A new polarimetric radio occultation metric is used to assess the fidelity of different microphysical schemes in multiple simulations of atmospheric rivers. The results seem robust showing that xsnow~0.1 (aggregates) provides the best results and it is a good demonstration of a novel measurement that could be used to test other schemes and build climatologies.

This work should be publishable subject to satisfying the following points.

We sincerely thank the reviewer for their constructive and insightful comments, which have helped us improve both the clarity and robustness of our work. In this revised version, we have made several important changes to enhance the physical consistency and transparency of the methodology.

(1) We calrify that there is no direct coupling between ARTS and WRF for the main part of the analysis. Instead, ARTS is used diagnostically to evaluate which paritcle habits are most compatible with the WRF-derived water content and the best x-parameter obtained when comparing with actual observations of differential phase shift.

(2) Second, the optimization process has been refined and is now carried out at two levels: (i) within each microphysics scheme, to obtain the optimal set of x-parameters; and (ii) across all schemes, to identify the combination of microphysics and x-parameters that minimizes the cost function.

(3) In the comparison between the best x-parameter and those derived from ARTS, we now generate two distinct look-up tables that relate Kdp and WC, each corresponding to a different assumed particle size distribution (PSD). This modification ensures a closer alignment between the scattering properties used in ARTS and the microphysical assumptions in WRF.

Main point.

This work is introducing and demonstrating a novel metric to be used to test cloud microphysics representations. To provide context and convince the reader of its value it would be good (necessary?) to also provide comparison to more traditional metrics. Comparisons should be made to readily available satellite derived precipitation and top of atmosphere broad band radiation, perhaps even vapor and liquid water path too. Ideally these comparisons will also demonstrate the Goddard scheme performing best and supporting the result of the PRO analysis.

This would just form an extra section and add some paragraphs the results/discussion and conclusions.

We agree that such comparisons are valuable for building a comprehensive analysis of the model performance. However, the main goal of this work is to introduce and demonstrate the potential of PRO as a novel diagnostic for cloud microphysics, rather than to perform a full evaluation of the schemes against several available observational instruments.

To provide some context, we already include in Figure 3 and Figure 4 comparisons between the integrated vapor transport (IVT) and the accumulated total precipitation for the different schemes against ERA5 and also against longwave IR brightness temperature. Extending the analysis to a full set of additional metrics would require a full dedicated study and it is out of the scope of this analysis.

Other points.

line 24. I did not see anywhere a discussion about the horizontal resolution of this approach. That needs to be delved into along with a discussion about the pros and cons of using these long path lengths (~200km?)

The horizontal resolution in the along direction of the radio occultations is large, being a limb sounding technique and specially in comparison with other spacial techniques like infrared or passive microwave imagery. However, the rays can be resolved with very fine resolution, and when interpolated into the model grid, the horizontal resolution becomes less relevant.

It is true that since the contribution to the differential phase shift at each ray can come from everywhere along the ray, values along the ray can be compensated and there is an intrinsic ambiguity.

line 70. Perhaps the spatial coherence of these phenomena is also good for this approach that has coarse horizontal resolution?

Exactly, the selection of Atmospheric Rivers was made principally because of their spatial extension that algins well with the larger horizontal resolution of radio occultations.

line 84. This section would benefit from outlining the geometry of the sampling, perhaps with a schematic? Section 2.1 and 2.3 should also be merged? I read 2.1 and wondered why the ice phase was being ignored.

We have added on Figure 1 a second peanel showing the geometry of the occultation rays for that specific case.

Sections 2.1 and 2.3 are separated to make more easily the understanding of the methodology, first by introducing the concept of radio occultations and then to explain the observational operator used.

We have rephrase the sentence at the beginning of section 2.1 to make clear that we are considering also frozen hydrometeors.

line 127-128. What were the problems - it did not seem to be mentioned again.

The problem with RRTMG arose in a small subset of simulations where the model integration terminated prematurely due, from our prespective, numerical instability. This instability we think it was related to the coupling between the RRTMG radiation and microphysics in those particular cases, and we were unable to resolve it. We can not confirm the exact cause but we suspect that could be because of the presence of strong moisture transport.

line 140. Can probably just omit 'new'?

Omitted

line 158. Was a single instantaneous out chosen? Or was it averaged in some way. How close was the sample to the output in time?

We used the WRF output closest in time to the PRO observation. The temporal mismatch was always less than +-30 minutes. We did not average over multiple outputs.

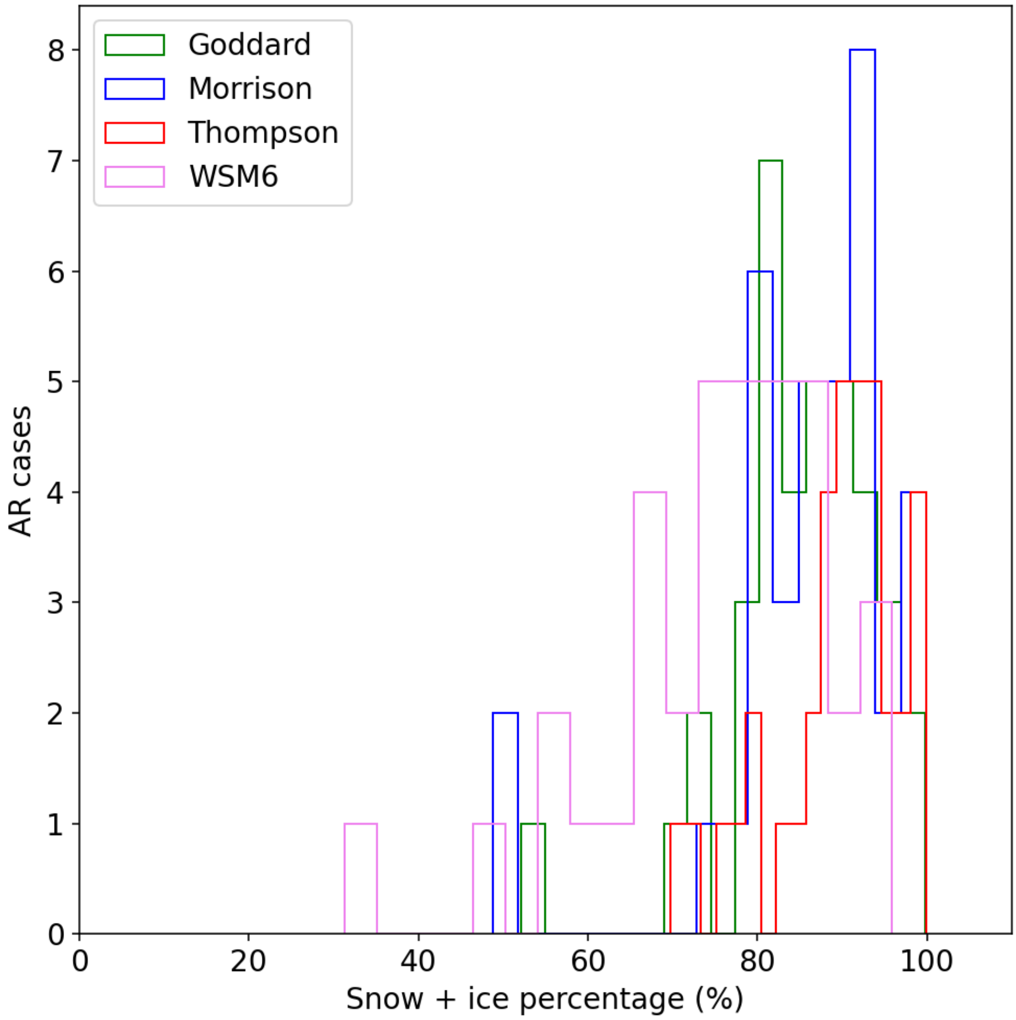
line 170. See the main point. This paper can be strengthened substantially with some additional comparisons to satellite data to support the findings from this paper.

See response to the main point

line 227. The use of 50 assumes that each point is truly independent. Are these 'points' next to each other as in the gray area in figure 1? If so then they seem to be spatial coherent over scales of order 100km or so and lower number than 50 might be more appropriate?

No, this not assumes that each point is truly independent. The measurements each 1 second have influence of the 50 points next to it. The measurement frequency os of 50Hz.

line 248 and throughout. These schemes all use different size or mass thresholds to determine what is ice and what is snow. That could lead to big apparent differences between them that are not necessarily incompatible. For figure 5 do the schemes look more similar if ice+snow if plotted?



Correct, it kind of looks more similar for the four schemes. Indeed, the distinction between "ice" and "snow" varies among the parameterizations, since each scheme applies different size or mass tresholds to separate the two categories.

line 262-265. I was confused here. Equation 7 define iWC and delta Phi as being the integral along the limb sounding path. Here it talks about vertical integral? I am assuming that the WRF output is integrated along similar curved paths through the domain? How is that done?

The WC employed in the analysis is obtained from the WRF simulations. In order to obtain the Kdp by multiplying WC by the x-parameters, we first interpolate the WC obtained from WRF into the PRO rays, and then integrate the WC along each ray.

figure 6. All 4 schemes are represented - does this mean that each one was best in different cases?

Maybe i am misinterpreting.

Correct. For the different observations the best combination of WRF+ARTS yields to an effective microphysics, in general the one obtained for the majority of the cases is the Goddard however, for several others the rest of the schemes perform better.

Why not show all results - maybe a different panel for each scheme if it gets too messy?

We are showing the best performing scheme for each of the cases, that's why we choose to represent all of them in the same figure. Our goal is to show that no aparent relation exists. If we plotted them all, some of them would correspond to really bad fits, that is, maybe one point in the plot would be the best for a particular microphysics scheme, but with a really bad fit to the data. Then, these type of points would not be representative.

line 272. DeltaPhi is defined as dependent on WC and x. Does a surface regression against WC and x result in a better result?

It is true that other approaches, such as a surface regression of $\Delta\Phi$ against both WC and x, could be explored. However, since the WC is a simulated vairable and $\Delta\Phi$ is the observable, our goal is to design the analysis so that WC from WRF is kept fixed, while the x-parameters are adjusted within physically plausible bounds. This separation allows us to asses which microphysics schemes provide water content fields that are more compatible with the PRO observations, without directly fitting WC to $\Delta\Phi$. A regression including both WC and x would risk absorbing biases in WC into the fit, making it more difficult to know whether discrepancies arise from one part or the other.

line 278. But how can they all have the same WC and delta Phi axes in figure 7? The different x values will mean that each species should have its own delta Phi for a give WC?

There are two different x axes in Figure 7, one for the represented integrated water content and the other for the differential phase shift (corresponding to the PRO observation). As i can be seen in equation 4, yes, the different hydrometeors have different contributions to the differential phase shift.

figure 7. Maybe some more discussion about how the data is used is needed. It looks like you are plotting the curved limb path iWC value at its lowest point(?)

In Figure 7 we plot the integrated water content along each of the PRO rays vs the differential phase shift obtained from the PRO observation. Remember that iWC means integrated along the full ray-path (not vertical integral).

line 282. Can you tabulate J and the error. Perhaps indicate where the lowest J is significantly lower than the others?

The Appendix already contains a plot showing the cost function values, showing where the maximum and minimum values are encountered. For this reason, we have not included an additional table, since it would essentially duplicate the information already provided in graphical form. Moreover, the errors in Figure 11 are directly correlated with the cost function, which allows the reader to identify where the lowest values of the cost function coincide with the smallest errors line 318-319. It would be useful to locate some ground based polarimetric studies of ARs to see if they agree with this result.

The study [1] was cited in relation with the WSM6 and Thompson schemes for their performance on Atmospheric River simulations.

[1] Jankov, I., Bao, J. W., Neiman, P. J., Schultz, P. J., Yuan, H., & White, A. B. (2009). Evaluation and comparison of microphysical algorithms in ARW-WRF model simulations of atmospheric river events affecting the California coast. *Journal of Hydrometeorology*, *10*(4), 847-870.