General: I think this is a nice study demonstrating the uncertainties not only with model microphysics parameterizations, but also the challenges with polarimetric radar forward operators. I have a few minor comments.

We are very thankful for the reviewer's comments. Below, we address the reviewer's specific comments (in bold blue).

Abstract: A reader approaching this for the first time does not know what Dice or Tgr are. Secondly, it is unclear what a 'low bias' is in the sense of polarimetric moments.

The text in the abstract has been modified for clarity:

Ln 5: "Modifying the critical diameter of particles for ice-to-snow conversion by aggregation (Dice) and the threshold temperature responsible for graupel production by riming (Tgr), was found to improve the synthetic polarimetric moments and simulated hydrometeor population, while keeping the difference in surface precipitation statistically insignificant at model resolvable grid scales. However, the model still exhibited a low bias (lower magnitude than observation)..."

Ln28 – remove extra ')'.

Corrected.

Ln 34: "..total mass peak" what is meant here? The model was not able to capture the correct height of the ice mass or the ice mass was too large?

Here, the "mass peak" refers to the peak of ice mass size distribution. The sentence has been rephrased in the revised manuscript for clarity.

Ln 37: "...model underpredicted total ice number concentrations and overpredicted the peak of mass size distribution..."

Ln 75 – I am confused about what 99.9 m MSL is indicating here. Is that the actual height of the BoXPol radar?

Yes, it is the height of the BoXPol radar with reference to the mean sea level (MSL). The sentence has been rephrased for clarity: Ln XXX: "..., University of Bonn at 99.9 m above MSL".

Ln 125: "it μ depends on" -> "it's μ depends on"

Corrected.

Ln 130: What is 'n'?

Here, "n" is the specific hydrometeor number, which is the ratio of hydrometeor number density to total density (air, vapor, hydrometeors). It is defined as such in the earlier paragraph.

Ln 149: I'm curious about the selection of Dice for the sensitivity studies. The range of values from 5.0 to 400.0 seem like rather large perturbations, and 400.0 seems excessive compared to the original 50.0 . I suppose such a large value essentially limits the production of snow from ice? How were these values selected for the sensitivity tests?

We conducted a sensitivity study using multiple values of Dice (e.g., 5, 50, 150, 400, 800 μm). And, for the upper extent from the default value, Dice = 400 μm showed considerable improvement in the synthetic polarimetric signatures at upper levels, and is used in this study. With Dice=800 μm, it starts to limit the production of snow from ice.

The following paragraph has been added in the revised manuscript to clarify the setup of model sensitivity study:

Ln 335: "For the cloud ice aggregation threshold, we conducted a sensitivity study using multiple values of Dice (e.g., 5, 50, 150, 400, 800 μm), the default value being 50 μm. For brevity, we only report on the results from one lower and one upper value as well as the default value. From these experiments, Dice = 400 μm showed the best improvement in the synthetic polarimetric signatures and is used as the upper Dice value in this study. Similarly, we varied Tgr from the default 0°C by reducing it by 5 and 3 °C respectively, to check the sensitivity of graupel production near the melting layer. The four experiments together constitute different combinations of aggregation (ice/snow partitioning) and riming (graupel production and rain gradient below melting layer)."

Ln 178: Put () around 2 to be consistent.

Added.

Ln 224: What temperatures (are they derived from a sounding, model, etc.) are used for the HCA-Dolan and HCA-Zrnic?

For HCA-Dolan and HCA-Zrnic the COSMO model temperatures were used. These were used for all calculations on the radar side to remain consistent.

Ln 227: "Problem description" seems a little ominous. Perhaps "Case description?"

We have changed it to "Case Description" in the revised manuscript.

Ln 230: Spell out 'approx.' to approximation.

Corrected.

Ln 308: I am confused here about the inter-changed language of mixing ratio and percentages. Are you able to derive an actual mixing ratio (g/kg) for the different hydrometeor types with HCA-Pejcic? Or is this more related to the probability of a given hydrometeor type within a volume?

It is not possible for us to calculate a mixing ratio in g/kg. What we can derive is the hydrometeor partitioning (PR) from the radar measurements. This is a relative mass contribution of specific hydrometeor types in %. The distance between a multidimensional measurement, in our case a vector consisting of horizontal reflectivity, differential reflectivity, specific differential phase, cross correlation coefficient and a temperature indicator, and a hydrometeor specific centroid derived from a clustering procedure is interpreted as PR. More details on the method can be found in Besic et al. 2018 (Unraveling hydrometeor mixtures in polarimetric radar measurements, https://doi.org/10.5194/amt-11-4847-2018) and on our refinement in Pejcic et al. 2021 (Polarimetric radar-based methods for evaluation of hydrometeor mixtures in numerical weather prediction models, https://doi.org/10.23919/IRS51887.2021.9466201).

For clarity, the "mixing ratio" in Section 3.2 has been rephrased to hydrometeor percentage.

Ln 357-360: Unless I'm missing something in the figures, the mean sizes are not shown? That is fine, but perhaps add (not shown) so the reader is not feeling like they are missing something from the figures.

Yes, for brevity, the figures of mean sizes are not shown. The sentence has been modified in the revised manuscript for clarity:

Ln 370: " ...the mean size of graupel around the vicinity of the melting layer is around 1.5 - 2 mm for CTRL and EXP1 (not shown here)."

Ln 424: This reads a little strange with "Since". Perhaps it would make more sense to start with "Even though…"

Reformulated.

Ln 430: It is hard to cross-reference these temperature on Fig. 11. Perhaps add a temperature scale on the right side as in Fig. 9?

To clarify, we added the respective information in the manuscript text:

Ln 444: "The snow-dominated lower levels (-3 to -13°C, approximately corresponding here to heights of 2.0 to 4.0 km) are characterized by a strong underestimation of Z_{DR}.."

Ln 465: If I'm not mistaken, Dice was increased in EXP1,3 not EXP2,3?

Yes, Dice was increased in EXP 1,3. Fixed.

Ln 489: EMA has not been defined.

We spell out the acronym here now:

Ln 503: "Concerning the effective medium approximation (EMA) for the ice-air mixture material of dry snowflakes..."

Figure 3: Perhaps state in the figure caption that the top row is the CTRL and the bottom is EXP3 for clarity.

Added.

Figure 6: I understand the logic of keeping the color scales consistent for each HCA, but it is a bit confusing that the different colors do not align with different species across the three HCAs. For example, it is great that RN and 'snow/Aggregates' are the same colors across all three, but melting snow is purple in a) and cyan in b) and c); light blue is ice crystals in a) and light rain/drizzle in b) and c), and perhaps most confusingly, graupel is orange in a) but ice crystals and vertical ice in c). I also find the purple for WS difficult to distinguish from the blue RN and black HA in a). Finally, in the caption for panel c) vertical ice is listed as VI but in the colorbar it is VC.

We have now adjusted the color bars. Wet Snow is now "cyan" in all three plots, Ice is now also grey in a) as in b) and c). Graupel in a) and Low Density Sleet (LG) in b) are now in the colour 'Fuchsia' like Graupel/Hail (GH) in c). The caption for panel c) has been changed from VI to VC, matching with the colorbar.

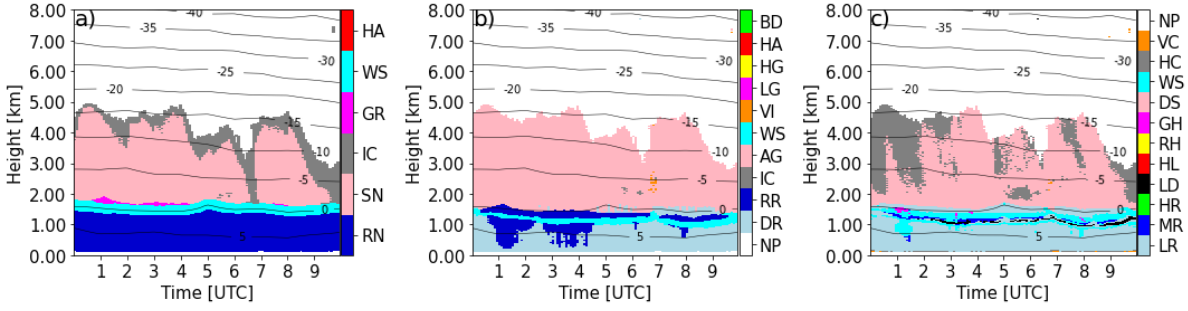

Table 1: are [m s-3] the correct units for vT?

Corrected to [m s^{-1}].

Table 2: The values for EMA are meaningless to me unless I dig up Blahak (2016). Is there a simple way to describe what these mean?

In the table 2 caption, we replaced the puristic reference to Blahak (2016) and added a short explanation of the three different EMA applied in the setup used in the study:

"The detailed meaning of the EMA and a complete overview over all options is given in Blahak (2016). All three settings applied here make use of the Maxwell-Garnett mixing rule (Maxwell Garnett, 1905). 'mas': ice-air mixture with air as matrix and spheroidal inclusions of ice. 'mis': similar, but with ice as matrix and air as inclusions. 'mawsms': a three-component (ice-water-air) mixture constructed as a two-fold two-component mixture, where spheroidal air inclusions are suspended in an ice-water matrix, the latter with spheroidal ice inclusions in a water matrix."