General Remarks

We thank both reviewers for their time and the constructive comments, which improve the quality of our manuscript significantly. The referee comments are formatted in grey and our reply in black with indentation and numbering (R##). The line and figure numbers refer to the revised version.

Specific Reply to Referee #1

In the manuscript "Continuous wintertime water vapor profiling by Raman lidar at Neumayer Station III, Antarctica: Characteristics of meridional moisture transport and assessment of ERA5 reanalysis" by Jakob et al., the authors provide significant and valuable measurements of Antarctica region. The dataset is highlighted by high-resolution water vapor profiles measured by ground-based lidar PollyXT covering an entire wintertime. Two characteristic synoptic regimes were identified as primary drivers governing moisture transport toward Antarctica's coast. The evaluation of ERA5's performance in the Antarctica region was also conducted by comparing with PollyXT. I recommend the acceptance of this manuscript after some minor revisions. The detailed comments are listed below:

1 line 50: Please also provide the periods of polar night and polar day, for the reader to understand Fig. 1c better.

> **R1:** The exact dates of polar night (25th May to 20th Jul) were added to the caption of Fig. 1.

2 line 123: I understand the authors averaged lidar profiles to match the ERA5 vertically and temporally. But how did the authors choose the "ERA5 profile"? Did the authors only choose one profile or averaging profiles of ERA5? Please clarify.

> **R2:** Thank you for raising this point. Only one ERA5 profile was used per comparison, the nearest gridpoint to Neumayer Station, without spatial averaging. We clarified this in line 124: "…, taking the ERA5 profile from the nearest grid point (70.75° S, 8.25° W)".

3 line 127: What and when is the automatic polarization calibration? How did this procedure impact the measurements?

> **R3:** The automatic polarization calibration is required to derive the particle depolarization ratio. The calibration is performed three times per day at 00, 12 and 18 UTC and takes 10 min. It causes a short measurement gap, which is automatically flagged in the data processing, but does not affect the 387 and 407 nm signal used here and therefore as no impact on the quality of the water vapor measurements. Further details are given in Freudenthaler (2016, https://doi.org/10.5194/amt-9-4181-2016), which we added as reference to the manuscript in line 127.

4 line 122: The low SNR results from the attenuation of the signal by the clouds?

> **R4:** Yes, the low SNR mainly results from the attenuation by clouds, but also aerosol and molecular extinction attenuate the signal, which is especially relevant for higher altitudes. Therefore, the Raman signal is generally insufficient above 6 km altitude to retrieve high-resolution water vapor data.

5 line 162: How did the authors get these proportions? How many cases did you analyze totally?

> **R5:** The proportions were derived by manually classifying the synoptic situation into pattern A, B or other at 6h intervals over the 4-month periods. This yields a total of 492 time steps, of which 112 time steps (=28 days) were classified as moist air events. We clarified this in line

149: “The classification was done continuously from May to August in 6h intervals”. See also R9 in Reply to Referee #2.

6 line 200: More details about “TRACE” should be introduced.

R6: We added a brief clarification in line 200: “…, which is based on FLEXPART backward simulations”, to indicate that TRACE relies on a Lagrangian particle dispersion model. Further details are provided in Appendix B and in the corresponding publication (Radenz et al., 2021a, b), which we also referenced in line 288. We consider this sufficient for the reader to follow the analysis presented here, and think that adding more detail would shift the focus away from the main analysis.

Specific Reply to Referee #2

This paper describes a 4-5 month water vapor mixing ratio profile dataset collected by a Raman lidar deployed to Neumayer Station III in April – August 2023. The paper is well-written and clear. I have a few items that need some additional clarification.

Line 69 – assuming that no overlap correction is needed seems possible, but in my experience somewhat unlikely. As Neumayer Station is launching radiosondes, why were those data not used to evaluate the need (or not) of the overlap correction?

R7: Thank you for raising this point, we indeed did not include these technical details in the manuscript, because they are discussed in previous literature. For the water vapor retrieval, we use the ratio of two signals (387 and 407 nm) which are both excited by the same wavelength (355 nm). The reason for not applying an overlap function is based on the fact that the same wavelength and telescope is used for both signals which yields the same overlap function and if the overlap functions of both channels are equal, they cancel each other out (Wandinger, 2005; Engelmann et al., 2016). Overlap effects cancel out down to an altitude of 200 m (Engelmann et al., 2016). In the lowermost 200 m, the overlap of telescope and laser beam is very low, leading to near-range effects, such as angle-dependent filter transmission and spatial inhomogeneities in detector sensitivity, which can lead to differences between the two channels (e.g., Freudenthaler, 2004, https://adsabs.harvard.edu/pdf/2004ESASP.561...37F).
Radiosondes were also used in the calibration procedure and the comparison between lidar and radiosonde profiles showed reliable measurements at heights above 200 m (see also Fig. 1b), which is the lowest height used for the calibration. We state now this lowermost altitude boundary in the manuscript, line 70: “…, allowing reliable WVMR measurements down to 200 m height” and also consistently use it as the lowermost height throughout the analysis. Specifically, we revised line 116 to read: “The second level, with a lower bound of 200 m above ground, was used as the lowest comparison height”. We also adjusted the scale of Figures 2 and 4 – 8 to make this boundary clearer.

How high above the surface did the Raman lidar need to profile in order for the sample to be considered comparable with the MWR’s retrieved IWV? Was the Raman lidar data screened first to make sure it achieved a certain SNR before it was used in calibration?

R8: The lidar signal ratio is integrated from the ground to 6 km to obtain the IWV, with the lowermost 200 m assumed well mixed and set to a constant value. Water vapor above 6 km is mostly negligible (Foth et al. 2015). Based on the available radiosonde data, the mean IWV was 4.33 kg/m^2 for the full atmospheric column versus 4.27 kg/m^2 when integrating up to 6 km only, meaning about 98.5 % of the total column water vapor is contained within the lowest 6 km.

Yes, before calculating the IWV the Raman lidar data undergoes quality checks, including SNR screening. If the SNR is not sufficient or clouds are present below 6 km, this particular profile is not used for the calibration.

Line 144: you indicate two dominant synoptic patterns were identified, and that this classification was done manually (line 147). However, it seems that only this relatively short 4-5 month dataset was used. I believe this work would be stronger if a multi-year ERA5 period was analyzed to classify synoptic conditions objectively (e.g., using a self-organizing map or other techniques); afterwards, you can then see how the data from the Raman lidar analysis period is classified.

R9: We agree that using an established weather pattern classification algorithm for a longer period instead of a manual classification would strengthen our analysis, and thank the reviewer for this suggestion. We would like to note that our initial motivation for identifying synoptic patterns was primarily to understand which large-scale patterns coincide with the observed moisture events and the moist air intrusion, and whether this is reflected in the 2D moisture structure. Therefore, the focus of this study remained on the water vapor observations rather than on the synoptic classification itself. Nevertheless, we followed the reviewer's suggestion and tested an existing, objective classification instead of relying on manual assessment, since training a self-organizing map was considered beyond the scope of this study. We applied the automated Jenkinson-Collison weather-typing scheme (Jenkinson and Collison, 1977; Jones et al., 1993, https://doi.org/10.1002/joc.3370130606; Fernández-Granja et al., 2023, https://doi.org/10.1007/s00382-022-06658-7), following the implementation of Herrera-Lormendez et al. (2023, https://doi.org/10.1002/joc.8033), and classified the synoptic pattern at Neumayer in a 6-hourly resolution based on the 850 hPa geopotential height field. Assigning northerly/north-westerly (N, NW) advection to pattern A and north-easterly/easterly/south-easterly (NE, E, SE) advection to pattern B, the automated classification agreed well with our manual classification yielding the same classification results for the moist air events. Applying it to the entire analysis period likewise reproduces our manually derived distribution closely, with only a slightly larger fraction classified as "other" at the expense of pattern B. We consider this a valuable independent confirmation of our manually derived classification and have added a corresponding sentence to the manuscript, line 150: “As an independent validation, this manual classification was compared to the automated Jenkinson-Collison weather-typing scheme (Jenkinson and Collison, 1977; Jones et al., 1993; Fernández-Granja et al., 2023; Herrera-Lormendez et al., 2023), applied to the 850 hPa geopotential height field at Neumayer, which showed good agreement with the manually assigned patterns.”

Figure 3: the upper panels need to have something on the x- and y-axes to indicate spatial distance; either lat/lon or km. I am unable to ascertain the scale of these images without this information.

R10: Thank you for this helpful suggestion. We added longitude and latitude labels to the gridlines.

Line 220: You are discussing the relative biases in the ERA5 from the daytime assimilation cycle (which has the radiosonde from Neumayer available) vs nighttime. Further, you state that the ERA5 bias is smaller at night, even though that cycle did not have a sonde to assimilate. However, this leads to two questions: (1) why is the vertical shape of the daytime vs nighttime so similar? (2) Why is the nighttime bias slightly smaller than the daytime bias? Could this be due to a bias in the Raman lidar data that results in misleading interpretation?

R11: 1) We attribute the similarity in vertical shape between the daytime and nighttime profile to the boundary layer parametrizations effects dominating over the influence of the assimilation cycle. To further investigate this, we compared ERA5 profiles to the daily radiosonde profiles using the same approach as for the lidar (see Fig. 1 below). This comparison yields a very similar vertical shape of the ERA5 bias, though of course with the much smaller radiosonde sample size.
2) We do not have a definite explanation on why the nighttime bias is slightly smaller than the daytime bias. One possibility could be that during the night more passive microwave sensors are assimilated, but we do not know that for certain. Unfortunately, the ERA5 reanalysis observation feedback files, which could provide us with such information, are not publicly available nor on request.

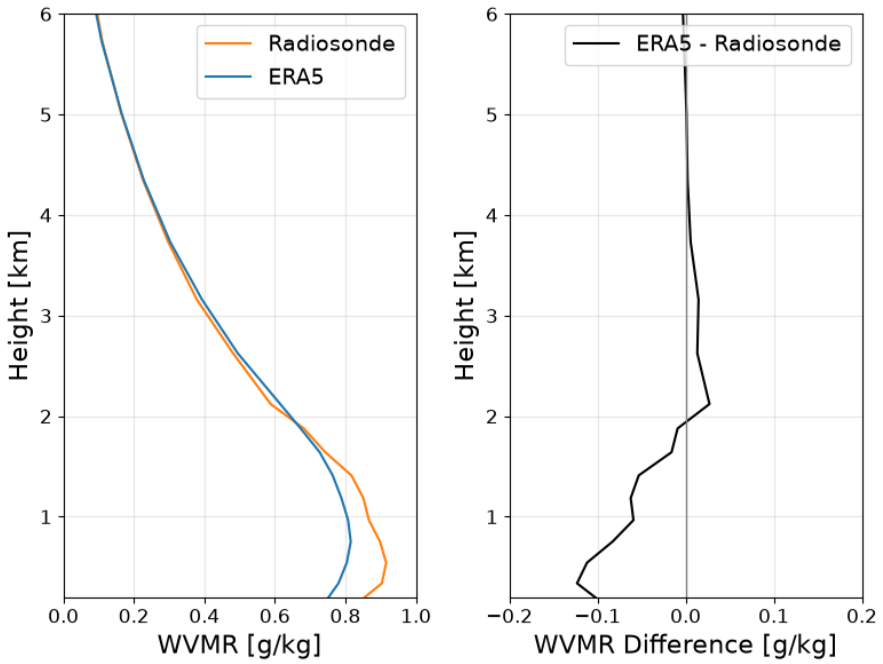

Figure 1. Height profiles showing the mean WVMR of ERA5 reanalysis (blue line) and radiosonde (orange line) (left panel) and the mean absolute difference in WVMR between ERA5 reanalysis and radiosonde (right panel) for the analyzed period (1st May to 31st Aug).

Same figure: The bias of 0.1 g/kg in the lowest 1 km seems to be a constant feature. Is this an artifact of the Raman lidar (e.g., because no overlap correction was applied)?

R12: As discussed in our responses above (R7 and R11), we are confident that any overlap effects are limited to below 200 m and do not affect the measurements above this height. The constant bias in the lower troposphere can be also seen in the independent comparison to radiosonde profiles, which supports our conclusion that this is a bias in the ERA5 reanalysis and not an artefact of the Raman lidar.

Line 273: it may not be only the boundary layer parameterization, but could also include errors in the surface layer parameterization and surface properties.

R13: Thank you for this valid suggestion. We fully agree and the manuscript was adapted as follows, line 270: "… and surface properties" was added, line 273: "… and surface layer parametrizations" was added.

Further changes

A few changes were incorporated due to additional feedback.

1. In line 185 there was a mistake in the analysis period: “1 May–30 September” was changed to “1 May–31 August”.
2. In the caption of Fig. 2 the following sentence was added: “Red arrows mark the example cases discussed in the text for patterns A and B.”, as this information was missing.