Reviewer I

The work is well written and presented. The validation of the CRTM radar operator along with an exploration of its sensitivities to microphysical assumptions is significant and of interest to the community. The CRTM radar model, being supported, open source and publicly available, is likely to be used in many future studies as well as in weather models. However, the study is currently incomplete because it is lacking any comparison between CRTM and the radar models being used by the level 2 data providers. This is the first place to look to explain the differences between the CRTM simulations and the observations. There are a number of other smaller issues to be addressed as well.

We would like to express our sincere thanks to Dr. Geer for his positive remarks and thoughtful suggestions. We fully agree that our findings are influenced by the retrievals, particularly by the radiative transfer code and its underlying assumptions. In the revised manuscript, we now provide a clearer discussion comparing the microphysical assumptions used in our study with those applied in the forward modeling components of the retrieval system.

Main points

The methodology of this study is slightly circular, as explained above. Although this approach is justifiable for a model validation study, it needs to be more clearly flagged to readers, and its implications followed up in more detail. In this context, if the the CRTM radar operator was identical to the forward model used in the level 2 retrievals, and if no information had been lost, then we would hope for an almost exact match between the CRTM simulations and observations. Already, the level of agreement between simulations and observations in terms of spatial structures is nearly exact, for example in Figure 1, which clearly shows that despite other passive data being used in the CloudSat retrievals, they must be very much dominated by the reflectivity.

We thank the reviewer for this insightful comment. We have added two paragraphs near the beginning of Section III to address this issue. While we agree that some dependencies exist, we still believe, as noted by the reviewer, that these retrieval products provide a unique opportunity for evaluating forward models.

Despite the close spatial agreement between simulations and observations, there are systematic differences in the reflectivities which are of interest and are most likely explained by differing microphysical assumptions in the two forward models, or by different model methodology (such as treatments for single or multiple scattering, or cloud overlap and inhomogeneity within the beam). Another possible explanation is if some of the reflectivity-generating hydrometeor mass has been lost between the retrieval and the input to the simulation (for example, if the snow mass has been somehow discarded in the DPR comparison, which only uses cloud and rain hydrometeor profiles). I would hope that all results could be explored in more detail in this framework, following these suggestions:

The discrepancies are likely due to a combination of differences in MP assumptions and the exclusion of some hydrometeor types. We have expanded our explanation of these points in the revised manuscript. As the reviewer also noted in later comments, if a hydrometeor type is not explicitly retrieved, it might be implicitly included in the water content profiles for other habits. We have also include two tables: one outlining the MP assumptions used in the retrievals and another summarizing the water content profiles and associated shapes employed in the CRTM simulations.

1) At a high level (e.g. abstract, introduction, conclusions) the work should indicate more clearly that the source of the hydrometeor profiles is the observations themselves, and that the study can be more clearly seen as a consistency validation against the existing radar models employed in the level 2 retrievals.

We thank the reviewer for this comment. We have discussed this in detail in Section III and also provided a brief explanation in the Introduction as follows.

> There are three main sources of input profiles for validating RT models: in situ measurements, NWP simulations, and retrievals from combined active–passive sensors. In this study, we use retrieval products to avoid temporal and spatial mismatches between the observations and input profiles. While these retrievals involve radiative transfer assumptions, they still provide a practical and reliable basis for evaluating forward models.

2) The work needs to detail more exactly the radar forward modelling being used by the data providers, especially the basic assumptions such as single particle scattering model or habit and particle size distribution, and any more advanced methods used for multiple scattering, cloud overlap or inhomogeneity. Also a little more information needs to be provided on the CRTM configuration in these areas. I imagine it should be possible to provide a new table comparing exactly the configuration of CRTM to the configurations of the forward models used in the Cloudsat, EarthCARE and DPR retrievals.

We've added a new table (Table 1) that compares the assumptions used in CRTM with those in the radiative-transfer models behind the retrievals. Where appropriate, we also updated the manuscript to reflect the information summarized in Table 1. In addition, we have included a second table that lists the water-content profiles used as input to CRTM.

3) The work needs to more clearly summarise the sources of hydrometeor mass being used in each comparison, since it differs. Again, this would be very helpful if presented in tabular form. It is also important to double check the retrieval methods and see if any hydrometeor mass has been missed. For the DPR example, where CRTM is only supplied with liquid water and rain profiles, if all the reflectivity above the melting layer has been attributed to supercooled droplets in the DPR retrieval, then even if in practice some of this reflectivity has been generated by frozen particles, it would not matter for the consistency validation, as long as the rain and cloud scattering models were identical. But if the DPR retrieval represents frozen particles separately, then mass may have gone missing in the comparison. Given the relatively good agreement shown in Figure 3, there is unlikely to be missing mass in the DPR comparison, but it would be important to be sure, and this would help reduce the number of knowledge gaps in the study.

We have included a new table showing the water-content profiles used as input to CRTM. In addition, we believe that the mass associated with excluded hydrometeor habits is not entirely missing but is implicitly represented by other habits. For instance, graupel water content may be absent from the retrievals, but it is likely incorporated into the ice or snow water-content categories.

4) Based on the comparisons listed at points 2 and 3 above, it should be possible to explain many of the systematic differences between simulations and observations. For example, the conclusion that the Thompson PSD worked better the Abel PSD in simulating DPR is not useful on its own (this is not an independent validation) but it is useful when compared to the assumptions used by the DPR retrieval team.

We have better explained the results but we believe the error due to Abel's model (more than 20 dBZ) is more than anything that can be explainable by differences in PSD distribution. This is now better discussed in "Section 4.2: Impact of PSD on Ku/Ka".

5) Since there is a large body of existing work on radar simulators, much of which has fed into the forward models used for the Cloudsat, EarthCARE and DPR retrievals, it would be good to acknowledge this and cover it in the introduction too.

We thank the reviewer for the suggestion. We have included a new paragraph in "Introduction" acknowledging the role of previous studies in radar scattering calculations.

Minor points

1) Line 15 - radiative transfer models also critically include the surface, not just the atmosphere

The introduction has been substantially revised, and although this specific sentence no longer exists in the revised version, the reviewer's point has been taken into account.

2) Line 35 referencing the ZmVar model: there is a more recent publication that could be useful to reference here: "Direct 4D-Var assimilation of space-borne cloud radar reflectivity and lidar backscatter. Part I: Observation operator and implementation, by MD Fielding, M Janisková - Quarterly Journal of the Royal Meteorological Society, 2020". Here, especially the two-column treatment of cloud overlap for the attenuation calculation is quite novel.

We have included a new paragraph in "Introduction" discussing the new publication by Fiedling and Janisková (2020).

3) Line 72, the radar equation needs a little more explanation, since this refers more precisely to the equivalent radar reflectivity (or reflectivity factor) rather than the actual reflectivity. Here there should also be some support from prior textbooks or papers: for example Grant Petty's book has a good introduction of these concepts.

We have completely revised the section describing the radar forward model. Please see the revised Section 2.2.

4) Line 225: "Discrepancies between observed and simulated reflectivities can be attributed to a combination of inaccuracies in the input profiles, forward model errors, and observational biases". Based on discussions above, I would not expect observational biases to have any relevance here.

We thank the reviewer and have revised the text as follows:

> Several factors contribute to discrepancies between observed and simulated reflectivities. These include differences in PSDs and particle shape assumptions used in the CRTM calculations compared with the assumption made during retrieving these products from observations, uncertainties in the input profiles, and the exclusion of certain hydrometeor types from the simulations.

5) Line 236, "CloudSat shows overall a better agreement. For both instruments, the simulations tend to overestimate reflectivity at lower values but underestimate it at higher reflectivity values." This discussion needs to be more carefully framed because the CloudSat

results are within tropical cyclones whereas the EarthCARE results are global. The difference in agreement could be purely due to the different datasets and have nothing to do with the retrievals or the sensors themselves.

Both CloudSat and EarthCare results are based on global datasets, and this has been reflected in both the text and the figure caption. This section has also been largely revised.

6) Figure 2a is hard to reconcile with Fig 1, which shows simulations in convective cores reaching 30 dBZ, compared to observations rarely exceeding 20 dBZ. Figure 2a shows the opposite and the text concludes that the simulations are too low, not too high.

Figure 2a corresponds to the EarthCARE CPR, whereas Figure 1 shows CloudSat CPR data. The appropriate comparison is therefore between Figure 2b and Figure 1. However, Figure 2b uses global data, while Figure 1 is based on a tropical-cyclone overpass. The section discussing the results has been substantially revised in the updated manuscript.

7) E.g. line 246 "included liquid and rain content". Partly repeating earlier main points, it would have been much easier to follow the results if the basic details of the CRTM simulations had been presented and summarised, ideally in a table. This would include basic details like particle shape and PSD choices for each hydrometeor type in each simulation, along with the hydrometeor types being used in each case.

We have included a new table that presents the CRTM assumptions and the water content profiles used in each set of simulations.

8) Figure 3 and others show a freezing level on the observations but not simulations, which makes it harder to compare the two. To make things more visually consistent and comparable, the line should ideally be on both panels.

We have included the freezing levels at both panels.

9) Line 267 "we focus on non-attenuated reflectivity". This is a bit confusing as, as far as I could see, the text has not yet clearly stated whether any of the results are based on attenuated or non-attenuated reflectivity. This needs a clearer signposting earlier in the text. I also don't like the term "non-attenuated reflectivity" relating to observations, because what these actually are is "attenuation corrected reflectivity".

We agree with the reviewer and have changed it to attenuation corrected reflectivity. We have also added the following statement at the beginning of the "Results" section to emphasize that the results are generally based on attenuated reflectivity: "The results presented here are based on attenuated reflectivity, unless noted otherwise."

10) Figure 4: the difference in the x-axis ranges makes comparison very hard. Please ensure consistency across all the panels, if possible.

We have revised the figure so that all plots use the same range.

11) Line 267 "do not account for frozen hydrometeors". This is picking up on a point made earlier, that the more important question is whether the retrievals attempt to represent frozen hydrometeors separately, or whether any reflectivity above the freezing level is assumed to be explained by supercooled water droplets (assuming also that melting particles are not separately represented either).

We have revised the Results to better explain the exclusion of certain hydrometeors from the simulations. We believe these habits are not being completely excluded, but are instead implicitly included within the retrieved water-content categories. For instance, if both supercooled water and frozen habits were truly excluded, then in Figure 3 we would not see any simulated reflectivity above the melting layer.

Technical and grammar

1) The second part of the final sentence in the abstract seems to convey no concrete meaning - please rephrase for clarity or remove: “underscoring their potential to improve the assimilation of spaceborne radar data in NWP models”. Exactly what has the potential to improve assimilation is not clear.

The abstract has been revised for clarity. The text now reads as:

> The sensitivity of forward radar simulations to microphysical assumptions, underscores their importance for the assimilation of radar observations in numerical weather forecast models.

2) Line 22 CRTM is a “pivotal collaborative model”. This doesn’t seem to convey much scientific meaning and should be removed or explained more clearly.

Thank you for noting this. We have completely revised the introduction for better consistency.

3) line 28: space radars “provide vertical profiles of cloud and precipitation structures” - this language is too loose. They provide reflectivity profiles giving information on vertical structures of cloud and precipitation.

We have fully revised the introduction for consistency and clarity of language.

4) line 50: “supports an operational environment” does not seem to convey a clear meaning and could be removed or rephrased.

Thank you for carefully reading the manuscript. As noted, we have fully revised the introduction for consistency and clarity of language.

5) equation 1: please use a more self-consistent mathematical notation both here and throughout the paper, and most critically of all, please explain what it is. Here the vector y is bold and the vector x is non-bold italic, for example. The use of bold capitals for both the observation operator H (presumably a nonlinear function) and the background error covariance B (a matrix) is potentially confusing. Typically bold capital H is used for the Jacobian matrix of the nonlinear operator H, which itself is usually written in non-bold italic capitalised or similar.

Thank you for noting the discrepancy. We have revised the text to prevent any confusion in the interpretation of the equations.

6) equation 4 and others: please explain or remove the comma notation (for multiplication?) which does not seem to be very standard.

We have revised all of these equations, as the commas were mistakenly added due to a LaTeX command that did not work correctly.

7) Line 89 V(D) is not strictly the particle volume, in the case of non-spherical particles.

The reviewer is correct, and the text has been updated as follows:

> where $m(D)$ is the particle mass (kg), $\rho(D) = m(D)/V(D)$ is the particle density, and $V(D)$ is the equivalent spherical volume. For convenience, $V(D) = \pi D^3/6$ is used to represent the volume of a sphere with diameter D, even though the actual particles may be nonspherical. This approximation provides a consistent way to relate particle mass and size through an equivalent-volume definition.

8) line 240: "as a result . . . DPR did not capture observations directly over the storm centre" is incorrect. The swath of DPR made it more likely to capture the storm centres (as opposed to nadir viewing sensors like CloudSat). In any case this is a separate message and requires a separate sentence for clarity.

What this really means is that the cyclone's eye appears missing because the storm center was not located in the middle of the scan, and a tilted scan may not clearly capture the eye. In some scans, the storm was entirely absent as well.