne, Switzerland, has launched Vaisala RS92 sondes biweekly at noon and midnight UTC since October 2011 to obtain GRUAN-certified profiles of temperature and humidity. Launches are co-located with RALMO. Unfortunately, not all of the radiosondes launched at
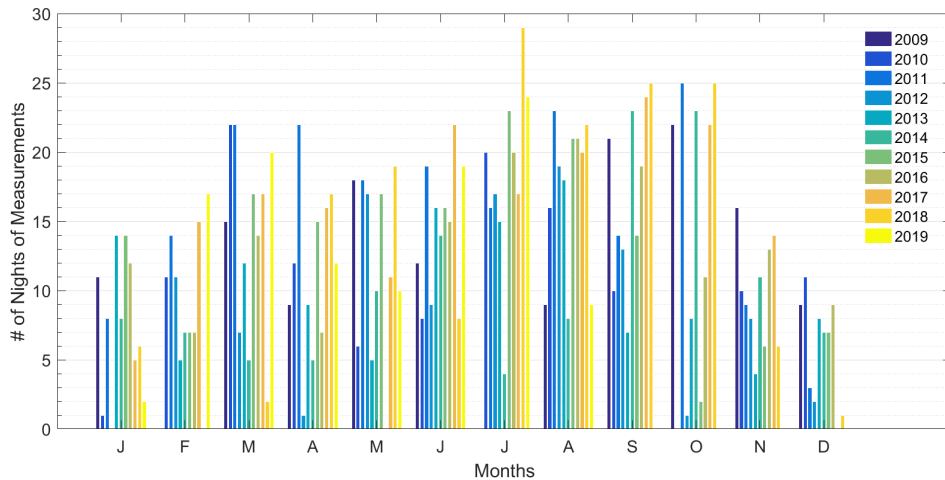
Figure 1. The total number of semi-clear nights with at least 30 minutes of profiles per month from 2009 through the middle of August 2019. Winter months have fewer measurements due to more cloud cover and precipitation. Most measurements are taken in the summer.

the station from 2009–2018 were GRUAN-compliant or GRUAN-corrected, therefore, we were unable to calculate a GRUAN radiosonde climatology or use them for trends. GRUAN-certified sondes were used to calculate the lidar calibration constants in Appendix A, as well as examine the uncertainty in the daily radiosondes (Hicks-Jalali et al., 2019).

2.2.2 Daily Radiosondes

5   Meteoswiss also launches an operational radiosonde every day at 11h00 and 23h00 UTC. The Meteoswiss operational radiosonde time series uses multiple radiosonde types: the SRS400 (Martin et al., 2006; Morland et al., 2009; Brocard et al., 2013b) from 2009–2010, SRS-C34 from 2010–Jan 2017, SRS-C50 from Feb 2017–March 2018 , and the Vaisala RS41 (Jensen et al., 2016) from March 2018 - present day. The C34 and C50 radiosondes are manufactured by MeteoLabor. The C34 uses a Sippican hygristor for humidity measurements which is quoted to have an accuracy of 2% RH (Meteolabor, 2010); however,

10   no public validation studies have been published. The C50 is the updated version of the C34, however, no public documentation exists on the C50 specifications. As these radiosondes are not GRUAN-certified we do not use them to compare with the individual lidar profiles. However, we have compared their PWV measurements to that of the coincident GRUAN-processed RS92 radiosondes. We compared the operational sonde PWV measurements to coincident GRUAN PWV measurements and found an average difference of $0.8\% \pm 3.7\%$ PWV between the two sets of sondes. The scatter in the radiosonde measurements

15   did not change between the different types of radiosondes, therefore we have not attempted any homogenizing between the different data sets. The average difference between the two sets of sondes was within the GRUAN uncertainty for PWV, therefore, we were comfortable using the operational sondes for PWV comparison with RALMO. We have also used the pressure measurements from these radiosondes to convert the retrieved water vapour lidar altitude profiles onto a standard pressure grid.

8
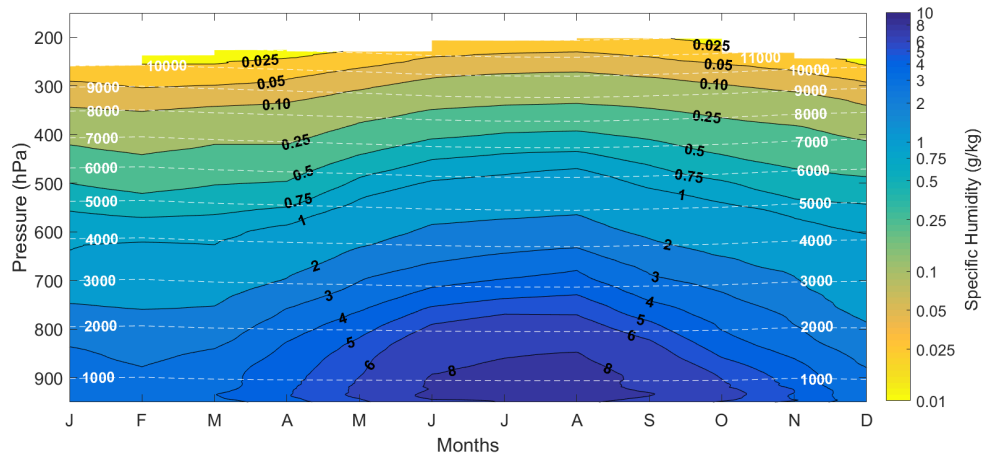
Figure 2. The monthly climatology for tropospheric water vapour from RALMO in units of specific humidity (g/kg). Contours for specific humidity are labeled in bold black. White dashed lines are the average altitudes for each pressure level and are in units of meters. Larger amounts of water vapour are seen in the summer months due to higher temperatures. The highest altitude/lowest pressure is 250 hPa during the summer months. In the winter, the lowest pressure is 280 hPa due to the lower water vapour content. Higher water vapour content is shown in darker blue colors and lower water vapour content in lighter colors.

3   A Monthly Tropospheric Water Vapour Climatology for Switzerland

Climatologies are extremely useful as they provide a baseline for daily measurement comparison and can be used as *a priori* information for forecast modeling and reanalysis models. To that end, we have calculated a monthly water vapour climatology using the RALMO measurements for Payerne, Switzerland. A monthly climatology was chosen instead of a daily climatology

5   due to the inability to retrieve enough coincident daily measurements to accurately represent a daily average.

A Raman lidar's native measurement grid is in altitude, however, models and satellites often work on pressure grids. Therefore, we have interpolated our final retrieved profiles onto a pressure grid to facilitate comparisons with other studies. The climatology on an altitude grid can be found in the supplementary materials section. We interpolated each profile onto a standard pressure grid from 950 hPa to 100 hPa using pressure and altitude data from the operational radiosondes launched

10   at midnight UTC. When surface pressures are higher than 950 hPa, the measurements are extrapolated. No surface pressure measurements were higher than 960 hPa.

Once the lidar profiles have been interpolated, all profiles in a month which pass the cost threshold of 3.5 are averaged together and weighted based on their statistical uncertainties. Due to the cloud masking applied to each individual profile, it is possible that some altitude bins will use fewer measurements than others. We required that each individual bin have

15   measurements from at least 5 different years, with at least 3 measurements per year to calculate a representative average state. This means that the minimum number of measurements is 15 per pressure bin. Figure 2 shows the resulting specific humidity climatology with respect to pressure. The corresponding average altitudes for each pressure are shown as white contour lines.

The climatology shows the expected seasonal cycle for water vapour with high concentrations (darker blues) in the summer and lower concentrations in the winter. The spring months are slightly drier than the fall months, with specific humidity values only starting to increase in April and high summer concentrations lasting through September. On average, the lidar is able to retrieve measurements consistently (at least 15 profiles per bin over 5 years) up to 280 hPa (roughly 10 km). However, in the summer months, the climatology reaches up to 250 hPa (roughly 12 km) due to higher lidar SNRs from the higher water vapour content.

The average statistical uncertainties for the climatology are consistent across all seasons, with the exception that the summer months and September show smaller uncertainties at lower altitudes due to their larger SNRs (Fig. 3). The average statistical uncertainties are calculated by taking the average of the uncertainties for all the profiles used in each month. The statistical uncertainties are presented as the uncertainty in the water vapour mixing ratio, and not the uncertainty in specific humidity. The OEM retrieval is in units of water vapour mixing ratio, and while we have converted the final retrieval profiles to units of specific humidity, even at the highest concentrations of water vapour the difference between the two units is less than 2%. Therefore, the uncertainties calculated for the water vapour mixing ratio are comparable to the uncertainties for the converted specific humidity profile. 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 the remaining 23% had uncertainties between 20 and 25%. The statistical uncertainty is less than 1% in the boundary layer but increases up to about 14% uncertainty at the lowest pressures. The average uncertainties in the winter are slightly higher in the boundary layer due to the drier winter air.
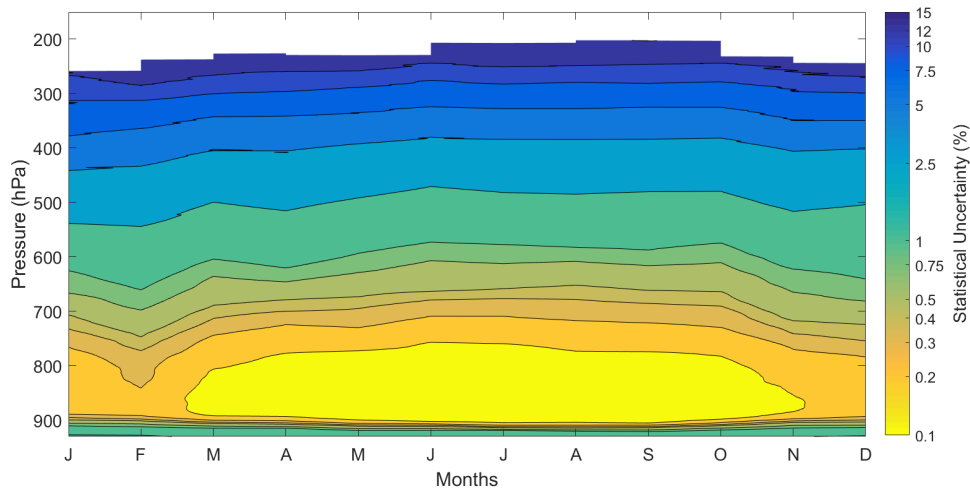
Figure 3. The average statistical uncertainties (represented as fractional percent uncertainties) for each month as a function of pressure. The highest statistical uncertainty is 14% of the water vapour mixing ratio between 275–250 hPa or 11–12 km.

In this study we refer to the individual retrieval uncertainties in water vapour mixing ratio due to uncertainties in non-retrieved forward model parameters as parameter uncertainties (Fig. 4). Parameter uncertainties are calculated via Gaussian uncertainty propagation for each retrieval parameter. Figure 4 shows the parameter uncertainties' effect on water vapour mixing ratio calculated for each retrieval, including: the calibration constant, Ångstrom exponent, NCEP air density, aerosol extinction

5 profile, overlap, and the Rayleigh cross-section. Details regarding the covariances of the system parameters can be found in Appendix 4B of Hicks-Jalali (2019). Every uncertainty profile used in the climatology is shown in gray dashed lines and the average uncertainty for each component is the black dashed line. Note that systematic uncertainties are on a different height scale for the aerosol extinction profile and overlap components. This choice is because these forward model parameters are also retrieval parameters for part of the profile. Overlap is retrieved from the ground to 400 hPa, therefore the uncertainty in

10 overlap on the retrieved water vapour mixing ratio is only calculated above 400 hPa. The opposite situation occurs for the retrieved aerosol extinction; it is retrieved above 400 hPa and therefore its uncertainty on the water vapour mixing ratio is only calculated below 400 hPa. The largest systematic uncertainty, which dominates the retrieval, is the 5% calibration uncertainty. Sica and Haefele (2016) originally assigned a value of 5% which was later confirmed by the study conducted in Hicks-Jalali et al. (2019). The residuals of the smoothing spline fit to the combined calibration time series also showed an average of

15 5% variation in the residuals (Appendix A). The second largest component is the uncertainty due to the assumption of the aerosol Ångstrom exponent. The OEM retrieval assumes that the Ångstrom exponent does not change with altitude, which can cause larger uncertainties in clouds where our assumed value may not be appropriate. We used a mean climatological value of 1.5 ± 0.5, which results in an average uncertainty on the order of 1% in the water vapour mixing ratio. In 4 out of the 1300 nights used in this study, the uncertainties caused by the uncertainty in the Ångstrom exponent were larger than 5% due to the

20 Ångstrom exponent choice being a poor fit in a persistent cloud. The uncertainty due to NCEP model air density is slightly

11
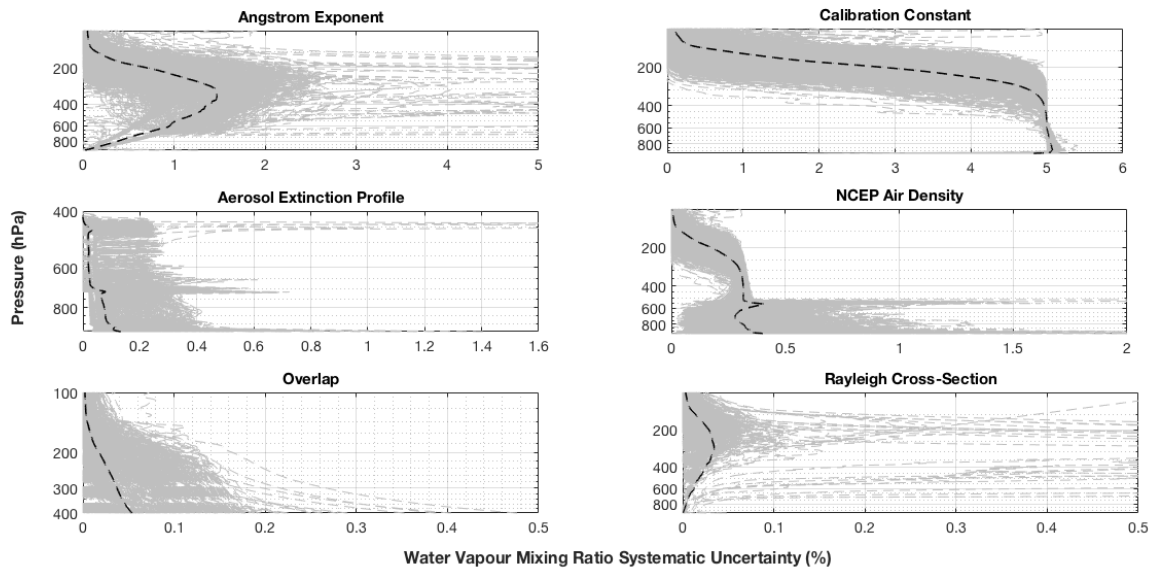
Figure 4. The water vapour mixing ratio systematic uncertainty budget for each night used in the water vapour climatology (gray dashed lines), and the average uncertainty contribution from each component (black dashed line). The largest contributors are the Ångstrom exponent and the calibration constant. Four nights out of 1300 had Ångstrom exponent uncertainties larger than 5% due to our assumed value not fitting with persistent clouds. The average uncertainty contribution from the NCEP air density is .25%. All other uncertainties contribute less than 0.1% on average.

higher than the Ångstrom exponent uncertainty at altitudes below 2 km, but with an average uncertainty of around 0.25% for all altitudes. All other systematic uncertainties including extinction, overlap, and the Rayleigh cross section are less than 0.1%. However, the Rayleigh cross section uncertainties also had 4 nights which exhibited uncertainties larger than 0.5%. The major improvement to the water vapour retrieval since Sica and Haefele (2016) is the retrieval of overlap from the surface to 6 km (if there are no clouds below 6 km). In Sica and Haefele (2016), overlap was not a retrieval parameter and added another 7 - 10% uncertainty below 3 km.

We also compared the lidar PWV measurements to the PWV column from the Payerne daily radiosondes during the same time period as the height-resolved climatology. The radiosonde measurements are systematically larger than than the lidar measurements, likely due to the fact that the lidar is not able to take measurements in the first 100 m from the surface. The lidar PWV climatology agrees with the operational radiosonde PWV climatology within their respective uncertainties (Fig. 5). The uncertainties for the climatology were calculated using the standard deviation of all measurements in one month.
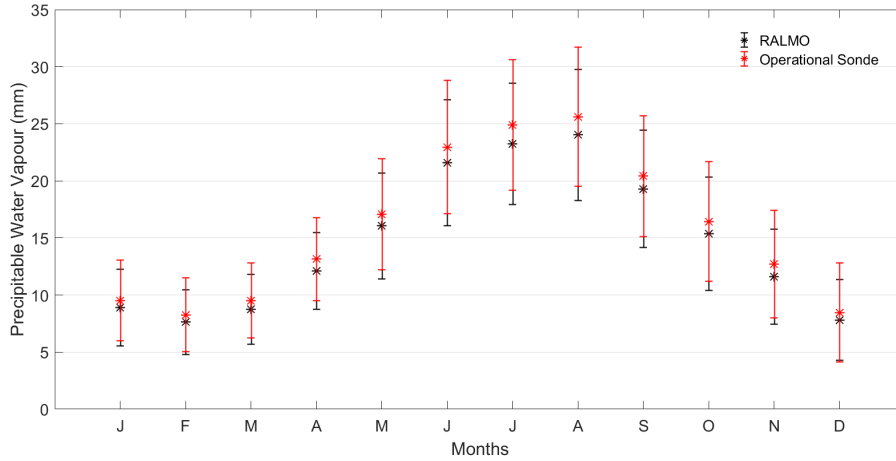
Figure 5. The PWV climatology of both the lidar and the operational radiosondes. The error bars are the 1σ standard deviation of all of the measurements in each month. The climatology was calculated using the same dates for both the radiosonde and the lidar measurements.

The PWV climatology shows the same behavior as the profile climatology already shown in Figure 2, with the largest amount of water vapour in the summer months and smallest amounts in the winter months.
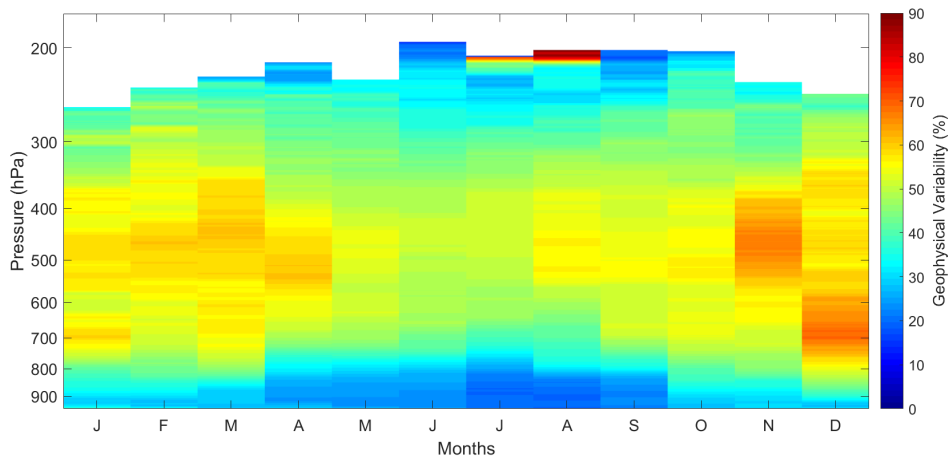
3.1 Geophysical Variability

Water vapour concentrations change by up to four orders of magnitude through the troposphere. The troposphere is also a region of many dynamic processes, including the hydrologic cycle as well as general global circulation. These processes have a direct impact on water vapour variability and make it difficult to measure. The abundance of measurements used in this study make it possible to estimate water vapour's variability as a function of altitude. The geophysical variability ($\sigma_{geo}(z)$) of the specific humidity can be estimated by taking the standard deviation of all profiles in one month ($\sigma_w(z)$) and removing the measurement variability, which is dominated by the statistical uncertainty ($\sigma_{stat}(z)$, Eq. 3, Argall and Sica (2007)). The remaining variability should then be physical in nature.

$$\sigma_{geo}^2(z) = 100\frac{\left(\sigma_w^2(z) - \sigma_{stat}^2(z)\right)}{\langle q(z) \rangle^2} \tag{3}$$

The percent geophysical variability for water vapour was calculated using Eq. 3 in which $\sigma_w^2(z)$ is the variance of all monthly profiles, $\sigma_{stat}^2(z)$ is the average of the statistical uncertainty variances of the retrievals, and $\langle q(z) \rangle$ is the climatology. Figure 6 is then the geophysical variability in units of percent variability from the mean state.

The smallest percent variability is in the boundary layer, typically in the summer. However, as this is a nighttime and primarily clear-weather climatology, the boundary layer should be stratified and either neutral or stable at these times. Therefore, Fig. 6 would represent the variability of the boundary layer during nighttime and primarily high-pressure conditions.

Above the boundary layer, from 700 hPa in the winter/spring/fall and up to 350 hPa there is much higher variability present. The largest variabilities of 60% or higher are located between 550 - 400 hPa during the spring and fall, and between 750 -

13

Figure 6. The monthly geophysical variability of water vapour measured by RALMO as a function of pressure. The largest amount of variability can be seen in the free troposphere between 600 and 400 hPa in the summer, and from 800 to 450 hPa in the winter.

600 hPa in December. The latter region in December is likely due to sampling bias since December had half the measurements as the summer months. However, the high variability regions between 600 hPa and 400 hPa are interesting and could represent dynamic processes at work. The first, and most straightforward, explanation for the high variability at these levels is the presence of mid-level layers of clouds or aerosols. These mid-level layers are often semi-transparent to the laser and as such they are not

5   removed or are only partially removed by the SNR mask at the pre-processing stage. The second possible explanation for the high variability is based on tropospheric dynamics. The region from 600 to 400 hPa is a region of active mixing of air from the upper troposphere or stratosphere and the surface. Simultaneously, planetary waves transport air horizontally from different air masses at these heights. Therefore, we would expect the region from 600 hPa to 400 hPa to exhibit a high amount of variability.

       We compared the lidar-measured geophysical variability to the water vapour geophysical variability measured by the op-

10   erational radiosondes. All available daytime and nighttime operational radiosondes were used to calculate the geophysical variability. The total value of the percent geophysical variability is only an estimate because we do not know the uncertainty of the daily radiosonde's humidity measurements. The maximum uncertainty reported by the GRUAN radiosondes in the troposphere is around 10% of the mixing ratio measured. Therefore, we assume a constant uncertainty in the operational radiosondes measurements of 10%, although this would be a lower limit of the uncertainty of the routine sondes because they have not been

15   corrected. The variability is then calculated in the same way as for the lidar (Fig. 7).

       The radiosondes measure similar behavior in the variability to the lidar, however, the magnitude of the variability is roughly 20% larger than the lidar's. As discussed earlier, half of this difference is likely due to underestimating the uncertainty in the operational radiosondes. However, it could partially be due to the fact that more measurements are available from the radiosonde than from the lidar. When comparing November and December profiles, it can be seen that the regions of large

20   variability in November and December in the lidar variability are likely due to sampling bias and not physical. However, it

14
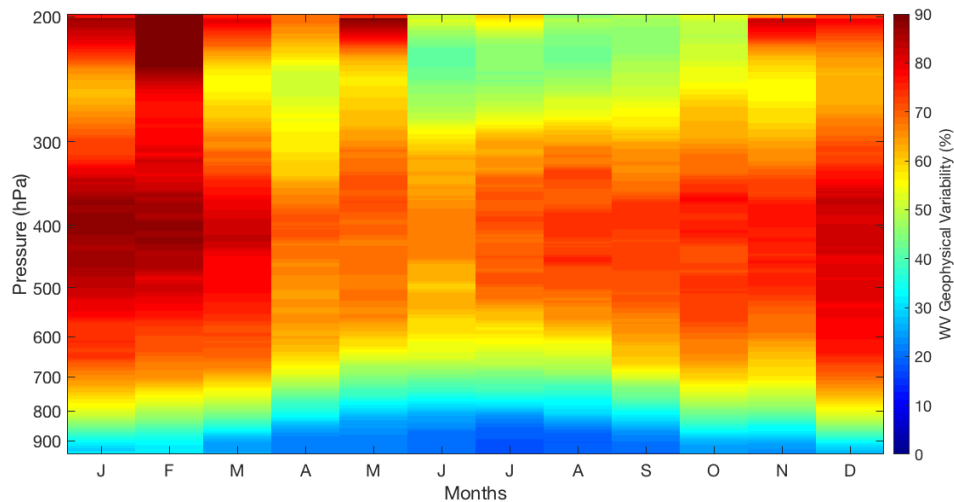
Figure 7. The water vapour geophysical variability estimated using all available daytime and nighttime operational radiosondes as a percentage and function of pressure. The radiosondes see the same behavior as the lidar, and the distribution of the variability is the same, although the magnitude is generally 20% larger. The larger magnitude is possibly due to the underestimation of the daily radiosonde uncertainties.

is encouraging to note that the same physical structure is seen in both instruments. This result would suggest that the large variability in the troposphere is not due to clouds or layers affecting the climatology, but likely due to the increased mixing of air between 600 to 400 hPa in the summer, and 700 to 450 hPa in the winter. In high pressure systems, the air from the stratosphere and high regions of the troposphere converge which causes air to subside at the surface (Holton, 2004). The colder

5   air in the stratosphere and lower troposphere is usually drier. This exchange of moist air and dry air in the free troposphere could explain the large variability. It is interesting to see that adding the daytime sondes to the variability calculation does not change the overall behaviour of the water vapour in the troposphere. We explored looking at the difference between the variability including daytime sondes and not including the sondes and found almost no difference in the variability when the daytime sondes were added.

10   The radiosondes and the lidar measure more variability in the winter months than in the summer months.The difference in variability between the two seasons could be due to the fact that the cold air in the winter cannot hold as much water vapour as the warm air in the spring and summer due to smaller saturation vapor pressures in the winter and larger in the summer. The smaller average concentrations of water vapour in the winter to a larger percent variability. European winter weather is highly influenced by the polar jetstream. An increase in the variability of water vapour during these months could be caused by an

15   increase in the exchange of air between the Arctic and the mid-latitudes.

15

4 Deriving Trends

There have been several tropospheric water vapour trend studies conducted over the course of the last few decades, only a few of which are cited here (Hense et al., 1988; Ross and Elliott, 1996; Ross and Elliot, 2001; Morland et al., 2009). Whiteman et al. (2011a) investigated the potential for water vapour Raman lidar trend calculations and carefully estimated the thresholds needed to calculate statistically significant trends in the upper troposphere. They found that with no statistical uncertainty, one would need between 10 to 12 years of daily measurements to calculate trends at 200 hPa due to water vapour's naturally large variability at that pressure. We have 11.5 years of measurements and we should be able to detect tropospheric trends at higher pressures than 200 hPa due to the higher concentration of water vapour in the free troposphere and boundary layer and smaller variability than those at 200 hPa. and are therefore in the realm of possible detection at the 95% level; however, whether or not we can detect trends depends on the variability of the atmosphere at the levels we choose and the magnitude of the trend itself. As we will show, the variability of the atmosphere in the free troposphere will determine our ability to detect trends at the 95% level. According to Weatherhead et al. (1998), trends of at least 5% per decade should be calculable in 10 years provided the uncertainty of the measurementsstandard deviation of the noise in the trend residuals is less than 6% and the autocorrelation of the trend fit residuals is less than 0.4 (Table 1 of Weatherhead et al. (1998)). Therefore, the lowest pressure we should be able to calculate trends of at least 5% is around 350 hPa (11 km) provided the autocorrelation is small and we have relatively little noise in our trend residuals.

4.1 Trend Estimate Methodology

The bootstrap method of Gardiner et al. (2008) makes use of a least-squares regression for a model which includes both the linear trend and the annual and semi-annual oscillations via a Fourier function (Eq. 4):

$$
\begin{aligned}
f(t, a, \mathbf{b}) = at + b_0 &+ b_1 \cos(2\pi t) + b_2 \sin(2\pi t) \\
&+ b_3 \cos(4\pi t) + b_4 \sin(4\pi t) \\
&+ b_5 \cos(6\pi t) + b_6 \sin(6\pi t),
\end{aligned}
\tag{4}
$$

where a is the trend in units of either g/kg per year or mm/year, the vector $\mathbf{b}$ is comprised of all the seasonal variation coefficients ($\mathbf{b_n}, \mathbf{n = 0, 1, ..6}$) as well as the offset ($\mathbf{b_0}$), and t is time. Gardiner et al. (2008) found that, for their study, it was not necessary to continue beyond the third order of the Fourier series, as doing so no longer decreased the RMS of the residuals. We also found no difference in the residual RMS by adding terms beyond the third order in the Fourier series. Indeed, the third order could probably have been removed as well, however, we have kept it to maintain consistency with other studies.

We calculated trends using monthly averages of nightly averaged measurements from RALMO as well as the nightly radiosondes. Monthly averages were used to reduce the autocorrelation of the measurements and increase the statistical signifi-

cance of the trends. The statistical significance of the trends was determined by using the uncertainty equation from Weatherhead et al. (1998) which includes the autocorrelation coefficient of the residuals:

$$\sigma_a = \frac{\sigma_N}{n^{3/2}} \sqrt{\frac{1+\phi}{1-\phi}}. \tag{5}$$

The uncertainty of the trend, σ_a, is determined from σ_N, the standard deviation of the residuals from the least squares fit, n, the number of years of measurements (in our case 11.5), and ϕ, the autocorrelation coefficient at lag 1. All uncertainties presented in this study are $2\sigma_a$ to determine if the corresponding trend is statistically significant at the 95% level.

As shown in Fig. 1, some months of data are missing from the full time series. However, there were no successive gaps in the monthly-averaged time series. We linearly interpolated the monthly-average time series before calculating the trend to fill the 6 gaps in the missing months and to calculate meaningful autocorrelation coefficients.

4.2 Precipitable Water Vapour Trends

We first calculated tropospheric precipitable water vapour (PWV) trends using RALMO and the daily radiosonde measurements to compare with the majority of previous studies where height-resolved measurements were not available. Precipitable water vapour values were calculated by integrating water vapour densities between the daily station surface pressure as measured by the daily radiosonde and 280 hPa to maintain the same pressure range over the entire year. Table 1 shows the value of the trends for the 30-day averaged RALMO and daily radiosonde nighttime measurements from January 2009 to August 2019. The PWV measurements, the function fits, and the trend lines are shown in Fig. 8. The radiosonde trends were calculated in two ways: the first used only the nights from the lidar measurements and the second used all available nighttime radiosonde measurements. We calculated four radiosonde trends to compare with the RALMO PWV trend. The first two methods used only nighttime data, but the first limited the data to the same nights as the radiosonde and the second used all available nighttime radiosondes. They are shown in rows 1 and 2 of Table 1, respectively. The difference between the two nighttime radiosonde trend methods could represent the bias from using only semi-clear nights during clement weather. However, the bias between the two radiosonde nighttime trends is more likely caused by the number of points used in each trend and the uncertainty in the radiosonde measurements. Roughly 25% of the possible nights are used in the trend analysis for the coincident lidar dates and as the radiosondes do have a larger and uncharacterized uncertainty it is more likely that the large trend value is due to a larger scatter in the radiosonde measurements. Given that the lidar trend is much closer to the radiosonde trends which use all available nights, including the radiosonde trends calculated using daytime and nighttime measurements, the large radiosonde trend using only nights consecutive with the lidar is more likely due to random error and not a difference in weather. We also calculated daytime radiosonde trends using all available daytime measurements as well as a combined trend using all available daytime and nighttime measurements (rows 3 and 4 of Table 1). The daytime radiosonde trends differ from the nighttime radiosonde trend using all available nights by 0.1 mm/decade and they agree within their 1-sigma uncertainties.

The lidar and radiosonde precipitable water vapour trends have varying levels of significance. The RALMO trend is significant above 90% while the radiosonde trends are significant above 95% (Table 1). The trend calculated using all available

17
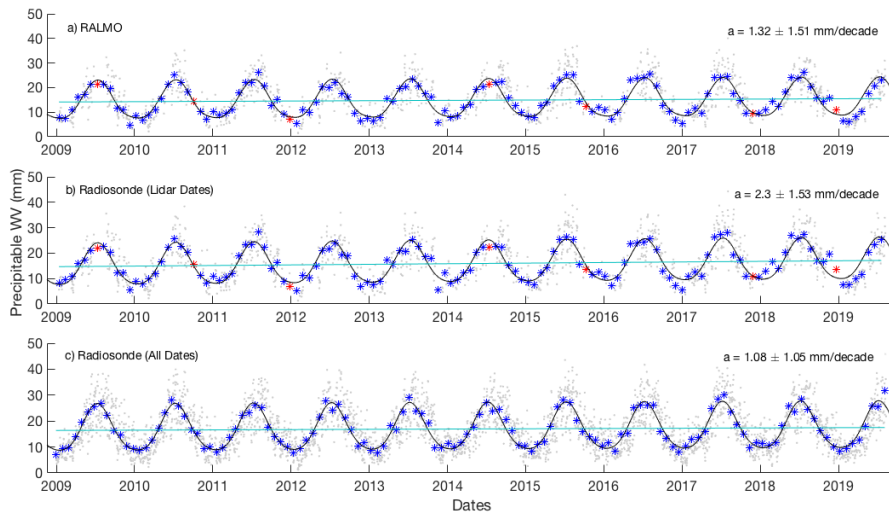
Figure 8. The nighttime precipitable water vapour measurements taken by RALMO and the daily radiosondes. For all subplots, the nightly measurements are the grey dots, 30-day averages are the blue stars with interpolated points in red, the fitted function is in black, and the trend line is in light blue. a) Ralmo PWV measurements, b) Radiosonde PWV measurements on coincident dates with lidar measurements, c) Radiosonde PWV measurements for all available dates.

Table 1. Table of precipitable water vapour trends calculated using RALMO and daily radiosonde monthly averaged measurements from January 2009 - August 2019. Trends are presented in units of mm per decade (mm/dec), percent change from the mean per decade (%/decade), and percent change in water vapour per °C. The trend values in column 3 were calculated by dividing the water vapour trend in column 2 by the surface temperature trend of 1.38 ± 1.41 °C per decade. Trend uncertainties are $2\sigma_a$ values and in in units of mm per decade and are calculated using Eq. 5.

Measurement Series	Trend (mm/dec)	Trend (%/dec)	Trend (%/°C)	Trend Uncertainty (mm/dec)	ϕ	Significance Level
RALMO	1.32	8.85	6.41	1.51	0.21	90%
Daily Radiosonde (Lidar dates)	2.31	14.45	10.59	1.53	0.16	95%
Daily Radiosonde (All dates, night only)	1.08	6.36	4.94	1.05	0.09	95%
Daily Radiosonde (All dates, day only)	1.18	7.20	5.27	0.97	0.05	95%
Daily Radiosonde (All dates, day and night)	1.19	7.20	5.27	0.98	0.07	95%

nighttime radiosonde measurements is slightly smaller than the RALMO PWV trend, but slightly larger than previously calculated trends for Payerne and the region (Morland et al., 2009; Wang et al., 2016; Nyeki et al., 2019).

The Clausius Clapeyron relationship suggests that the water vapour in the atmosphere should increase by roughly 7.5% per 1°C, or more depending on the initial temperature (Held and Soden, 2000). Therefore, such large PWV changes should be correlated with proportionally large increases in temperature. We used the temperature measurements at 500 m (950 hPa) a.s.l. from the daily radiosonde to compare with the lidar water vapour trend because the lowest available measurements are at 550 m. We assume that the lidar PWV measurements should correspond best with the temperatures at that altitude. The same nights used in the lidar and radiosonde water vapour trend analysis were used for the radiosonde temperature measurements. The 950 hPa temperature trend was statistically significant at 95% with a slope of 1.38 ± 1.41 °C per decade. The temperature trend we measure is larger than that measured by Morland et al. (2009) and Nyeki et al. (2019). However, the percent change in water vapour per degree for all PWV trend values is between 5 - 10%, which is consistent with the expected change assuming relative humidity is conserved. The differences between our temperature trend measurements and other studies will be discussed further in Sect. 6.

4.3 Height-Resolved Trends

Here we show height-resolved water vapour trends measured at 11 layers of the troposphere from 950 hPa to 250 hPa. Each layer is an average of measurements from ± 10 hPa from the center (e.g. measurements at 900 hPa are an average between 890 - 910 hPa). The layers were averaged to reduce the variability due to choosing a single pressure level. We calculated the trend for each layer using monthly averages of the time series. We calculated the specific humidity height-resolved trends using the same bootstrapping method that was used for the PWV trends. An example of the trend at 800 hPa is shown in Figure 9. The blue asterisks are the monthly averages, with the interpolated points shown in red. The nightly points are shown in light gray. The trend and seasonal fit are in light blue and black, respectively.
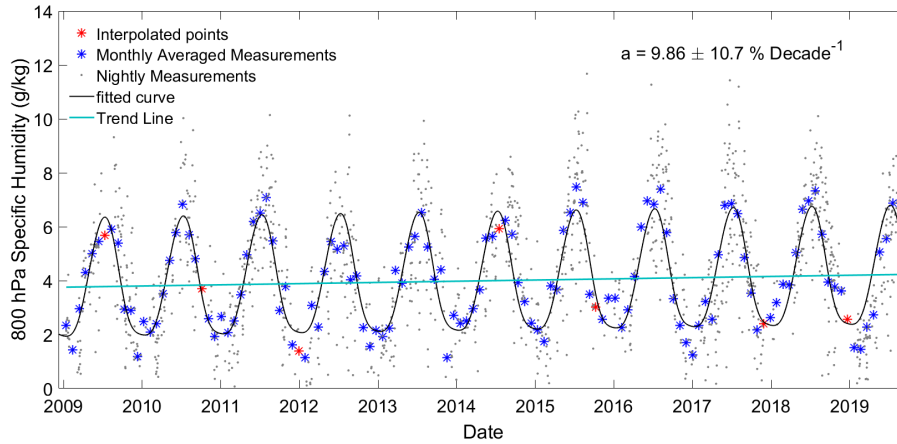
Figure 9. An example lidar specific humidity trend at a fixed pressure level of 800 hPa showing the seasonal fit (black) and the linear trend line (light blue). Nightly measurements are in gray, monthly averages are blue asterisks with the 7 interpolated monthly points in red. The trend is calculated over the series which includes the interpolated points.

The trends at each layer are shown in Table 2. The trend at 950 hPa is statistically significant at the 95% level, while trends at 800, 700, 600, and 350 are significant at the 90% level. The trend at 400 hPa is statistically significant at the 80% level. The rest of the trends have significance levels of less than 80%. The trends in Table 2 are presented in absolute units per decade, percent change from the mean per decade, as well as percent change from the mean per degree. The uncertainties are presented

5   as $2\sigma_a$ in units of percent change from the mean per decade. All trends were positive, but varied between 3 and 16% per decade depending on the layer. The trend at 950 hPa is statistically significant, but the noise in the residuals is less than the noise at 900 hPa.

To calculate the change in the specific humidity per degree, we also calculated layered temperature trends using the daily radiosondes. The radiosonde temperature measurements were also averaged over the same pressure ranges as the specific

10   humidity trends and averaged monthly. The temperature trends are presented in Table 3. All available nights were used to calculate the temperature trends. We found all temperature trends to be positive, with the largest 1.38°C/decade at 950 hPa. All of the temperature trends were statistically signficant at the 95% level. The temperature trends are larger by about 0.4 - 0.8 °C/decade than what was reported in Morland et al. (2009) and Brocard et al. (2013a). Interestingly, dividing the height-resolved water vapour trends by the temperature trends does not produce the expected 7%/°C relationship which was seen in

15   the precipitable water vapour trends, with the exception of the trend at 950, 300, and 250 hPa. Instead, we see increases on the order of 10 - 17%/°C, between 800 and 300 hPa. The rest of the layers present humidity changes with temperature that would indicate that relative humidity may not be conserved at the individual layers. The PWV trends are likely dominated by the surface water vapour trend, hence why it PWV and 950 hPa trends are similar.

Table 2. Table of RALMO specific humidity trend calculations for each pressure layer. The first column is the specific humidity trend in units of g/kg per decade, the second column is the specific humidity trend in percent per decade, the third column is the specific humidity trend in units of percent change per degree. The fourth and fifth columns are the uncertainties of the trend in units of percent per decade, and the autocorrelation value at lag 1. Trends marked with a *, **, and *** are statistically significant at the 80%, 90%, and 95% level, respectively.

Pressure (hPa)	Trend (g/kg Decade^{-1})	Trend (%/Decade)	$2\sigma_a$ (%/Decade)	ϕ
950***	0.74	12.11	10.25	0.25
900	0.05	0.84	8.15	0.23
800**	0.44	9.83	10.45	0.22
700**	0.35	13.94	14.36	0.18
600**	0.22	15.85	16.92	0.19
500	0.03	4.80	16.19	0.26
400*	0.03	10.87	15.79	0.23
350**	0.02	12.59	14.07	0.25
300	0.007	7.08	14.01	0.23
275	0.002	3.19	11.6	0.22
250	0.002	3.55	11.48	0.25

Table 3. Table of temperature trend calculations for each pressure layer using the operational radiosondes. The second column is the temperature trend in units of degrees per decade, the third column is the $2\sigma_a$ trend uncertainty in units of degrees per decade, the fourth column is the autocorrelation coefficient at lag 1, and the fifth column is the percent change in specific humidity per degree C. Trends marked with a *, **, and *** are statistically significant at 80, 90, and 95% level respectively.

Pressure (hPa)	Trend (°C Decade^{-1})	$2\sigma_a$ (°C Decade^{-1})	ϕ	Trend (%q °C^{-1})
950***	1.38	1.02	0.14	8.80
900***	1.20	1.02	0.24	0.68
800***	1.05	1.02	0.21	9.39
700***	0.99	0.91	0.19	14.09
600***	0.90	0.86	0.19	17.64
500***	0.91	0.84	0.15	5.33
400***	0.95	0.84	0.15	11.44
350***	1.06	0.79	0.14	11.88
300***	1.06	0.61	0.13	6.67
275***	0.95	0.52	0.13	3.34
250***	0.52	0.55	0.17	7.10

5 Summary

We use the calibrations done by Hicks-Jalali et al. (2019) (Appendix A) in conjunction with the updated water vapour OEM retrievals from Sica and Haefele (2016) to reprocess 11 and a half years of water vapour measurements from RALMO (January 2009 - August 2019). We calculated a monthly tropospheric climatology from the reprocessed measurements which shows that we have more water vapour in the summer than winter, as expected with the larger summer temperatures. The climatology also reaches lower pressures during the summer when more water vapour is present. The statistical uncertainty at the lowest pressures is 14% of the water vapour content. The total uncertainty at those pressures is on average 20% when considering the systematic uncertainties. Uncertainties at lower altitudes are dominated by the uncertainty in the calibration constant and are a constant 5% until the statistical uncertainty is larger (roughly above 350 hPa).

We also calculated the geophysical variability of water vapour in the troposphere. The geophysical variability is a selective representation since the measurement intervals in this study were restricted to relatively clear nighttime conditions with most low-level clouds and optically thick mid-level clouds removed from the reprocessing. Therefore, the variability that is presented is representative of the nighttime structure during clear to semi-clear conditions. The same variability study was done using the nighttime operational radiosondes and no restrictions were placed on the presence of clouds or precipitation. The radiosonde and the lidar studies agree in the boundary layer and both datasets show that the nighttime boundary layer does not exhibit a large percent variability. The magnitude of the variability in the radiosonde dataset is larger than in the lidar study in the free troposphere, but they see similar behavior in the variability. This bias is likely due to the fact that the water vapour variability measured by the radiosonde can only be estimated since their measurement uncertainties are not reported. We are likely underestimating the uncertainties in the radiosonde which would create the difference in the variabilities measured by the two instruments.

Most previous tropospheric water vapour trend studies focus on precipitable water vapour. Therefore, we calculated PWV trends using the bootstrap method for both the radiosonde and RALMO measurements. The PWV trends calculated by the lidar and the radiosondes were 1.32 ± 1.51 mm/decade and 2.31 ± 1.53 mm/decade, respectively when using the same dates. When the radiosonde trend includes dates in which clouds or precipitation are present, the PWV trend decreases to 1.08 ± 1.05 mm/decade. The RALMO PWV trend is statistically significant above 90%. Both radiosonde PWV trends are significant to 95%. These trends are large and correspond to 8.9% per decade from the mean for the lidar and 15.5% per decade from the radiosonde. These large percentages can be explained by correspondingly large temperature changes. We calculated a surface temperature trend using the measurements from the radiosonde and found an increase of 1.38°C per decade from January 2009 - August 2019. The PWV trends are then consistent with the fact that a change in 1 degree Celsius will result in a roughly 7% change in water vapour at atmospheric temperatures of around 300 K, assuming relative humidity is conserved Held and Soden (2000).

Lastly, we investigated the specific humidity trends at 11 pressure layers in the troposphere, each 20 hPa in thickness. No trends significant to 95% were found at 900 and 500 hPa, or above 350 hPa. However, all other layers exhibited positive trends with statistical significance at or above 80%. The surface specific humidity trend followed the expected Clausius Clapeyron

relationship and had a trend of 8.8% change in specific humidity per degree Celsius. All other trends were larger than 7% per °C, with some as large as 17% per °C (column 5 of Table 3). We compare the trends determined in this study with other studies in the following section.

6 Discussion

We have calculated a lidar climatology of the water vapour distribution above Payerne, Switzerland using measurements from 2009 to 2019 from the RALMO lidar, as well as a precipitable water vapour climatology. RALMO is the onlyone of a few lidars which has produced a published high vertical resolution water vapour climatology of the troposphere with 11.5 years of consistent measurements. , the climatology in Hadad et al. (2018) (H2018) is the only other ground-based Raman lidar published climatology (410 m, 45.77°N, 2.96°E). The climatology in H2018 shows a wet bias of roughly 2 g/kg at the surface in September and October compared to RALMO. This difference is likely due to the fact that they only have 6 years of measurements from 2010–2016. The remaining months show good agreement between the RALMO and H2018 lidar climatologies. Our RALMO climatology agrees very well with the H2018 Atmospheric Infra-Red Sounder (AIRS) on the Earth Observing System (EOS) climatology which had more measurements from 2002–2017. The AIRS water vapour climatology was calculated for the 100 km radius around Cezeaux, France. The differences between the AIRS and RALMO climatologies are less than 1 g/kg.

H2018 also calculates a standard deviation for their climatology. However, it is important to note that this is not the same as the geophysical variability that we have calculated, as they do not subtract the variability due to their measurement noise. Additionally, they do not look at the relative variability, but the absolute. They show large standard deviations in the boundary layer and lower troposphere and very small standard deviations in the upper troposphere. However, dividing their standard deviation by their climatology produces variability measurements comparable to ours of around 20 to 30% in the boundary layer and between 40 to 80% in the free troposphere. H2018 states that they see more variability in the boundary layer, however, we would argue that this is because they are looking at the absolute value of the variability, and not the percentage. The free troposphere actually has larger variability, since it has much smaller concentrations of water vapour. In summary, both the H2018 and our variabilities are consistent with each other.

The RALMO climatology is a nighttime and clear-sky only climatology. The AIRS climatology in H2018 is not cloud-filtered (or at least no mention is made of any cloud-filtering), appears to use both day and night measurements, and differs from the RALMO climatology by less than 1 g/kg. The question is whether or not a nighttime climatology for Payerne can serve as a true representation of the average state. Morland et al. (2009) and Hocke et al. (2017) calculated the diurnal PWV cycle for Bern using the TROWARA instrument and GNSS measurements. Hocke et al. (2017) found that the average PWV diurnal amplitude was 2% of the monthly mean. In other words, the total water vapour content on average changes by 2% over the course of the day. The maximum change in water vapour was 4% over the course of the day in June, and the minimum was less than 1% in the winter months. While PWV cannot accurately represent the changes at individual layers, it does serve to show that as a whole the amount of water vapour does not change significantly over the course of the day. Perhaps the more

important fact is that the phase of the diurnal cycle is such that the maximum amount of water vapour peaks around 20hrs LST and then decreases to reach a minimum at 8hrs the next morning. This would suggest that the average nighttime profile over a month is probably a good representation of the average water vapour content and therefore a nighttime only water vapour climatology is probably representative of the average water vapour content. It is important to note that this is the case for Payerne, but would not necessarily apply to other locations.

It is difficult to compare the RALMO profile climatology with anything other than the operational radiosondes. Unfortunately, the routine radiosondes do not provide an uncertainty product, and therefore are not ideal for water vapour profile analysis. In the coming years, it should be possible to do a profile climatology comparison with the GRUAN radiosonde product from Payerne which may soon become available on a daily basis. However, there is currently not enough of a sample to accurately create a representative average, as the results would most certainly be biased towards the last two years of measurements. Most satellite climatologies of water vapour only extend down to 300 hPa (Hegglin et al., 2013). The AIRS instrument on Aqua is capable of reaching into the lower troposphere, and we have compared our results with the AIRS climatology produced in H2018. They are more suited to climatology comparisons of the upper troposphere and lower stratosphere.

Most water vapour trend studies of the troposphere focus on PWV measurements or surface humidity measurements. To compare with the literature, we also calculated precipitable water vapour trends using both RALMO and operational radiosonde measurements and found statistically significant trends of 1.3 ± 1.53 mm/decade (8.8%/decade, 90% significant) and 2.3 ± 1.51 mm/decade (14.4%/decade, 95% significant). These trends are larger than those calculated previously in Switzerland by Morland et al. (2009) and Hocke et al. (2011). Morland et al. (2009)'s nighttime trends were all positive, but no larger than 0.6 mm/decade ranging between 0.3 and 0.9 mm/decade with uncertainties between 0.3 and 0.6 mm/decade depending on the instrument. Morland et al. (2009)'s midnight radiosonde trend using measurements from the surface to 200 hPa was 0.87 ± 0.46 mm/decade and was statistically significant at 90%. Hocke et al. (2011) found no trends in water vapour using microwave radiometer PWV measurements from 2004–2009. Nyeki et al. (2019) measured temperature and PWV trends from four locations in Switzerland, including the Payerne and Jungfraujoch research stations, with GPS measurements from 1996–2015. They calculated statistically significant (90%) PWV trends at Payerne of 1.03 mm/decade during cloud-free conditions using Sen's slope method. Interestingly, the all-sky (cloudy and clear) condition trend produced a smaller trend value of 0.80 mm/decade (90% significance). We also saw a similar drop in trend magnitude when including all-sky conditions in the radiosonde PWV trend with a change in trend value from 1.34 ± 1.51 mm/decade to 1.08 ± 1.05 mm/decade. Bernet et al. (2019) has published an update to the Nyeki et al. (2019) GNSS PWV trends and the Morland et al. (2009) TROWARA PWV trends using measurements from 1995–2018. They used Payerne GNSS measurements to calculate a PWV trend of 1.14 ± 0.93 mm/decade or 7.3 ± 5.3 %/decade which is statistically significant at 95%. Their GNSS trend used all-weather and both night and day measurements. Their PWV trend calculation is consistent with the RALMO and radiosonde trends at the 1-sigma level.

We have discussed the clear sky bias in our lidar PWV trends and found no statistical difference between the trends using all-weather measurements and our clear-sky only PWV trends. The nighttime bias in the RALMO dataset must also be discussed in comparison with the other literature and our radiosonde trends. We have calculated a statistically significant daytime-only

PWV trend using the operational radiosondes with a magnitude of 1.19 ± 0.99 mm/decade. We also calcualted a mixed PWV trend using both daytime and nighttime measurements which was also 1.18 ± 0.97 mm/decade. The nighttime only radiosonde trend is 0.1 mm/decade smaller than the daytime only trend and the RALMO trend is 0.14 mm/decade larger than the daytime trend. The GNSS trend calculated by Bernet et al. (2019) uses both daytime and nighttime measurements and agrees with our radiosonde trend which uses both day and night measurements. Bernet et al. (2019)'s trend is also not statistically significant from the RALMO PWV trend at the 1-sigma level. Based on these results we do not detect a bias using only nighttime measurements. While there may be a difference in the daytime and nighttime trends it would be difficult to detect with such a short time series. Hocke et al. (2017) calculated an average PWV diurnal cycle amplitude of 2% for Switzerland. Therefore, it would probably take a very long time series with very low scatter in the residuals to significantly detect a difference between day and night trends. Unlike our analysis which only included nighttime measurements, Nyeki et al. (2019) used both daytime and nighttime measurements for the water vapour and temperature trends. Despite this difference, our PWV trends agree within 1σ. Therefore, while there certainly is a natural selection bias due to only using nighttime measurements in our study, the magnitude of the nighttime bias is not currently detectable.

The PWV trends we have calculated over Payerne likely differ from earlier studies due to the fact that the temperature trend has changed between the respective time periods, and the fact that we are measuring during a relatively short time period.The PWV trends over the last 20 years have been increasing. One reason for the increase in the trends could be caused by a corresponding increase in the surface temperature trend over the same time period. We calculated a surface temperature trend we measured with the Payerne daily radiosondes measurements from 2009 - 2019 of 1.38 ± 1.41 °C per decade. The MeteoSwiss 2017 Klimareport (MS2017, Bundesamt für Meteorologie und Klimatologie (2017)) reported an average surface temperature trend over Switzerland of 0.34 °C/decade from 1961 - 2017, which was strongly significant. However, a steeper increase in temperature can clearly be seen in Figure 4.1 of MS2017 for the period of 2009 - 2017, which would correspond with our larger trend value. Additionally, the MS2017 trend is for all of Switzerland, not just for Payerne. Nyeki et al. (2019) reports a surface temperature trend of 0.79°C/decade (90% significant) during cloud-free conditions and a trend of 0.59°C/decade (90% significant) during all-sky conditions at Payerne using measurements from 1996–2015. The last 4 years have included 3 of the warmest years on record for Europe, which may have contributed to the increase in our temperature trends with respect to Nyeki et al.'s (Copernicus Climate Change Service, 2019).

Brocard et al. (2013a) conducted a detailed study of height-resolved temperature trends and found several break points in Payerne temperature trends. While the temperature in the troposphere increased by 0.64 degrees from 1980 to 2000, the temperature trend decreased to almost zero from 2000-2010. Therefore, measurements which include the time period from 2000–2010 will likely have smaller trend values. The temperature trend appears to have now increased again after 2010, but it is not possible with our measurements to determine the most recent break point. Despite the difference in our temperature trends from the previous literature, when the PWV trends are measured in terms of the temperature trend they agree with the dependence of saturation vapour pressure with temperature that water vapour should increase by roughly 7% for every degree. Therefore, the current PWV trends measured by RALMO and the operational radiosondes appear reasonable.

There are very few studies of height-resolved water vapour trends over the last few decades. Most tropospheric water vapour studies focus on surface or total column changes, with a couple including trends at 850 hPa (H2018, Serreze et al. (2012)). While the majority of water vapour resides in the bottom 3 km, it is still of high importance to characterize and understand water vapour's behavior throughout the troposphere. Height-resolved water vapour trends would benefit climate and forecast models. Additionally, Raghuraman et al. (2019) suggests that different height regions in the troposphere contribute unequally to the greenhouse effect. Their study showed that the middle tropospheric water vapour contributed the most to the water vapour component in the greenhouse effect. Lidars are well-suited to height-resolved trend measurements and the authors hope that this is the first of many similar studies which will appear over the next few years. We can compare our surface specific humidity trends to the surface water vapour trends measured in Nyeki et al. (2019). They measured statistically significant surface specific humidity trends in Payerne, during cloud-free conditions, of 0.18 and 0.23 g/kg (depending on the method). We measured surface specific humidity trends of 0.74 ± 0.46 g/kg. Both studies detected trends which agree with the assumption of conservation of relative humidity with Nyeki et al. (2019) detecting a change of 6%/°C and RALMO with 8.8 ± 10.2%/°C. The difference between the two surface trends is likely due to the difference in temperature trend and the time period being used; however, when converting the trends to % per °C the uncertainties in our surface trend encompass the Nyeki surface trend value and the differences are not significant. We can also use Nyeki et al. (2019)'s Jungfraujoch trends to compare with our 700 and 600 hPa specific humidity and temperature trends. The Jungfraujoch Observatory is at 3,580 m or about 650 hPa. The Nyeki Jungfraujoch cloud-free specific humidity trend is 0.19 g/kg per decade while RALMO's trends at 700 and 600 hPa are 0.35 ± 0.36 and $0.22 \pm .24$ g/kg per decade. The trends are the same within their respective $1\sigma_a$ uncertainties. Interestingly, neither our trends nor Nyeki's Jungfraujoch trends conform to the hypothesis that relative humidity is conserved. Both humidity trends with respect to temperature are at or above 14% per degree, however, they are not statistically significant above 95% in those units.. The relative humidity may be conserved over the troposphere, or at the surface, but not in individual layers. Raghuraman et al. (2019) recently tested the assumption of the conservation of relative humidity in a changing climate and found that in small geographical regions relative humidity may not be conserved over time. However, globally, relative humidity is generally conserved. Height-resolved relative humidity trends would help to determine where relative humidity is conserved in the atmosphere.

The percent variability of water vapour through the troposphere is greatest in the free troposphere. The large variability limits the ability to calculate trends in the free troposphere, as is evident by the lack of 95% significance in the trends calculated above the surface. As Whiteman et al. (2011a) and Weatherhead et al. (1998) suggested, the variability and therefore the trend residuals drive the uncertainty in the trend and ultimately determine how long is needed to calculate the trend. The noise in the residuals of the height-resolved trends varies between 11% and 24%. Given the noise at the individual levels in our height-resolved trends, detecting a statistically significant trend at 95% of 10% per decade could take between 12–24 years of measurements (Equation 2, Whiteman et al. (2011a)). Longer years are needed at the levels with the most noise, namely between 600–400 hPa. That is not to say that one should wait that long to try and calculate a trend. The equation in Whiteman et al. (2011a) and Weatherhead et al. (1998) to estimate the number of years to calculate a trend assumes a 90% probability of a statistically significant detection (Tiao et al., 1990; Weatherhead et al., 1998). Therefore it is an upper limit on the number of

years needed. It may be possible to calculate a statistically significant trend sooner, as we have done here. In fact, that equation predicts 13 years are needed for RALMO to calculate a trend at the surface at the current trend magnitude, and yet we still found a statistically significant trend. Another point worth considering is that the amount of noise may partially be determined by the vertical resolution chosen for this study. We chose to use higher resolutions, but reducing the vertical resolution of the measurements could possibly reduce the noise in the residuals thus increasing trend significance levels. We leave this experiment to future studies.

7 Summary and Conclusions

We used the calibrations done by Hicks-Jalali et al. (2019) (Appendix A) in conjunction with the updated water vapour OEM retrievals from Sica and Haefele (2016) to reprocess 11 and a half years of water vapour measurements from RALMO (January 2009 - August 2019). We calculated a monthly tropospheric climatology from the reprocessed measurements which shows that we have more water vapour in the summer than winter, as expected with the larger summer temperatures. The climatology also reaches lower pressures during the summer when more water vapour is present. The statistical uncertainty at the lowest pressures is 14% of the water vapour content. The total uncertainty at those pressures is on average 20% when considering the systematic uncertainties. Uncertainties at lower altitudes are dominated by the uncertainty in the calibration constant and are a constant 5% until the statistical uncertainty is larger (roughly above 350 hPa).

We also calculated the geophysical variability of water vapour in the troposphere. The geophysical variability is a selective representation since the measurement intervals in this study were restricted to relatively clear nighttime conditions with most low-level clouds and optically thick mid-level clouds removed from the reprocessing. Therefore, the variability that is presented is representative of the nighttime structure during clear to semi-clear conditions. The same variability study was done using both daytime and nighttime operational radiosondes and no restrictions were placed on the presence of clouds or precipitation. The radiosonde and the lidar variability studies agree in the boundary layer and both datasets show that the nighttime boundary layer does not exhibit a large percent variability. The magnitude of the variability in the radiosonde data set is larger than in the lidar study in the free troposphere, but they see similar behavior in the variability. This bias is likely due to the fact that the water vapour variability measured by the radiosonde can only be estimated since their measurement uncertainties are not reported. We are likely underestimating the uncertainties in the radiosonde which would create the difference in the variabilities measured by the two instruments. In general we see higher variability in the free troposphere than in the boundary layer. Additionally, the lidar and radiosondes measure an increased variability in water vapour during the winter months.

Most previous tropospheric water vapour trend studies focus on precipitable water vapour. Therefore, we calculated PWV trends using the bootstrap method for both the radiosonde and RALMO measurements. The radiosonde trends were calculated using nighttime and daytime measurements. The nighttime PWV trends calculated by the lidar and the radiosondes were 1.32 ± 1.51 mm/decade (8.9% per decade) and 2.31 ± 1.53 mm/decade (15.5% per decade), respectively when using the same dates. When the radiosonde trend includes all available data, including those in which clouds or precipitation are present, the PWV trend decreases to 1.08 ± 1.05 mm/decade or 6.36% per decade. The RALMO PWV trend is statistically significant

27

above 90%. Both radiosonde nighttime PWV trends are significant to 95%. We also compared the nighttime only trends to daytime-only PWV trends using radiosonde measurements. We measured a daytime PWV trend of 1.18 ± 0.97 mm/decade or 7.20%/decade which was statistically significant at 95%. When combining daytime and nighttime radiosonde measurements we detected a similar trend of 1.19 ± 0.98 mm/decade or 7.20%/decade. Due to the large uncertainties in the trends we cannot

5   detect a bias between the daytime and nighttime trends at this time. These large changes in water vapour can be explained by correspondingly large temperature changes. We calculated a surface temperature trend using the measurements from the radiosonde and found an increase of $1.38°C$ per decade from January 2009 - August 2019. The PWV trends are then consistent with the fact that a change in 1 degree Celsius will result in a roughly 7% change in water vapour at atmospheric temperatures of around 300 K, assuming relative humidity is conserved Held and Soden (2000).

10   Lastly, we investigated the specific humidity trends at 11 pressure layers in the troposphere, each 20 hPa in thickness. No statistically significant trends were found at 900 and 500 hPa, or above 350 hPa. However, all other layers exhibited positive trends with statistical significance at or above 80%. The surface specific humidity trend followed the expected Clausius Clapeyron relationship and had a trend of 8.8% change in specific humidity per degree Celsius; however, when converting to % per degree the trend is no longer significant. All other trends were larger than 7% per °C, with some as large as 17% per °C (column 5 of

15   Table 3).

We have shown that Raman water vapour lidars are useful for making detailed tropospheric water vapour measurements and can be used for long-term analysis. The water vapour concentration over Payerne is changing at twice the rate compared to trend measurements from 10 years ago. The change in precipitable water vapour is consistent with a temperature gradient of $1.38°C$ per decade, assuming that relative humidity is conserved. These large changes in temperature and water vapour could

20   have concerning impacts on the climate in Payerne over the coming decades. According to Held and Soden (2006), as water vapour increases in the atmosphere the circulation of water vapour increases and wet regions become even wetter. The increase in saturation vapour pressure with the increasing temperatures could lead to stronger precipitation events (Sherwood et al., 2010). For the Swiss plateau, MeteoSwiss detects a positive, though non-significant, trend of 1.4% per decade in precipitation for the period 1961–2017, and the IPCC reports that, in general, precipitation amounts are increasing in the latitude band from

25   30°N - 60°N (Bundesamt für Meteorologie und Klimatologie, 2017; Hartmann et al., 2013).

We show for the first time, height-resolved water vapour trends in the troposphere. These trends suggest that relative humidity may be conserved at the surface, but not necessarily aloft. More measurements from Payerne as well as other lidar sites of height-resolved water vapour trends are needed in order to investigate our results. As of this writing, radiosondes are not yet well enough characterized to calculate height-resolved trends with high confidence (Elliott and Gaffen, 1991; McCarthy et al.,

30   2009; Miloshevich et al., 2009), while satellite water vapour measurements typically have low vertical resolution and high uncertainty in the troposphere. The higher variability of water vapour in the free troposphere compared to the surface and boundary layer means that longer timelines and/or measurements with low uncertainty are required to establish statistically significant trends at the 95% level.

Appendix A: The RALMO Calibration Time Series

Hicks-Jalali et al. (2019) discussed the trajectory calibration technique which we used to calibrate RALMO from 2011 to 2016 using GRUAN-certified sondes. This method is valuable because it allows for the drift of the sonde away from the launch site. The method does not require the use of GRUAN sondes specifically, but would work with any sonde that reports wind speed and direction, or latitude and longitude coordinates. However, the use of GRUAN-certified radiosondes is preferable because it allows the researcher to calculate uncertainty budgets for every calibration. We found that the average total uncertainty in our calibration was around 5% for both the trajectory method and the traditional techniques and the uncertainty did not increase over the time period of the calibration study. The majority of the uncertainty (4%) is due to the uncertainty of the radiosonde measurement. While the uncertainty of the calibration constant did not increase over the course of the study, the value of the calibration constant did increase by at least 30% over the course of the five years. It is thought that this change in the calibration constant is due to a differential aging in the nitrogen and water vapour photomultipliers (Simeonov et al., 2014).

It has been well discussed in the lidar and trends communities that abrupt changes in calibration or any discontinuities in time series should be avoided (Whiteman et al., 2011b; Weatherhead et al., 1998). Therefore, the 25 calibration nights from the Hicks-Jalali et al. (2019) study would not be enough to use for a trend analysis, as they are too sparsely distributed and do not encompass the entire time series. They also do not fully characterize the evolution of the calibration factor over the course of RALMO's lifetime. For this reason, internal calibration methods are extremely useful as they can provide a continuous calibration function (Leblanc and McDermid, 2008; Venable et al., 2011). Leblanc and McDermid (2008) found that a hybrid method of combining an internal lamp calibration and an external calibration with a radiosonde decreased their calibration uncertainty to less than 2%.

In 2014, a UV lamp was installed at RALMO to internally calibrate the lidar following the procedure outlined in Simeonov et al. (2014), however it lacked sufficient stability and was removed from the system. An alternative method called the "Background Calibration Method" (BCM) was implemented instead. The BCM uses solar background to compute the relative calibration coefficient for RALMO (Voirin, 2017). The solar background calibration method was first introduced in Sherlock et al. (1999). We will summarize and then apply the technique to the RALMO water vapour time series used in this paper.

The calibration coefficient is the constant which converts the relative lidar profile into physical units, such as mixing ratios (g/kg). When measuring with a constant light source, such as a lamp, the ratio of the signals between the nitrogen and water vapor channels is a measurement of the ratio of each channel's efficiency. Measuring the ratio over time creates the relative calibration time series $r_{solar}(t)$:

$$r_{solar}(t) = \frac{N_{H,const}(t)}{N_{N,const}(t)}, \tag{A1}$$

where $N_{X,const}(t)$ is the water vapour or nitrogen signal measured with a constant light source (either a lamp or solar background). This function can then be normalized by an external calibration measurement taken at a time t_0 or by normalizing using an entire time series of measurements, such that the new calibration time series becomes:

$$C(t) = C_{ext}(t_0)r_{solar}(t). \tag{A2}$$
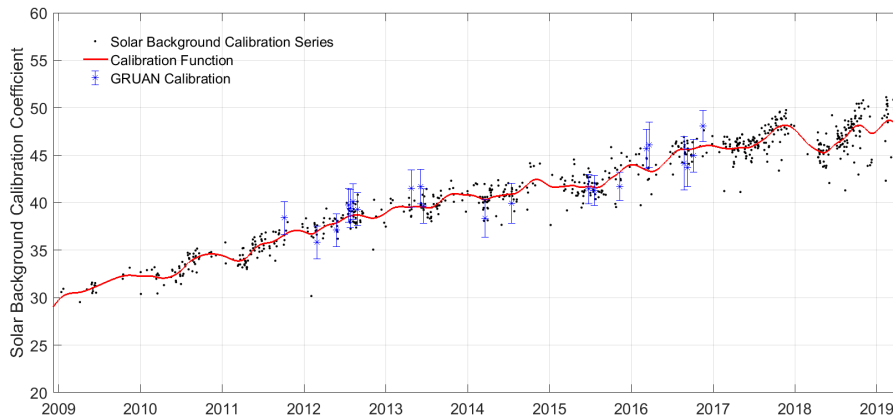
Figure A1. The final normalized calibration time series. Black points are the solar calibration values from January 2009 to August 2019. Blue points are the trajectory calibration values with their uncertainties. The red line is the smoothing spline fit to the solar calibration points.

The external calibration, C_{ext}), at a time t_0 can be chosen at any point in the time series, or be an average of several points. The solar background signal for both the nitrogen and water vapour channels can be calculated by taking the average value of the background signal above 50 km. At these altitudes in a raw 1-minute profile, the signal is completely due to background solar radiation and not photons emitted by the laser. Therefore, during the daytime, all photons should be solar photons or tube background. One must be careful to consider both the diurnal and seasonal solar cycles when using this technique, therefore, it is important to use the same solar zenith angle each day at the highest possible angle. we only use the solar background corresponding to The highest solar zenith angle on the winter solstice (20°) is 20°. Therefore, the calibration is conducted each cloudless morning when the sun is at a solar zenith angle of 20°. The ratio of the nitrogen solar background and the water vapour solar background were then used to calculate the r_{solar} function. We do not correct the solar background time series for differential transmission between the nitrogen and water vapour channels. A sensitivity study using over three years of RALMO data was used to assess the value of $\Delta\alpha$, where $\Delta\alpha = (1 - \tau_{N_2})/(1 - \tau_{H_2O})$. The study found that the overall impact of the differential transmission correction amounts to a maximum of 5%, with an average of 2%. These findings are in agreement with what was found in the study by Whiteman and colleagues (1992). However, the GRUAN radiosonde calibrations were corrected for differential aerosol transmission (Hicks-Jalali et al. 2018), and the relative calibration time series is scaled by the GRUAN radiosonde external calibration. Scaling the calibration time series with the GRUAN radiosonde calibration creates a continuous calibration time series from 2008 until the end of 2018 (Fig. A1). We used the calibration time series from Hicks-Jalali et al. (2019) to scale the relative solar background time series to create a continuous calibration time series from 2008 until the end of 2018 (Fig. A1).

The black points in Fig. A1 are the solar background calibration time series. The background values clearly exhibit a linear component which would indicate that the nitrogen and water vapour photomultipliers are differentially aging (assuming that the aging is linear). However, there is also a more complicated behavior embedded in the time series, particularly towards

the end of 2017 and 2018 which would make it inappropriate to simply fit a linear function to the time series. Therefore, we chose to fit a smoothing spline to the time series, which is represented by the red line in Fig. A1. The blue points show the calibration values from Hicks-Jalali et al. (2019) with their respective uncertainties. The abrupt dip in calibration values after 2018 may be due to the installation of a new laser at that time, but the source has not yet been determined. The uncertainty

5   in the calibration function was estimated using the standard deviation of the de-trended background time series. The percent uncertainty is the standard deviation divided by the mean calibration value. It agrees with the average uncertainty of 5% from the GRUAN calibrations in Hicks-Jalali et al. (2019). While this method did not decrease the uncertainty of the calibration, it allowed the use of a continuous and consistent calibration constant for the entire water vapour time series.

References

Argall, P. S. and Sica, R. J.: A comparison of Rayleigh and sodium lidar temperature climatologies, Annales Geophysicae, 25, 27–35, https://doi.org/10.5194/angeo-25-27-2007, 2007.

Bernet, L., Von Clarmann, T., Godin-Beekmann, S., Ancellet, G., Barras, E. M., Stübi, R., Steinbrecht, W., Kämpfer, N., and Hocke, K.: Ground-based ozone profiles over central Europe: Incorporating anomalous observations into the analysis of stratospheric ozone trends, Atmospheric Chemistry and Physics, 19, 4289–4309, https://doi.org/10.5194/acp-19-4289-2019, 2019.

Brocard, E., Jeannet, P., Begert, M., Levrat, G., Philipona, R., Romanens, G., and Scherrer, S. C.: Upper air temperature trends above Switzerland 1959-2011, Journal of Geophysical Research Atmospheres, 118, 4303–4317, https://doi.org/10.1002/jgrd.50438, 2013a.

Brocard, E., Philipona, R., Haefele, A., Romanens, G., Mueller, A., Ruffieux, D., Simeonov, V., and Calpini, B.: Raman Lidar for Meteorological Observations , RALMO-Part 2 : Validation of water vapor measurements, Atmospheric Measurement Techniques, 6, 1347–1358, https://doi.org/10.5194/amt-6-1347-2013, 2013b.

Bundesamt für Meteorologie und Klimatologie: Klimareport 2017 MeteoSchweiz, Tech. rep., MeteoSchweiz, https://doi.org/https://doi.org/10.18, www.bundespublikationen.admin.ch, 2017.

Copernicus Climate Change Service: European State of the Climate 2018, Tech. rep., ECMWF, https://climate.copernicus.eu/sites/default/files/2019-04/Brochure{_}Final{_}Interactive{_}1.pdf, 2019.

Dai, A., Trenberth, K. E., and Karl, T. R.: Effects of clouds, soil moisture, precipitation, and water vapor on diurnal temperature range, Journal of Climate, 12, 2451–2473, https://doi.org/10.1175/1520-0442(1999)012<2451:eocsmp>2.0.co;2, 1999.

Dessler, A. E., Schoeberl, M. R., Wang, T., Davis, S. M., and Rosenlof, K. H.: Stratospheric water vapor feedback, Proceedings of the National Academy of Sciences, 110, 18 087–18 091, https://doi.org/10.1073/pnas.1310344110, http://www.pnas.org/cgi/doi/10.1073/pnas.1310344110, 2013.

Dinoev, T., Simeonov, V., Arshinov, Y., Bobrovnikov, S., Ristori, P., Calpini, B., Parlange, M., and Van Den Bergh, H.: Raman Lidar for Meteorological Observations, RALMO-Part 1 : Instrument description, Atmospheric Measurement Techniques, 6, 1329–1346, https://doi.org/10.5194/amt-6-1329-2013, 2013.

Dirksen, R. J., Sommer, M., Immler, F. J., Hurst, D. F., Kivi, R., and Vömel, H.: Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde, Atmospheric Measurement Techniques, 7, 4463–4490, 2014.

Elliott, W. P. and Gaffen, D. J.: On the Utility of Radiosonde Humidity Archives for Climate Studies, Bulletin of the American Meteorological Society, 72, 1507–1520, https://doi.org/10.1175/1520-0477(1991)072<1507:otuorh>2.0.co;2, 1991.

Farhani, G., Sica, R. J., Godin-Beekmann, S., and Haefele, A.: Optimal estimation method retrievals of stratospheric ozone profiles from a DIAL, Atmospheric Measurement Techniques, 12, 2097–2111, https://doi.org/10.5194/amt-12-2097-2019, 2019.

Gardiner, T., Forbes, A., De Mazière, M., Vigouroux, C., Mahieu, E., Demoulin, P., Velazco, V., Notholt, J., Blumenstock, T., Hase, F., Kramer, I., Sussmann, R., Stremme, W., Mellqvist, J., Strandberg, A., Ellingsen, K., and Gauss, M.: Trend analysis of greenhouse gases over Europe measured by a network of ground-based remote FTIR instruments, Atmospheric Chemistry and Physics, 8, 6719–6727, https://doi.org/10.5194/acp-8-6719-2008, 2008.

Goldsmith, J. E. M., Bisson, S. E., Ferrare, R. A., Evans, K. D., Whiteman, D. N., and Melfi, S. H.: Raman Lidar Profiling of Atmospheric Water Vapor: Simultaneous Measurements with Two Collocated Systems, Bulletin of the American Meteorological Society, 75, 975–982, https://doi.org/10.1175/1520-0477(1994)075<0975:RLPOAW>2.0.CO;2, 1994.

Hadad, D., Baray, J. L., Montoux, N., Van Baelen, J. V., Fréville, P., Pichon, J. M., Bosser, P., Ramonet, M., Kwok, C. Y., Bègue, N., and Duflot, V.: Surface and tropospheric water vapor variability and decadal trends at two supersites of CO-PDD (Cézeaux and Puy de Dôme) in central France, Atmosphere, 9, 1–22, https://doi.org/10.3390/atmos9080302, 2018.

Hartmann, D. L., Klein Tank, A. M. G., Rusticucci, M., Alexander, L., Bronnimann, S., Charabi, Y. A.-R., Dentener, F. J., Dlugokencky, E. J., Easterling, D. R., Kaplan, A., Soden, B. J., Thorne, P. W., Wild, M., and Zhai, P.: IPCC Climate Change 2013: The Physical Science Basis. Chapter 2: Observations: Atmosphere and Surface, Tech. rep., IPCC, Cambring, UK and New York, NY, USA, https://doi.org/10.1017/CBO9781107415324.008, 2013.

Hegglin, M. I., Tegtmeier, S., Anderson, J., Froidevaux, L., Fuller, R., Funke, B., Jones, A., Lingenfelser, G., Lumpe, J., Pendlebury, D., Remsberg, E., Rozanov, A., Toohey, M., Urban, J., Von Clarmann, T., Walker, K. A., Wang, R., and Weigel, K.: SPARC Data Initiative: Comparison of water vapor climatologies from international satellite limb sounders, Journal of Geophysical Research Atmospheres, 118, 11 824–11 846, https://doi.org/10.1002/jgrd.50752, 2013.

Hegglin, M. I., Plummer, D. A., Shepherd, T. G., Scinocca, J. F., Anderson, J., Froidevaux, L., Funke, B., Hurst, D., Rozanov, A., Urban, J., Von Clarmann, T., Walker, K. A., Wang, H. J., Tegtmeier, S., and Weigel, K.: Vertical structure of stratospheric water vapour trends derived from merged satellite data, Nature Geoscience, 7, 768–776, https://doi.org/10.1038/NGEO2236, 2014.

Held, I. M. and Soden, B. J.: Water Vapor Feedback and Global Warming, Annual Reviews of Energy and Environment, 25, 441–75, 2000.

Held, I. M. and Soden, B. J.: Robust Responses of the Hydrological Cycle to Global Warming, Journal of Climate, 19, 5686–5699, https://doi.org/10.5220/0001648603850388, 2006.

Hense, A., Krahe, P., and Flohn, H.: Recent fluctuations of tropospheric temperature and water vapour content in the tropics, Meteorology and Atmospheric Physics, 38, 215–227, https://doi.org/10.1007/BF01054574, 1988.

Hicks-Jalali, S.: A Tropospheric Water Vapour Climatology and Trends Derived from Vibrational Raman Lidar Measurements over Switzerland, Ph.D. thesis, University of Western Ontario, https://ir.lib.uwo.ca/etd/6375, 2019.

Hicks-Jalali, S., Sica, R. J., Haefele, A., and Martucci, G.: Calibration of a Water Vapour Raman Lidar using GRUAN-certified Radiosondes and a New Trajectory Method, Atmospheric Measurement Techniques Discussions, pp. 3699–3716, https://doi.org/https://doi.org/10.5194/amt-12-3699-2019, 2019.

Hocke, K., Kämpfer, N., Gerber, C., and Mätzler, C.: A complete long-term series of integrated water vapour from ground-based microwave radiometers, International Journal of Remote Sensing, 32, 751–765, https://doi.org/10.1080/01431161.2010.517792, 2011.

Hocke, K., Navas-Guzmán, F., Moreira, L., Bernet, L., and Mätzler, C.: Diurnal Cycle in Atmospheric Water over Switzerland, Remote Sensing, 9, 909, https://doi.org/10.3390/rs9090909, http://www.mdpi.com/2072-4292/9/9/909, 2017.

Holton, J. R.: An introduction to dynamic meteorology: Fourth edition, Elsevier Inc., Burlington, MA, fourth edn., https://doi.org/10.1016/C2009-0-63394-8, 2004.

Immler, F. J., Dykema, J., Gardiner, T., Whiteman, D. N., Thorne, P. W., and Vömel, H.: Reference quality upper-air measurements: Guidance for developing GRUAN data products, Atmospheric Measurement Techniques, 3, 1217–1231, https://doi.org/10.5194/amt-3-1217-2010, 2010.

Jalali, A., Sica, R. J., and Haefele, A.: Improvements to a long-term Rayleigh-scatter lidar temperature climatology by using an optimal estimation method, Atmospheric Measurement Techniques, 11, 6043 – 6058, https://doi.org/10.5194/amt-11-6043-2018, 2018.

Jensen, M. P., Holdridge, D. J., Survo, P., Lehtinen, R., Baxter, S., Toto, T., and Johnson, K. L.: Comparison of Vaisala radiosondes RS41 and RS92 at the ARM Southern Great Plains site, Atmospheric Measurement Techniques, 9, 3115–3129, https://doi.org/10.5194/amt-9-3115-2016, 2016.

Jin, S., Park, J., Cho, J. H., and Park, P. H.: Seasonal variability of GPS-derived zenith tropospheric delay (1994-2006) and climate implications, Journal of Geophysical Research Atmospheres, 112, 1–11, https://doi.org/10.1029/2006JD007772, 2007.

Khosrawi, F., Lossow, S., Stiller, G. P., Rosenlof, K. H., Urban, J., Burrows, J. P., Damadeo, R. P., Eriksson, P., García-Comas, M., Gille, J. C., Kasai, Y., Kiefer, M., Nedoluha, G. E., Noël, S., Raspollini, P., Read, W. G., Rozanov, A., Sioris, C. E., Walker, K. A., and Weigel, K.: The SPARC water vapour assessment II: Comparison of stratospheric and lower mesospheric water vapour time series observed from satellites, Atmospheric Measurement Techniques, 11, 4435–4463, https://doi.org/10.5194/amt-11-4435-2018, 2018.

Kiehl, J. T. and Trenberth, K. E.: Earth's Annual Global Mean Energy Budget, Bulletin of the American Meteorological Society, 78, 197–208, 1997.

Leblanc, T. and McDermid, I. S.: Accuracy of Raman lidar water vapor calibration and its applicability to long-term measurements, Journal of Applied Optics, 47, 5592–5603, https://doi.org/https://doi.org/10.1364/AO.47.005592, 2008.

Leblanc, T., Sica, R. J., Van Gijsel, J. A., Godin-Beekmann, S., Haefele, A., Payen, G., and Liberti, G.: Proposed standardized definitions for vertical resolution and uncertainty in the NDACC lidar ozone and temperature algorithms-Part 3: Temperature Uncertainty Budget, Atmospheric Measurement Techniques, 9, 4051–4078, https://doi.org/10.5194/amt-9-4051-2016, 2016.

Mahagammulla Gamage, S., Sica, R. J., Martucci, G., and Haefele, A.: Retrieval of temperature from a multiple channel pure rotational Raman backscatter lidar using an optimal estimation method, Atmospheric Measurement Techniques, 12, 5801–5816, https://doi.org/10.5194/amt-12-5801-2019, 2019.

Mahagamulla Gamage, S., Sica, R. J., Haefele, A., and Martucci, G.: A 1D Var assimilation for relative humidity using Raman lidar1measurements and the ERA5 dataset, Journal of Atmospheric and Oceanic Technology, Under Review, 2020.

Martin, L., Mätzler, C., Hewison, T. J., and Ruffieux, D.: Intercomparison of integrated water vapour measurements, Meteorologische Zeitschrift, 15, 57–64, https://doi.org/10.1127/0941-2948/2006/0098, 2006.

McCarthy, M. P., Thorne, P. W., and Titchner, H. A.: An analysis of tropospheric humidity trends from radiosondes, Journal of Climate, 22, 5820–5838, https://doi.org/10.1175/2009JCLI2879.1, 2009.

Melfi, S. H.: Remote Measurements of the Atmosphere Using Raman Scattering, Journal of Applied Optics, 11, 1605–1610, 1972.

Meteolabor: MeteoLabor SRS-C34 Digital Radiosonde, Tech. rep., MeteoLabor, meteolabor.ch, 2010.

Miloshevich, L. M., Vömel, H., Whiteman, D. N., and Leblanc, T.: Accuracy assessment and correction of Vaisala RS92 radiosonde water vapor measurements, Journal of Geophysical Research, 114, 1–23, https://doi.org/10.1029/2008JD011565, 2009.

Morland, J., Coen, C. M., Hocke, K., Jeannet, P., and Mätzler, C.: Tropospheric water vapour above Switzerland over the last 12 years, Atmospheric Chemistry and Physics, 9, 5975–5988, https://doi.org/10.5194/acp-9-5975-2009, 2009.

Nilsson, T. and Elgered, G.: Long-term trends in the atmospheric water vapor content estimated from ground-based GPS data, Journal of Geophysical Research Atmospheres, 113, https://doi.org/10.1029/2008JD010110, 2008.

Ning, T. and Elgered, G.: Trends in the atmospheric water vapour estimated from GPS data for different elevation cutoff angles, Atmospheric Measurement Techniques Discussions, pp. 1–25, https://doi.org/10.5194/amt-2018-279, 2018.

Nyeki, S., Wacker, S., Aebi, C., Gröbner, J., Martucci, G., and Vuilleumier, L.: Trends in surface radiation and cloud radiative effect at four Swiss sites for the 1996-2015 period, Atmospheric Chemistry and Physics Discussions, pp. 1–22, https://doi.org/10.5194/acp-2018-1096, 2019.

Raghuraman, S. P., Paynter, D., and Ramaswamy, V.: Quantifying the Drivers of the Clear Sky Greenhouse Effect , 2000-2016, JGR: Atmospheres, https://doi.org/10.1029/2019JD031017, 2019.

Reichardt, J., Wandinger, U., Klein, V., Mattis, I., Hilber, B., and Begbie, R.: RAMSES: German meteorological service autonomous Raman Iidar for water vapor, temperature, aerosol, and cloud measurements, Applied Optics, 51, 8111–8131, https://doi.org/10.1364/AO.51.008111, 2012.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding, World Scientific, 2000.

5   Ross, R. J. and Elliot, W. P.: Radiosonde-based Northern Hemisphere tropospheric water vapor trends, Journal of Climate, 14, 1602–1612, https://doi.org/10.1175/1520-0442(2001)014<1602:RBNHTW>2.0.CO;2, 2001.

Ross, R. J. and Elliott, W. P.: Tropospheric WAter Vapor Climatology and Trends over North America: 1973-93, Journal of Climate, 9, 3561 – 3574, 1996.

Serreze, M. C., Barrett, A. P., and Stroeve, J.: Recent changes in tropospheric water vapor over the Arctic as assessed from radiosondes and

10   atmospheric reanalyses, Journal of Geophysical Research Atmospheres, 117, 1–21, https://doi.org/10.1029/2011JD017421, 2012.

Sherlock, V., Hauchecorne, A., and Lenoble, J.: Methodology for the independent calibration of Raman backscatter water-vapor lidar systems., Applied optics, 38, 5816–5837, https://doi.org/10.1364/AO.38.005816, 1999.

Sherwood, S. C., Roca, R., Weckwerth, T. M., Andronova, N. G., Change, C., Wales, S., and Arbor, A.: Tropospheric Water Vapor, Convection and Climate, Reviews of Geophysics, pp. 1–30, 2010.

15   Sica, R. J. and Haefele, A.: Retrieval of temperature from a multiple-channel Rayleigh-scatter lidar using an optimal estimation method, Applied Optics, 54, 1872–1889, https://doi.org/10.1364/AO.54.001872, 2015.

Sica, R. J. and Haefele, A.: Retrieval of water vapor mixing ratio from a multiple channel Raman-scatter lidar using an optimal estimation method, Journal of Applied Optics, 55, 763–777, https://doi.org/https://doi.org/10.1364/AO.55.000763, 2016.

Simeonov, V., Fastig, S., Haefele, A., and Calpini, B.: Instrumental correction of the uneven PMT aging effect on the calibration constant of

20   a water vapor Raman lidar, Proc. of SPIE, 9246, 1–9, https://doi.org/10.1117/12.2066802, 2014.

Tiao, G., Reinsel, G. C., Xu, D., Pedrick, J., Zhu, X., Miller, A., DeLuisi, J., Mateer, C., and Wuebbles, D. J.: Effects of Autocorrelation and Temporal Sampling Schemes on Estimates of Trend and Spatial Correlation, Journal of Geophysical Research, 95, 507–520, https://doi.org/10.1029/JD095iD12p20507, 1990.

Trenberth, K. E., Fasullo, J., and Smith, L.: Trends and variability in column-integrated atmospheric water vapor, Climate Dynamics, 24,

25   741–758, https://doi.org/10.1007/s00382-005-0017-4, 2005.

Trent, T., Schröder, M., and Remedios, J.: GEWEX Water Vapor Assessment: Validation of AIRS Tropospheric Humidity Profiles With Characterized Radiosonde Soundings, Journal of Geophysical Research: Atmospheres, 124, 886–906, https://doi.org/10.1029/2018JD028930, 2019.

Venable, D. D., Whiteman, D. N., Calhoun, M. N., Dirisu, A. O., Connell, R. M., and Landulfo, E.: Lamp mapping technique for independent

30   determination of the water vapor mixing ratio calibration factor for a Raman lidar system, Journal of Applied Optics, 50, 4622–4632, https://doi.org/https://doi.org/10.1364/AO.50.004622, 2011.

Voirin, J.: Water Vapor Calibration of the Raman LIDAR for Meteorological Observation: Master Internship Report, Tech. rep., MeteoSwiss, 2017.

Wang, J., Dai, A., and Mears, C.: Global water vapor trend from 1988 to 2011 and its diurnal asymmetry based on GPS, radiosonde, and

35   microwave satellite measurements, Journal of Climate, 29, 5205–5222, https://doi.org/10.1175/JCLI-D-15-0485.1, 2016.

Weatherhead, E. C., Reinsel, G. C., Tiao, G. C., Meng, X.-L., Choi, D., Cheang, W.-K., Keller, T., DeLuisi, J., Wuebbles, D. J., Kerr, J. B., Miller, A. J., Oltmans, S. J., and Frederick, J. E.: Factors affecting the detection of trends: Statistical considerations and applications

to environmental data, Journal of Geophysical Research, 103, 17 149, https://doi.org/10.1029/98JD00995, http://doi.wiley.com/10.1029/98JD00995, 1998.

Weckwerth, T. M., Wulfmeyer, V., Wakimoto, R. M., Hardesty, R. M., Wilson, J. W., and Banta, R. M.: NCAR-NOAA Lower-Tropospheric Water Vapor Workshop, Bulletin of the American Meteorological Society, 80, 2339–2357, https://doi.org/10.1175/1520-0477(1999)080<2339:NNLTWV>2.0.CO;2, 1999.

Whiteman, D. N.: Examination of the traditional Raman lidar technique II. Evaluating the ratios for water vapor and aerosols, Applied Optics, 42, 2593, https://doi.org/10.1364/AO.42.002593, 2003.

Whiteman, D. N., Melfi, S. H., and Ferrare, R. A.: Raman lidar system for the measurement of water vapor and aerosols in the Earth's atmosphere., Applied optics, 31, 3068–82, https://doi.org/10.1364/AO.31.003068, 1992.

Whiteman, D. N., Vermeesch, K. C., Oman, L. D., and Weatherhead, E. C.: The relative importance of random error and observation frequency in detecting trends in upper tropospheric water vapor, Journal of Geophysical Research Atmospheres, 116, 1–7, https://doi.org/10.1029/2011JD016610, 2011a.

Whiteman, D. N., Vermeesch, K. C., Oman, L. D., and Weatherhead, E. C.: The relative importance of random error and observation frequency in detecting trends in upper tropospheric water vapor, Journal of Geophysical Research Atmospheres, 116, 1–7, https://doi.org/10.1029/2011JD016610, 2011b.

Whiteman, D. N., Cadirola, M., Venable, D., Calhoun, M., Miloshevich, L., Vermeesch, K., Twigg, L., Dirisu, A., Hurst, D., Hall, E., Jordan, A., and Vömel, H.: Correction technique for Raman water vapor lidar signal-dependent bias and suitability for water vapor trend monitoring in the upper troposphere, Atmospheric Measurement Techniques, 5, 2893–2916, https://doi.org/10.5194/amt-5-2893-2012, 2012.

Wulfmeyer, V., Girolamo, P. D., Schlüssel, P., Baelen, J. V., and Zus, F.: Reviews of Geophysics of water and energy cycles, Reviews of Geophysics, pp. 819–895, https://doi.org/10.1002/2014RG000476.Received, 2015.