Paul Ockenfuß et al.: "First Nationwide Analysis of Riming Using Vertical Observationsfrom the Operational German C-Band Radar Network"

Overall assessment

The authors use the vertical (zenith) "birdbath" scans of the 17-station German operational polarimetric C-band radar network (operated by Deutscher Wetterdienst, DWD). The birdbath scan is typically used for calibration, but here is used to retrieve microphysical information (riming in mixed-phase clouds).
Taking operational C-band birdbath scans and using them for microphysical retrievals is a novel and elegant reuse of existing infrastructure.
This is a highly relevant topic, and the manuscript is generally well written, in most parts easily understandable, and I highly recommend publication of the paper after addressing the comments below.
Figures are generally clear and well designed. A few adjustments to captions and colormaps could further improve readability and accessibility.

General comments

- The introduction would benefit from a stronger motivation for why detecting riming and understanding its drivers is important for cloud physics, precipitation processes, and numerical weather prediction.
- use either the "German Meteorological Service" or "German Weather Service" consistently throughout the manuscript
- as the other referee notes, your perspective is very Germany-centered and only in the last paragraph do you expand your view to the broader (European) perspective. *Which other national meteorological services operate similar C-band networks? Do any routinely perform birdbath scans?* Clarifying this would help position the work within a broader operational context.
- The original Kneifel & Moisseev (2020) retrieval used an average of MDV over a relatively long (several minutes) window to get rid of turbulence effects and wavy structures. How does this scale with the 15 rays every 5 minutes data from the C-Band radar network? Maybe I missed it in the text, but I think there should be a brief description of the Kneifel & Moisseev (2020) retrieval, how it was improved in Ockenfuß et al. (2025) and how it is finally applied to the C-Band data. You could list (maybe in the table) how many MDV values are averaged for each of the radar data sets, if you use a boxcar filter, and which MDV-FR relation from Kneifel & Moisseev (2020) is used for which of the data sets.

Specific comments

- line 19f: Clarify whether research cloud radars *scan* in the vertical or whether they typically perform vertically pointing observations without scanning. I thnk that "vertically pointing observations" (and not scans) are the standard measurement strategy.
- what should be Eq. (1), the number is missing here for some reason: I think that the subscript "c" for "crystal" is potentially misleading, as "crystals" typically denote pristine individual ice particles. Since the retrieval mainly targets riming of snowflakes, consider using notation referring to "unrimed snow," e.g. *us* (as in Moisseev et al., 2017), or something similar refering to both unrimed aggregates and crystals.
- l. 62: The first figure referenced is Fig. 7a,b, which disrupts the flow. Consider omitting the figure reference here and describing the retrieval conceptually, returning to the examples later.
- Figs. 3 and 4 and 7: use the same colormap for the Doppler velocity. I generally like the use of colorblind friendly color schemes as in Fig. 3. In this context, please also ensure Fig. 1 is accessible to readers with color-vision deficiencies.

- l. 62: Fig. 7a and b : Please provide site and instrument information for the frontal systems shown in the referenced figures in the text to improve understandability.
- l. 71: there is a reference missing.
- l. 74f: “In the vertical, the intrinsic data resolution is 60m, but sampling is performed at 25m resolution (Gergely et al., 2022).” Please clarify how a dataset with intrinsic 60 m resolution is sampled at 25 m, are spectra interpolated or is the FFT performed on overlapping parts of the time series? And does it make sense to look at the data in 25 m resolution then?
- l. 92 : I doubt a bit that it is possible to apply the EDR retrieval by Borque et al. (2016) to the 5min resolution birdbath scan, even if you took relatively long (e.g. 20 minute) windows to apply the retrieval. I think it would still be too few data points to fit turbulence spectrum slopes reliably. Please check this, or remove the reference here.
- l. 93 “equipped”
- l. 102ff: Why are you addressing the points 1-3 in reverse order, i.e. start with the last point and then moving on to the second and finally the first? Consider changing the order and separating the considerations regarding each of the three points with a line break to improve readability.

Mockup data
- 105ff: how are you upsampling the 35 m resolution MDV data to 25 m? Please state which kind of interpolation you are using in the text.
- You are not taking every 10th cloud radar data profile but every 10th cloudnet profile, correct? I understand that cloudnet categorize data are already 30 s averages of the raw MDV: https://github.com/actris-cloudnet/cloudnetpy/blob/main/cloudnetpy/categorize/radar.py#L347
I guess this has implications for the mockup data and all the error considerations downstream (possibly the kappa filter): If the mockup data are 30 second averages, you miss small-scale variability that you would have in unaveraged MIRA-35 data. The goal is to use data similar as possible to the C-Band observations.
<u>I strongly recommend having another close look at the mockup data and making sure that it is really comparable to the C-Band birdbath observations.</u> My guess is that the number of individual samples, coherent and uncoherent averages, going into one time-height observation in the C-Band radar data should be as close as possible to the number of samples in one time-height observation of the mockup data. Are the C-Band data averaged over 15 s more similar to the 30 s cloudnet resolution or to the “raw” MIRA-35 observations? Or does it make sense to aggregate a couple of raw MIRA measurement time steps to get to the resolution of the C-Band 15-s-data? My guess would be that 30 seconds of MIRA observations contain more than your 15*1024 pulses. It might make sense to actually dig into the spectra and (visually) make sure that the data look similar.
- For creating the mockup data, why didn’t you take into account the lower sensitivity of the C-Band radar and the blind range, i.e. artificially worsen the Ka-Band data? I think it wouldn’t be a big deal to exclude pixels below the height-dependent Ze threshold, or, for selected cases such as in Fig. 7, even go into the Doppler spectra and cut them at a higher threshold and then recompute MDV to see what happens when you miss all the liquid water peaks.

Some more specific comments
- l. 110ff: I do not fully understand the meaning of these sentences. Please clarify: Are you saying that you are using the fit coefficients by Kneifel and Moisseev (2020) derived for X-Band on your C-Band data (but probably the ones for Ka-band on the mockup data?) and that X-Band and C-Band are sufficiently close together, the transition size where Mie effects play a role in X-Band would be 1 cm, and this can be neglected except for very large aggregates? You could add the information which FR-MDV relation is used, to Table 1.
- l. 131 “for most of the height range”
- l. 140: missing reference.

Melting layer detection

The Cloudnet melting layer detection uses a combination of a peak in LDR (is this available for your cloudnet data?) and the criterion of MDV above 1 m/s at melting layer base, and, as a fallback if no LDR is available, only profiles of MDV, and, if available, spectrum width. This should be stated somewhere in the text.

Why develop a new algorithm instead of using or extending the Cloudnet MDV/ width one? In general though, it looks like the melting layer detection method proposed here seems to work very well, and this might also be of interest for Cloudnet developers. This could motivate a small comparison study, like in Fig. 3.

The cloudnet melting layer code detection is here:
https://github.com/actris-cloudnet/cloudnetpy/blob/main/cloudnetpy/categorize/melting.py#L128

More specific comments

- l. 157: fallspeeds (lower case) and "criterion" instead of "criteria"
- l. 174: "In Fig. 2c, only winter months (Nov to Apr)… ": this sentence is somewhat out of context here. I would mention this earlier, when discussing Fig. 2 (without having explained the melting layer detection yet but referring the reader to the next section), and maybe also write it in the caption of Fig. 2.
- Side question: Why consider only winter months? You are mentioning it later in the Discussion, that it is to exclude convection, but I think this should be explained much earlier.
- l. 176f: "Only if no reliable detection was possible, we extrapolate the last value up to 1 h and otherwise use the value from the nearest sounding station as a proxy. In the latter case, at most 12 h are tolerated between sounding and radar profile." How far are the nearest sounding stations for the different radars?
- l. 186: replace the "original cloud radar resolution" with the cloudnet categorize/ classification time resolution or something similar
- approx. l. 200 (line numbering is broken there): collocated, not colocated
- Fig. 5a and b are referenced in opposite order in the text, consider switching them
- l. 204: 20,000 instead of 20.000
- l. 212 remove duplicate "percent"
- l. 222ff: Your point 3 is scientifically meaningful and worth further discussion.
- Fig. 6: Maybe I missed it somewhere in the text, but I am missing a definition of "winter" and a discussion why these months were chosen and not differently. Sometimes "winter" refers to the months of DJF only, but it is different in your case. In one place, you write November to April, in another "14 months between October 2021 and April 2023".
- Also Fig. 6: Even though it's kind of obvious, I would also mention why there is a vertical dashed line at 0 degrees (no riming at higher temperatures)
- Fig. 7: This is a matter of taste, but I think it would benefit from adding the site and date as column headers.

- Section 3.3:
- beginning of Section 3.3 (the line numbers are broken again): I would mention the second outlier, Neuhaus, much earlier in the text. You mention that the Feldberg site is influenced by orography. Any ideas why the Neuhaus site sticks out? Interestingly, it does not stick out with respect to temperature distribution of the riming onset in Fig. 13. Any ideas why?
- last 2 lines before the equation: For better readability, I would refer to the letters in the equation in the explanatory text, i.e. "we decompose the total precipitation per site, T, into the product of the number of hours with precipitation, N and the average hourly precipitation rate ^R:

- l. 272ff: Equation numbers are missing again.

- Fig. 10: In 10b, it’s not so easy to see the wet south and dry northeast. Consider changing the colorbar a bit
- l 273ff: The discussion of Figs. 11 and 12 could use a bit more detail. I think these are some of the main results in this paper and I find it a bit hard to follow. In particular Fig. 12b is only discussed very briefly.
- l 290: “who” instead of “which”

Fig. 13 and vertical air motions
- l. 316ff: vertical air motions: The kappa filter removes updrafts, but may leave stationary or slowly varying downdrafts unflagged. Stationary downdrafts could be misinterpreted as riming by the algorithm. I guess that those downdrafts however would not shift the entire temperature distribution in Fig. 13 to lower temperatures for the Feldberg site, as you observe, but might increase riming frequency equally across most of the temperature range (if we assume that the downdrafts influence the entire atmospheric column similarly). In connection with the lower number of observations at higher temperatures due to higher elevation of the Feldberg site, it’s a bit hard to wrap the head around the different effects different factors could have on the temperature distribuion curve.

While for the other (not orography influenced sites) I would expect a probability of 0 at very low temperatures, say below -20°C, at the Feldberg site, there could be also riming at temperatures below that range, maybe even below -40°C, due to misclassifications, e.g. if the entire atmospheric column is in a downdraft for several minutes. It is somewhat eye-catching that fbg is the only curve extending down to the x-axis lower limit, and maybe even beyond.

How does the curve look like for lower temperatures below the x-axis lower limit in Fig. 13? This can give a hint on the uncertainty of the method introduced by orography.
- *If you are not normalizing to an area of 1 below the curve but rather look at the likelihood of riming per temperature bin as Kneifel & Moisseev (2020) did in their Fig. 8, how does fbg then compare to the other sites?*
- l. 331ff: the idea that stronger updrafts might be responsible for the shift in the temperature distribution towards colder temperatures at first sounds plausible, but aren’t these the cases that the kappa filter flags as convective/ updraft?

More general comments:
- did you look into whether there is a correlation of FR and precipitation intensity?
- Conclusions: One point from earlier in the manuscript could be brought up again here: Saving high-resolution model profiles of T, p, q and wind for all the weather radar sites should not be a big deal as they are also being saved for Cloudnet sites, correct?
- In general, the Conclusions are beautifully written, and motivating.