We thank the reviewer for the valuable feedback on the revised manuscript. In the new revision, we have incorporated the suggestions from the reviewer and performed a language review.

Please find our reply in blue to the individual issues raised by the reviewer. Note that the line numbering refers to the lines in the new revision.

Review of the revised version of the following manuscript

Title: Combining commercial microwave links and weather radar for classification of dry snow and rainfall
Author(s): Erlend Øydvin et al.
MS No.: egusphere-2024-2625
MS type: Research article

I would like to thank the authors who did a considerable extra effort to reply to the reviewers comments. The new data classification strategy is straightforward and results are presented now in a more clear form.

However, I think some (small) work is still needed.

Major point 1 has been properly addressed.

Major point 2 has been properly addressed after the authors moved to the hourly integration time scale.

Major point 3: I think it has been addressed. I have only an editorial comment on Fig. 7 (same for Fig. 2): I think it would better to use a different colormap on Fig. 7b and Fig. 7c, because it's not easy to distinguish between small and high fractions. For instance, using 5 distinguishable tones of red for snow and of blue for rain (0-20, 20-40,40-60, 60-80 and 80-100) would help in better understanding your statement on lines 259-260: "Additionally, for events where the CML estimates more rainfall 260 (negative bias), there are more rainy hours as observed by the disdrometer."
Good idea. In the revised manuscript we have changed the colormap in Fig. 2 and Fig. 7 to use a colormap that divides the colors into distinguishable tones.

Major point 4: I do not completely agree with the authors' reply on 4a. It makes sense to feed the RT method with ERA5 re-analysis data and indeed I didn't ask the authors to feed the RT method with ground data. I think it would be useful to check ERA5 against some available ground data. The authors discuss in depth about possible errors of disdrometer but do not talk about possible errors in Td estimates.
We agree that we should discuss the uncertainties of the temperature data. The temperature data is a downscaled version of ERA5 temperature data that is combined with ground observations of temperature. This information is included in the sources, but it was not made explicitly clear in the previous version of the manuscript. Lussana et al., (2019) perform an extensive analysis of this dataset with some case studies. To better highlight the uncertainties of the temperature data we suggest the following changes to the manuscript:

Add the following to the methods section (L147):

- "Temperature and humidity data were downloaded from THREDDS (2024). The temperature data is a downscaled version of ERA5 data that is combined with ground observations on a 1 km grid with a temporal resolution of 1 hour (MET, 2024; Lussana et al., 2021, 2019). Lussana et al., (2019) provide an extensive analysis of this data, and the uncertainty of the data depends on several factors like distance to the closest observation station, terrain complexity and model assumptions.

Add the following to the discussion (L375):

- "Another source of uncertainty lies in the temperature data used for the RT method. The temperature data is a downscaled version of ERA5 data that is combined with ground observations. Lussana et al., (2019) found that the expected RMSE of the temperature data ranged between 1-2 °C in observation dense areas and 2-2.5 °C in observation sparse regions. The RT method performance could thus be less good in areas with complex terrain and sparse ground observations."

Comments on the new version:

• Fig. 6 c) deserves an explanation. Why CR method detects a lot of snow events with Td well above 0°C (there is a peak around 8-9° C!). This is an important fact spotted on lines 250-251 and quantified in Figure 9f). However, there is only a short comment on lines 324-326 later on in the discussion. Is this really an algorithmic problem, as the authors suggest (hence, it is anything that can be mitigated) or, rather, is it due to an inner limitation of CML sensors, i.e. a limited sensitivity due to length-frequency characteristics + signal quantization? Unfortunately, there is not an analysis divided by rainfall intensity class, which I think would help in understanding. At least, please add this possible explanation related to a limited sensitivity to the CML.

Good point. There are several things that could affect the CML. We suggest adding the following to L322:

- "The large number of false snow events estimated by the CR method, also observable above 2 °C (Fig. 6), might be due to several factors. Low intensity rainfall events could fail in triggering the CML rainfall detection algorithm, for instance due to the quantization of the CML signal. Further, due to the spatial difference between the radar beam and the CMLs, the precipitation might hit the radar, but miss the CML, triggering the CR method to estimate snow. Finally, hardware issues with the CML, or database errors, could result in a flat signal level, causing the CR method to misinterpret conditions and estimate snow. Better quality control of the CMLs, for instance by checking their correlation against the weather radar during rainfall events could improve the CR estimates."

• Fig. 2 caption: what do the author mean by "(distorted) positions of the CMLs" ?

We suggest rephrasing the line to read:

- "White circles indicate the distorted positions (i.e. randomly shifted position to prevent exact retrieval of coordinates) of the CMLs."

• Lines 230-236 the description of what's happening around 18:00 does not fit Fig. 3 ("Around 18:00, the CR method, RT method, and disdrometer all estimate rainfall")

It's true that the RT method also estimates some snow. For making the text simple to read we suggest deleting this line, as the rest of the text sufficiently well describes the behaviour of CML1 and CML 2.

Conclusions/1: Just counting the misclassified samples on Fig. 9 (i.e. non-diagonal elements of the matrices), for Td>=2°C I see roughly 35k misclassified hours by RT and 39k by CR, while if Td<=-2°C we have 30k against 32k respectively and if |Td|<=2°C we have 22k against 21k. Either method has its own strengths and weaknesses, which brings to the next bullet.
We agree that each method has its own strengths and weaknesses. See also the next comment.

• Conclusions/2: I would like that the authors comment about whether merging RT + CR methods would be beneficial. Maybe yes, because, for instance, Td info would dump to zero all that huge CR misclassification of dry/rain as snow at large positive Td values, which is the minus of CR.
Agree, we suggest to add the following to the discussion (L341):

- "For classifying snowfall and rainfall, both methods have their own strengths and weaknesses. The CR method shows a better ability to classify precipitation in the interval -2 to 2 °C (MCC = 0.39), but falsely estimates a large number of snowfall events above 2 °C. The RT method, on the other hand, provides reliable precipitation classification below -2 degrees (MCC = 0.30) and above 2 °C (MCC = 0.34), while its performance is not as good within the interval -2 to 2 degrees (MCC = 0.28). Consequently, combining the RT and CR method would be optimal. This could, for instance, be done by using the CR method in the interval -2 to 2 °C and the RT method below -2 °C. Above 2 °C, precipitation could be classified as rainfall if either the RT or CR method detects rainfall."