Manuscript Review

Title: Cluster Analysis of Vertical Polarimetric Radio Occultation Profiles and Corresponding Liquid and Ice Water Paths From GPM Microwave Data

Authors: Jonas Katona, Manuel de la Torre Juarez, Terence L. Kubar, F. Joseph Turk, Kuo-Nung Wang, and Ramon Padulles

Reviewer Recommendation: Minor Revision

Overall Quality:

In this study, the authors evaluate the ability of k-means clustering to find relationships among polarimetric phase difference, refractivity, liquid water path (LWP), ice water path (IWP), and water vapor pressure using over two years of data matched between the GPM mission and ROHP-PAZ. They develop a refractivity model to ascertain how different types of vertical thermodynamic profiles that can occur during different precipitation scenarios are related to changes in the polytropic index and thereby vertical heat transfer rates. The authors' main conclusion is that clustering in the manner introduced in this study confirms its value as a tool for quality control of profiles and can automate the classification of physical phenomena found across large datasets, thereby avoiding the need to inspect and compare profiles individually. I believe that this work is, overall, a well-written manuscript and will provide benefits to both the RO/PRO and precipitation communities, although it will first be necessary to clarify some issues and expand on details/explanations. My recommendation for this paper is that a minor revision is necessary for publication and my review is followed.

Specific comments:

1. The abstract is well-written, although there should be a more definitive concluding sentence summarizing the overall main conclusion of the work (e.g., what is the usefulness or benefits of this study toward the science?)
2. Lines 39-47: There should be a brief summary of the RO 1D-Var method if the authors are going to mention how water vapor information can be extracted from refractivity profiles. A good resource to use is from Wee et al. 2022 (**https://doi.org/10.3390/rs14215614**).
3. Lines 87-88: What are the collocation criteria used in this study? These values need to be explicitly defined in this sentence (e.g., 3 hours/300 km - whatever values you used).
4. Figure 1 caption: Sampling distribution of what? This caption needs to be more detailed. I assume this is showing the locations/times of the collocations, but readers shouldn't

have to guess. Also, this also raises another question that I do not see answered in the methodology: What tangent point lat/lon is chosen for this figure? Nearest to surface? 10 km? Please add this to the paper.

5. Line 90: Why is there a data gap in Jan/Feb 2019?
6. Lines 100-101: The authors write "By checking when the retrieved temperature is above or below 273 K, we partition this integrated water content into LWP and IWP, respectively". Is it appropriate to simply use the freezing point of water to differentiate between LWP and IWP when there are often supercooled water droplets in clouds?
7. Lines 177-180: Why do the authors use this definition of the tropopause to identify its location, rather than use an established and commonly used definition, such as the lapse-rate tropopause definition from the WMO: "the lowest level at which the temperature lapse rate decreases to 2 K/km or less. To fill this condition, the average lapse rate between this specific level and all higher levels within 2 km should not exceed 2 K/km."
8. Line 197: What are some examples of "nonphysically high water vapor pressure values (> 300 hPa)" and what are some possible causes for these erroneously retrieved values?
9. Lines 229-232: The authors write a summary from previous studies: "Later studies (Muller et al., 2009; Holloway and Neelin, 2010; Emmenegger et al., 2022) demonstrate a positive relationship between precipitation and total column water vapor in the tropics, where under a certain total column water vapor value, precipitation is generally negligible in a given profile, and above a "pickup" threshold, precipitation may become non-negligible and increases exponentially." I expected to see the authors apply this same idea towards their own study, with a focus on Figure 3, especially because in lines 233-234, the authors write "we look for the precipitation pickup pattern (Holloway and Neelin, 2010) in the relationship between the total column water vapor and the total column of the PRO observable". However, I don't really see that analysis provided. As readers, yes, we can attempt to visualize a relationship in Figure 3. But I'm not sure I'm convinced by a precipitation pickup pattern in any of those panels. For example, in each of those panels, at what total column water vapor value do the authors see this "precipitation pickup pattern"? It's not easily discernable to me. Additionally, how much are those moving averages influenced by a handful of large phase difference outliers? I

think this specific analysis needs to be thought about more carefully and expanded upon in order to be considered a useful result.

10. Lines 272-273: The authors write “Furthermore, Nmodel is fit across most of the troposphere down to 2.5 km. Hence, Nmodel is most effectively sensitive to concentrated moisture anomalies within narrow bands of the troposphere.” I’m struggling to understand this statement. Wouldn’t it be the opposite (e.g., the Nmodel would be least sensitive to concentrated moisture anomalies since it is fit across most of the troposphere, therefore missing those thin moisture anomaly layers)? The modeled N should not be as effective in regions where large and sharp moisture anomalies/gradients are present, such as in deep convection, where rapid vertical moisture transport commonly occurs. Can the authors elaborate on what their statement means (both in response to this question as well as in the paper)?
11. Line 279: The authors write “Fig. 4 shows three examples where N − Nmodel does not correlate strongly with ΔΦ.” However, it is not clear to me how Fig. 4a follows that statement – how would the authors expect those refractivity profiles to look if they were correlated with the phase difference profile?
12. Lines 291-292: Do the authors have any suggestions as to what the possible retrieval issues might be?

Technical Corrections:

1. Lines 32-38: You don’t need to start new paragraphs so often, as it is unnecessary to have a separate paragraph for a single sentence.
2. Lines 36-38: This has also been done in recent years using both TRMM and GPM for both tropical and midlatitude deep convection – add Johnston et al. 2018 (**https://doi.org/10.1002/2017JD027120**) and Johnston et al. 2022 (**https://doi.org/10.3390/atmos13020196**).
3. Line 51, along with many other locations: The authors seem to be differentiating between the liquid and ice phases of water by consistently writing “*precipitation or ice*”. Precipitation generally encompasses any form of water that forms and falls to the earth, regardless of whether it is ice or liquid. Thus, I recommend changing any of these instances in the paper to “liquid or ice” or something along these lines. Or are the authors

trying to differentiate between precipitating and non-precipitating features? Because non-precipitating features can also be either liquid or ice.

4. Lines 74-75: delete “from PRO derived refractivity and $\Delta\Phi$, to model-inferred water vapor, water path, and ice path”.
5. Line 260: change “relative humidity” to “specific humidity”. You can still have a higher relative humidity in regions where there is very little moisture present, and as a result, you would not see a big impact to a refractivity profile.