Review comments #1:

Summary:

The paper describes an instrument designed to continuously profile water vapor in the lower atmosphere at short time scales (10s averaging for Nyquist frequency of 0.05Hz). The instrument design and relevant specifications are described. An hour of high resolution data in the PBL and above are shown as examples. The goal is to resolve turbulence and an analysis of turbulence spectra are highlighted. The instrument is compared to radiosondes and state-of-the-art Raman lidar systems. Overall the paper achieves its objectives but there are some concerns and issues that need to be addressed.

Many thanks for this review. The comments helped to significantly improve this manuscript. We carefully revised the manuscript based on the comments by both reviews. When going through the individual points of criticism, we came across a small but relevant error in the previous calculation of water vapor absorption coefficients. Therefore, all plots including DIAL data were recalculated and are now in their updated version. As a result, some of the comments do not apply anymore.
A detailed explanation on the changes made, due to the recommendations and requests, is following below.

Specific comments: major issues and concerns:

Calibration of the DIAL needs to be discussed further. The thinking and explanation are not very scientific

- (line 87) "The DIAL technique is advantageous for measuring water vapor for several reasons, most important because it is inherently self-calibrating by its working principle". Then on line 205: The Water vapor DIAL is found to have a bias at low ranges. Then Line 210, the biases are calibrated away using radiosonde data. This removes one of the most useful benefits of DIAL, so Please discuss the magnitude of the problem, and calibration procedure further.
    - This issue is now obsolete. After the recalculation of the data, in order to meet your reasonable doubts on the correction procedure, we decided to present the data without any correction towards radiosonde profiles. Corresponding changes have been made within Sec. 2.3.
- Line 206 "However, due to the DIAL principle and the instrumental setup, this cannot be a classic overlap issue." Why not?
    - A sentence of explanation to this thought has been added to the manuscript. (l. 241-243)
- Line 207 "Therefore, we assume that there has been an issue with a detector overload which leads to this artifact." Shouldn't you be able to tell if the detector is saturating? It later is indicated that the feature that would allow the authors to tell if there are issues with clouds was turned off, is that correct?

    - An explanation to the issue of “unrecognized” detector saturation has been added at the very beginning of Sec. 2.3 and is discussed during the data analysis within Sec. 5.2.
- The calibration makes many of the comparisons not very compelling, such as at Line 604. “4 systems show good agreement in the lowest 2 km above ground”. Have not all of these systems been calibrated to the radiosonde below 1.1 km?
    - Due to the detected error in the calculation of the absorption coefficients, this valid point is now obsolete. The DIAL data isn’t calibrated anymore at all. Nevertheless, you are right in the sense that the Raman lidar data has been calibrated to radiosonde ascents. However, this calibration has only been done for one radiosonde, the data is not calibrated to every new sounding.

Rayleigh-Doppler errors in DIAL

1. In the simplified version of the DIAL equation (Eq 1) it seems the outgoing and return absorption coefficients being additive is not correct. [See Bösenberg 1998 Eq. 10 and 11] Furthermore it is suggested to have the G term in Eq 1 written out or referenced.
    - With respect to the absorption coefficients, the given sum of outgoing and return absorption coefficients equal Eq. 11 in Bösenberg 1998. . This becomes evident within the 15th line after Eq. 11 in Bösenberg 1998. However, in the former manuscript version, a negative sign before the term d/dr was missing. This has now been added.
      The term G is now directly referenced to Bösenberg 1998.
1. The authors discuss that DIAL is subject to RD errors under two regimes, (molecular backscatter higher than aerosol, and at strong aerosol gradients), then proceed to not apply a correction due to its difficulty (and/or its introducing more uncertainty). At this point, it seems relevant to note several papers in the recent literature that solve the Rayleigh Doppler problem in DIAL by simultaneously measuring the molecular to aerosol scattering ratio (backscatter ratio). This was done with a high spectral resolution lidar (HSRL) channel. This seems particularly applicable to the ATMONSYS instrument as it is well positioned to measure the backscatter ratio at 355 nm from the Raman lidar, or at 532 nm with an I2 HSRL channel (likely better as closer in wavelength). Suggest that the authors acknowledge this method as a possible means to have a reliable RD correction. This would also allow ATMONSYS to offer robustly calibrated aerosols measurements and remove reliance on Klett inversions (one of the many drawbacks of this technique shows up in Figures 6 and 7 - and noted in the caption of Figure 7)
    - Thank you very much for this interesting input. We haven’t been aware of the corresponding development/publications. We’ve now included information on this methodology with corresponding references. (l. 200-203)
1. The above is relevant as discrepancies with the radiosondes and Raman lidars may be due to gradients (RD error) as mentioned in line 630. And the proper implementation of the correction term calls for further research (line 634). The “shreds of clouds” mentioned in line 635 as the most probable cause of discrepancy seems also to be gradient problems (RD errors) but this could be addressed with an HSRL channel.
    - We agree, this is something we have to consider for the future.

Line 39 Numerical Weather Prediction. There are confusing science drivers for this instrument application.

- This instrument is well-suited for short time scales – the science driver mentioned in the opening sentences of the abstract. And again in line 27 understanding the humidity transport process (process studies) is a very reasonable science objective.
    - Ok.
- But the science driver given on line 41 needs clarification. The WMO OSCAR requirements have uncertainty, temporal, vertical and horizontal requirements (note these are listed as 'goal', 'breakthrough',and 'threshold'). So is this uncertainty requirement for process studies, or for regular observations? This instrument is not well suited to improve the numerical weather prediction for routine monitoring as it would be impractical to meet the horizontal requirements. In the same sense that it is not economically feasible to make the radiosonde observations at sufficient horizontal spatial scales to improve weather forecasts. Line 391 again references the WMO criteria <5%. And finally, line 660 references this criteria again, and alludes to monitoring and data assimilation. This does not appear to be a realistic science driver for this instrument.
    - The reference to the WMO requirements seems to have been misleading. As the ATMONSYS lidar is an experimental development, we don't claim it to be an operational instrument that immediately helps in the improvement of numerical weather prediction. However, from a developers perspective, the WMO requirements can be seen as goal system performance parameters. Recent studies incorporating single high-resolution water vapor lidars show persisting effort in the improvement of model parameters based on enhanced knowledge gained by lidar systems. Therefore, we would argue that the WMO observational requirements are a fair motivation.
      A respective sentence, clarifying that the ATMONSY lidar isn't thought to be operational, has been added (l. 42-44/736-737).

Line 197 Spectroscopic T and P dependency

- Line 197. Please explain the rationale behind using radiosondes to inform the water vapor spectroscopic line parameters. The data is short, presented from 1 hr of a single day. Yet the radiosonde becomes uncorrelated in time (the radiosonde apparently was at 10.75 UTC and used to evaluate the period 11.6 to 12.6 UTC, correct?). Would not surface measurement of the T and P (assuming a lapse rate and hydrostatic equation to get profiles) provide better results? Reanalysis data would yield even higher quality data, if that was required.
    - T and p values over the full measurement range are important in order to precisely calculate the pressure and Doppler broadened absorption lines at different altitudes. Radiosonde data, even if it is some hours old, is a better foundation for the vertical p and T distribution than just assuming a standard lapse rate based on ground measurements. Especially in higher altitudes, one can assume that also the spatial differences are not that pronounced

anymore and persist for at least some hours (under the prerequisite of no major weather change).
We cannot answer the question on whether or not reanalysis data would be more beneficial than the up-to-date in-situ information. But sure enough, the differences should be marginal as they incorporate the radiosonde data itself.

Planetary Boundary Layer heights

- Throughout the measurement section (lines 272 - 354) the planetary boundary layer heights (PBLH) are discussed. Moisture and aerosol gradients are used synonymously with PBLH. But, as is well known, these methods are proxies for the PBLH and can fail for a variety of reasons. The aerosol lidar community often overlooks this issue. The authors have the means to measure the PBL height directly using thermodynamic buoyancy (virtual potential temperature from radiosondes) or kinematics (Doppler lidar). For example, the vertical wind velocity in Figure 7 provides evidence that the PBL is around 1 km above ground level at 12 UTC (automated methods to derive the top of the PBLH from this data based on the bulk Richardson number exist). At minimum, why not use these direct methods as proof that the proxy gradient methods are correct for the time shown? This is important to provide more confidence for claims as in Line 396 "This is again an indication for the position of the PBL top" and the analysis that follows.
    - We agree with your concerns.
      The PBL height determination purely based on the gradients of aerosol/humidity may lead to deviations in comparison to the thermodynamic/kinetic energy PBL height.
      To our understanding, a proper PBL height determination based on wind data requires horizontal wind information to identify the low level jet nose. However, the Doppler wind lidar next to the ATMONSYS system has been operated only at vertical stare. Calculating the temporal standard deviation of the vertical wind on each height leads to vertical profiles of σ_w. However, it seems to be the case that an automated PBL height determination based on this measure very much relies on personal choices for thresholds (e.g. : https://doi. org/10.16993/tellusb.1876).
      As a work around, we performed PBL height calculations based on the bulk Richardson number with the available radiosonde data. The respective values are included into the manuscript (Tab. 3).
      Above that, there was a VAD scanning Doppler lidar approx. 7 km away producing horizontal wind speed data. The sparse data that is available from those measurements, however, confirms a maximum in wind speed at around 1500m AGL during that time.
      Therefore, we trust the radiosonde data as a rough estimation of the PBL height development.
      A table with the calculated values as well as explaining sentences have been added to Sec. 3/ "Measurement day: 18 Jul 2021".
- In cases where this is not possible (perhaps the overview section around Figure 5), state that PBL heights were assumed using gradient methods and, as such, might not be the actual PBL.

    - Has been added to subsection “Measurement day: 18 Jul 2021” and all further PBL height occurrences..
- Line 370. The rationale for the negative water vapor sounds reasonable but likely incomplete. Would not RD error be expected at the steep gradient? How about the effect of cloud heterogeneity? Furthermore, quality controlling the data by masking out negative water vapor might introduce problems from binning/smoothing. Why not use a gradient method to remove clouds before retrieval of the DIAL to avoid any smoothing issues?
    - The chosen method is fairly rudimentary, but quite efficient for excluding cloud-inflicted profiles.
      Questions on biased data due to RD and/or complicated cloud heterogeneity would of course arise if the data would not be entirely neglected. However, there isn’t any temporal smoothing applied beyond 10s (integration time) and the sign changing effect by signal saturation is by far stronger than any RD effects.
      The application of a cloud detection algorithm e.g. based on the computational quite intensive Haar-wavelet-transform, for this specific analysis, in our opinion, doesn’t have much benefit.

Section 4.3 Turbulence spectra

- As a suggestion, since the frequency response doesn’t have much overlap with the expected trend, perhaps plotting this data as an Allan Variance (two sample variance vs integration time) would be easier to interpret. In this case the Kolmogorov constant is +⅔.
    - Thank you for this suggestion - this method hasn’t been within our attention. However, for reasons of convenient inter-comparison with previous studies, we would rather stick to the chosen way of representation as the referenced publications by Wulfmeyer et al. 2024, Mauder et al. 2020 and Senff et al. 1994 use the same way of visualization.
- Of course it is possible the frequency rolloff at longer integration times may be due to instrument instability beyond 1 minute or longer. But the most compelling rationale for the DIAL accuracy is the similar lack of low frequencies seen in Doppler winds. The rest of the discussion regarding the reasons for the non-Kolmogorov atmosphere is tangential.
    - True. A respective sentence has been added to l. 563-566.
- Line 465 Do you mean altitude 973 m AGL and not 500 m? It is hard to tell which altitude is deviating from which. But it is clearly different at all altitudes from the Doppler wind spectrum.
    - In comparison to the other lines, from our perspective, the energy drop in the 500m line seems to be most clear, but you’re right that this feature can be observed in the other lines as well. The sentence has been changed accordingly.
- Line 662. “ A spectrum analysis of the DIAL showed good agreement with Kolmogorovs…” This conclusion is unjustified. What was shown was good agreement with the Doppler lidar frequency spectrum. And that the Kolmogorov inertial subrange rolled off at low frequencies for some reason or other (which is not really necessary to explain)

- The respective sentence has been changed accordingly.

Technical corrections: minor grammar, misspellings, or strange word choices

Line 10. ‘Evaded’ perhaps 'explained' would be better?

Not sure whether “explained” is correct either, we changed it to “overcome”.

Line 16. Shreded is misspelled, but suggest changing to ‘broken clouds’

Ok.

Line 46. Deeply requested. Suggest changing to ‘often requested’

Ok.

Line 140. ‘Begin of the lidar range’. Suggest ‘start of the lidar range’

Ok.

Line 166. ‘Renowned DIAL equation’. Suggest changing to ’well-known DIAL equation’

Ok.

Line 281. ‘Surpassing the lidar’. Suggest “passing over the lidar”

Ok. Has been changed at all 3 initial occurrences.

Line 336 and 633. Straight forward should be one word, straightforward

Ok. Has been changed at all 3 initial occurrences.

Line 370 ‘Supersaturated’. Not a good word choice in English for this condition. Is it something in the electronic gain saturated or perhaps a non-linear response of the detector (perhaps some combination of both?). Suggest ‘saturated non-linear response’

Ok.

Line 379 ‘Spread is weaker’. Suggest ‘spread is reduced’

Ok.

Line 621. 40 min hour. It seems the word hour is unintended

Indeed. Thanks for noticing.

Line 636 and line 675. ‘Shreds of clouds’. Suggest ‘wisps of clouds’

Ok.

Line 639 ‘Such flags have not been set in the corresponding time’, Unclear what meaning is desired here.

Ok, “by the transient digitizer” has been added in order to make the meaning better understandable.

Line 644 ‘Proof’ is the wrong word. Use ’prove’

Yes, thanks.

Line 663 ‘Interludes’: Suggest ‘portions’. Actually the spectrum doesn’t agree well beyond at time scales longer than approximately 1 minute

Ok.

Citation: https://doi.org/10.5194/amt-2024-168-RC1