Review of Hicks-Jalali et al."A Raman Lidar Tropospheric Water Vapour Climatology and Height-Resolved Trend Analysis over Payerne Switzerland" by David N. Whiteman

The subject paper makes use of a 11.5-year time series of measurements from the RALMO Raman Lidar in Payerne Switzerland to calculate a nighttime water vapor climatology and height-resolved trends. The authors use the interesting new Optimal Estimation Method technique for Raman water vapor lidar retrievals introduced by Sica et al. a few years ago. The work is welcomed as extended data sets from Raman lidar have been under utilized in the past for this kind of analysis and I encourage the authors to continue this valuable work. I have a few major concerns and several minor ones and I look forward to seeing the revised manuscript.

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.
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?
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.
4. The authors refer to Table 1 of Weatherhead et al., 1998 and seem to be referring to the σ_N 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 σ_N that ranged between ~20-80% were found. I would expect values of σ_N much larger than the 6% used here.

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.
2. P 2, line 18. Suggest change to “Due to this huge variation, quantifying water vapor trends requires ...”
3. P2, line 32. AIRS is referred to as a “satellite”. It is an instrument on the Aqua satellite.
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”.
5. P3, line 5. Suggest change to “...trend measurements are ...”
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.
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.
8. P3, line 20. Suggest change to “...found a positive but insignificant trend ...”
9. P3 lines 25-26. The PWV trend values from Cezeaux have already been mentioned above. Delete this redundant information.
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
11. P3 line 34. This study is described here as being 10 years in length but elsewhere 11.5 years is used. Please reconcile.
12. P3. Line 35. This is the place to first note that this study uses only nighttime measurements.
13. P4 line 14. Suggest change to “...laser operating at 30 Hz...”
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.
15. P 5, line 15. unit is used of “counts/bin/m”. I suspect this should be “counts/bin/min” instead. Right?
16. P5, line 21. Suggest change to “...water vapor profiles have ...”
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.
18. P6, line 12. Suggest change to “Measurements from GCOS … highest quality radiosonde data product available”.
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 …”
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?
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.

22. P7, line 26. Suggest change to “This means that the minimum ...”
23. P9, Fig 3. Caption states fractional uncertainties but figure is labelled in %. Please reconcile.
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?
25. P9, line 10. Change “ration” to “ratio”.
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)?
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.
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.
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.
30. P11, line 15. “The see ...” ??
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.
32. P13, line 7. Suggest change to “...in the free troposphere could explain ...” since you are speculating here.
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.
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.
35. P 15, Fig 8. Suggest you add the equations of the linear trend fit to the figure for each of a, b, c.
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.
37. P15, line 13. The RALMO calculated PWV trend is not significantly different from 0 at the 95% confidence level. Right?

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.
39. P19, line 23. Authors give calculated trend values of 1.3 and 2.3. Please add uncertainties to these values.
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.
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.
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.
    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.
    3. Authors state in conclusion at the end of this paragraph: “Therefore, while there certainly is a natural selection bias due to only using 10 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.”
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.
44. P22, line 11. Suggest change to “Interestingly, neither our trends nor … conforms to ...”
45. P23, line 25. Change to “ratios”.
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.
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?