The authors would first of all like to thank the reviewers for their thoughtful and constructive review. We note that the goal of our manuscript was to evaluate the performance of a novel method for classifying precipitation types, that works by combining estimates from weather radar and commercial microwave links (CR method), and compare it to ground truth disdrometer estimates located along roads in Norway. This method was compared to estimates from weather radar and a downscaled ERA5 temperature model (RT method) against the same disdrometers. Since the different data sources have different temporal resolutions and spatial representations, they must be aggregated in order to be comparable. Our main take away from the review was that this aggregation should be performed in a clearer and more understandable manner, and that parameters used in the CR method (length of wet period) and RT method (temperature threshold) should be better justified.

In the original manuscript the observation methods (CR method, RT method and disdrometer ground truth) were compared using 1 minute resolution and the observation methods estimates were resampled so that they could be compared at that scale. While we think that this is a valid approach, we acknowledge that it would be easier for the reader to understand the aggregation method if the observation methods were compared at 1 hour resolution, as suggested by one of the reviewers. One aspect that arises from aggregating the disdrometer data to hourly resolution is how to classify hours where there have been observed a mixture of rain and snow. An intuitive approach would be to classify these hours as mixed precipitation. However, none of the observation methods used in this experiment directly estimates true mixed precipitation, such as wet snow, and it is therefore not known how these true mixed events show up across the different observation methods. Introducing a mixed class thus requires tuning of each of the observation methods. We have experimented with this mixed class, but found that since it lacks a true physical reference, it is prone to overfitting as it would require extensive tuning to the specific dataset used in this study.

We therefore suggest aggregating the observation methods to hourly resolution using a simpler approach where we classify hours with any rainfall as rainy, hours with no rainfall but with snow as snowy, and hours without any precipitation as no precipitation. We have experimented with this approach, and find that for the overall evaluation of the CR and RT method, the results and conclusions are approximately the same as in the original manuscript.

As a result of the simplified aggregation method, and to other comments below, we propose to following major changes to the figures:

- **Figure 1:** We suggest showing the number of wet hours as observed by the observation methods.
- **Figure 2:** We suggest including, in addition to MCC, accuracy, precision, recall and F1 score.
- **Figure 3:** This figure shows the MCC score as a function of temperature for the RT and CR methods. As one of the reviewers highlighted, this figure does not add much more than what is already presented in Figure 2. We suggest replacing this figure with a confusion matrix, so that readers can better understand how the CR and RT classifications differ (also suggested by the same reviewer). Additionally, we propose adding a table with corresponding accuracy, precision, recall, F1, and MCC scores.

We have experimented with this and find that the general results and conclusions remain the same.

- **Figure 4:** This figure shows how the CML overestimates precipitation amounts around zero degrees Celsius compared to the radar reference. We suggest removing the left column, as we mainly discuss the right column. We also suggest removing the panels showing mixed precipitation, as the simplified aggregation method, suggested above, does not include mixed precipitation.
- **Figure 5:** This figure shows a map of the CMLs and the estimated precipitation type, including mixed precipitation. One of the reviewers suggested including more information about the topography in the map. Since we suggest not including mixed precipitation directly in our analysis, we propose changing this figure to show the interpolated number of rainy and snowy hours estimated by the CR method, RT method, and disdrometer ground truth over a two-day period. Further, by zooming out and distorting the CML coordinates, we can plot the CML locations as well as the Norwegian coastline without revealing the exact CML positions, showing larger trends and allowing for a better understanding of the climatic differences. We have experimented with this and find that the patterns and differences between the RT and CR methods become more pronounced.
- **Figure 6:** This figure shows a time series of the CR and RT method. Both of the reviewers struggled to see the benefit of adding road surface conditions and mixed precipitation, but they found the figure very useful for understanding how the CR and RT method actually works. We therefore suggest simplifying this figure by removing the road surface conditions, and move the figure to the beginning of the results section so that the readers are exposed to this figure first. Further, to make the figure more relevant we suggest to instead plot the timeline of one or more of the CMLs shown in the map (previously Figure 5). We have experimented with this, and find this approach better connects the figures, making it more engaging for the reader.

Please find our reply in blue to the individual issues raised by the reviewer.

Review 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

This manuscript proposes simple methods to classify the precipitation type (i.e. rain/snow/dry) based on opportunistic data collected by wireless microwave links in combination with radar data (what the authors called “CR method”), relying on the different sensitivity of microwave links and weather radars to snow.

Gathering valuable information about snow from CML data has not been really addressed so far. Hence, the topic of this manuscript is interesting. Using joint radar and CML data for meteorological purposes is interesting as well. So, I think there is enough novelty and interest in this contribution. However, I see a number of points that should be addressed by the authors before publication. Therefore, I recommend to accept the manuscript after a major revision.

Major points

1. Datasets. The authors put together data from 2179 CMLs across Norway, later reduced to 550 and 435 for the winter and summer datasets (Sec. 2.3). However, they didn't provide much information about the characteristics of these links (they just mentioned lower and upper bounds on lines 135-136 of their manuscript). In particular, it would be good to provide the frequency vs length distribution and the quantization error, which determine the minimum detectable rainfall intensity as well the accuracy of CML measurements. I know that CML metadata have some issues due as the owners are private companies as Ericsson, but I saw these data published in several papers on this subject. Hence, we do not know which is the sensitivity of these links to light precipitation rates, which are typical of the winter period where also snow is present in cold regions. This would be also important to understand to what extent the scatter in the data highlighted for instance in Figs. 2-and 4 is due to some/several/many CMLs performing worse than others as their sensitivity is lower, or whether it is due to the classification method itself.

We agree that the characteristics of the CMLs should be shown, and suggest to include a plot in the methods section showing this information. The quantization of the CML signal level is the same for all CMLs (0.3 dBm), except for one. We suggest removing this CML so that all CMLs have comparable quantization.

2. Methods. In particular, the way labels (i.e. precipitation type) were assigned. If I got it correctly, CML data are sampled every 1-min, radar data are available every 5-min, disdrometer data are provided every 10-min and meteorological data come every 1 hour. Due to the different sampling rate of the sensors involved, it is necessary to define a suitable time window within which labels are assigned.
    a. I would like that the authors clearly state how they put together the 1-min CML slots over the integration window (arguably 1-hour) to decide whether an hour is rain or anything else. The same should be done for the other sensors. I think these concepts are written in Sec. 3.4, but it's not the right place in my view. This is about methods. Moreover, the author should provide some justification for all those threshold values they used. Another important detail on methods is somewhat hidden as it is delayed to lines 290-292.

We agree that the way the aggregation of the different observations is described is hard to follow. Moreover, as the results can be somewhat tuned based on how the aggregation is done we suggest simplifying the aggregation method so that we classify hours with any rainfall as rainy, hours with no rainfall but with snow as snowy, and hours without any precipitation as no precipitation. We suggest to describe this aggregation in the methods section as follows:

- "Since the weather radar, temperature model, CML, and disdrometer operate on different locations and at different time resolutions, their estimates might not be synchronized, potentially leading to erroneous comparisons. Aggregating the disdrometer, CR, and RT estimates to hourly intervals can help smooth out these differences. However, some hours may contain both snow and rain, complicating the aggregation process. One solution is to introduce a mixed class, classifying hours with both snow and rain as mixed precipitation. However, it remains unclear whether true mixed precipitation, like wet snow, would consistently cause the disdrometer, CR, and RT methods to alternate between detecting rain and snow, which could make the mixed class less physically meaningful and lead to inconsistent representations across different estimation methods. For instance, the RT method uses temperature data with hourly resolution, resulting in each hour being classified as either snowy or rainy. Mixed precipitation could be introduced by setting precipitation within a fixed temperature interval as mixed, yet the actual precipitation within this temperature interval could still be purely rain or purely snow, leading to inaccurate classifications. Additionally, since the radar might estimate precipitation slightly before the CML, the CR method is prone to estimate snow before rainfall events, leading to an overestimation of mixed precipitation. This could be addressed by aggregating the CML wet period so that the radar precipitation estimates fall inside the wet period, but this would require further tuning to avoid estimating too many rainfall events at the expense of fewer true mixed events. Another way to aggregate the estimates to hourly resolution is to drop the mixed class and classify each hour based on the most frequently estimated precipitation type. However, this approach has its own issues. True mixed precipitation, like wet snow, may not consistently show up as a mix of rain and snow in the observation methods, leading to inaccurate classifications. Additionally, this method would require fine-tuning of the disdrometers, CR, and RT methods, which adds complexity and uncertainty. Thus, to simplify the aggregation approach and ensure more consistent and accurate classifications, we have chosen to classify hours with any rainfall as rainy, hours with no rainfall but with snow as snowy, and hours without any precipitation as no precipitation for the disdrometers, CR method, and RT method. Note that switching the role of rain and snow would lower the performance of the CR method, as it often predicts snow due to the CML and radar not being synchronized."

Lines 290-292 mention that the RT method is evaluated at the CML midpoint, we agree that this information should be clearly stated in the methods section.

b. Related to previous point a): it is not easy (at least to me) to understand how the authors assigned the precipitation type labels to get the results in Fig. 2 just based on what the authors wrote in previous Sec. 2.

We hope that the suggested simplified aggregation can better help the reader understand what is going on.

c. Not clear which is the sampling rate of disdrometer data. On line 128 it is written that "The disdrometers […] provide an estimate of the precipitation type every 10 minutes". However in Fig.1 it seems that data are every 1-min. Please clarify.

The way this was done was to just interpolate the CML estimate, so that if the disdrometer estimated 10 minutes of snow, then snow would be assigned to every one of these minutes. However, we suggest simplifying the aggregation process, see comment above so that this will be less of an issue.

3. Results: I think the results in Sec. 3 are not presented in the best way. Reading the abstract, the purpose of this paper is clear: "This study introduces a new approach to improve rainfall and dry snow classification by combining weather radar precipitation

detection with CML signal attenuation […]. Both methods were evaluated using ground measurements from disdrometers.”. Hence, I expect a basic performance assessment of the RT and CR classification methods proposed by the authors against the ground truth. However, looking at Figs. 1-6, well it’s not very clear to me.

a. I agree MCC is a comprehensive performance indicator but it’s not easy to understand how MCC values in Fig. 2 turn into good or bad labelling of data. I think a simple contingency table with indexes as Specificity and Recall for either method would help.

Good point. In addition to MCC we suggest adding accuracy, precision, recall, F1 scores as well as a confusion matrix.

b. Fig. 3 puts together hundreds of CMLs (which I guess have different performance as rainfall sensors). The trends in the figure can be identified looking at Fig. 2. I don’t think it brings a lot of extra information. Indeed there are just a couple of short statements in the text that comment this figure.

We agree to remove this figure and replace it with a confusion matrix.

c. Fig. 4: I didn’t get well what is shown in rows 2-4 of this figure from the explanation in the text. On lines 205-207 it is written that “In the second, third and fourth row, we have plotted the fraction of hours within the bins where the disdrometer recorded at least 10 minutes of rain, snow and both snow and rain (mix) respectively.” The term “bins” is maybe inappropriate. I would use it for an histogram. Let me see if I got it. First, the counts in the top row are now hours rather than minutes. Right? Then in rows 2-4, these counts have different colours according to the fraction of time hours were flagged as rain/snow/mix by the disdrometer. So, if an hexagon in the first row has a color corresponding to 10 counts, that is 10 hours, the same point in the second row is colored according to the fraction of this 10 hours flagged as rain by the disdrometer? I don’t know if I got it. Whatever is the case, please explain it as I tried to do, as it is hard to understand it from the short explanation in the text. Moreover, the colorbar is not the best. I would have used a blue scale for fractions < 0.5 and a red scale for values > 0. Having stated this, the results in Fig. 4, at least what one can notice first sight, is expected in my view: a lot of rainy time above 0°C and a lot of snowy time below 0°C and finally an uncertainty region around 0°C. Moreover, we cannot say whether a 2.5 mm/h difference in CML vs radar accumulation is large or not if we do not know the exact magnitude of rain accumulation. Looking at some finer features: it is a bit strange that radar overestimates so much in several cases at large positive Td values. Is it maybe that those were high precipitation summer events? Hence, the fractional difference Radar-CML is much less.

We acknowledge that it is hard to interpret this figure. However, by changing the aggregation method to the simpler one suggested above, we think it would be easier to understand it. We therefore suggest to simplify the figure by:

- removing the mixed class, as we disregard the mixed class in the simplified aggregation.
- remove column 1, as wet/dry classification by the radar is not important for understanding the RT and CR method.

d. Fig. 5: I see really little information here. What's the purpose of showing this figure? On the other hand, I think Fig. 6 is very useful as it shows the whole story as seen by the different sensors. I would have moved this one forward as Fig. 2 because it is very easy to understand and helps the reader in interpreting better the scatterplots. Good job here!

We suggest changing the map by zooming to a larger area and counting the number of rainy and snowy hours over a 2 day period. Focusing on a larger area will let us plot the CML distorted locations while also showing large scale trends in the distribution of snow and rainfall.

4. Discussion. More than issues, here I am just pointing some other possible explanations of the results.
    a. Sec. 4.1. The authors argue that disdrometer could fail as ground truth in some cases. In addition, should we 100% trust estimates of Td based on ERA5? These are not ground data measured by weather stations. Maybe it would be good to check ERA5 RH and Ta outcomes against some weather stations that are for sure available (for instance as you did in the third row of Fig. 6). Indeed, looking at Eqn. (1), Td can range between -2 and 0°C with Ta=0°C and RH ranging from 90 and 100%. It means that a 10% error on RH measurement/estimate turns into a 2°C error in Td estimate.

We agree that using data from ground based sensors would improve the RT method estimates. However, the intention of using the downscaled ERA5 estimates was that these are available everywhere, and thus provide the most realistic competitor to the CR estimates. Thus we do not think a comparison of ERA5 data to sensors on the ground is necessary. Instead we suggest adding the following to the methods section:

- "It should be noted that model data carries inherent uncertainties due to factors like model assumptions and the downscaling process."

We also suggest adding a brief discussion on this when discussing the suggested confusion matrix (see summarizing comment) in the discussion section.

    b. Sec. 4.2 (but also Sec. 4.1). The only way to check the effect of spatial distance between CMLs and disdrometers is to select a subset of CMLs with a disdrometer in the neighbourhood and calculate the MCC:

We have selected disdrometers within 8 km. This was a tradeoff between having enough data and not having disdrometers too far away from the CML. Since we evaluate the RT and CR estimates at similar distances from the disdrometer, the methods should have the same bias. We have experimented with plotting the MCC for different CML-disdrometer distances, and we find that the MCC only decay slightly with distance. Moreover, the CR method really only provides an improvement at temperatures around 0 degrees, making temperature a more interesting variable. We suggest not to show a figure of the metrics for different distances, as it does not directly contribute to better understanding the difference between the CR and RT method.

    c. About wet snow. Looking at Fig. 4 wet snow seems to correlate with Td (as expected). I expect wet snow to occur at small negative values of Td, while dry snow to occur at colder temperatures. Below -5°C it's mostly dry snow according to Fig. 4. Can the author refine a little bit their decision algorithm including a Td threshold to discriminate between wet and dry snow? Could it be that some unexpected results are due to the way data with different sampling rates were combined together in the hourly time windows and the threshold values used to flag rain/snow /mix/dry intervals)

That would indeed be interesting. Unfortunately the disdrometers do not provide estimates about dry and wet snow, making such an experiment difficult to do. In the above comments we have suggested simplifying the classification method so that the assumptions made are more explicit. This would still allow for a discussion about the effect of wet snow, but more related to the performance of the CR and RT method.

Minor comments:

Line 44: “Human observations can be subjective and aren’t suitable for continuous high frequency monitoring” The term “high frequency” is a bit ambiguous in this context. I guess you mean high-rate monitoring.

We suggest rephrasing the line to read

- “Human observations can be subjective and aren’t suitable for continuous high-rate monitoring.”

Lines 174-175: I cannot understand the statement. “Our dataset consists of CML-disdrometer pairs from the summer dataset and the winter dataset. Every minute each pair provides several different observations such as disdrometer observed precipitation type, dew point temperature and CML signal loss.”. Isn’t the dew point temperature taken from the ERA5 dataset? ”Our dataset” in my eyes are the data “we produced ourselves”, but it’s not the case. Why radar maps are not part of “our dataset”?

We suggest simplifying the aggregation process, see major points 2 and 3. The radar maps are converted to rainfall rates along each CML.

Line 174: you state that disdrometer data are taken every 1-min while previously (line 128) you stated that precipitation type classification from disdrometers is available every 10-min. Please clarify.

We suggest to simplify the aggregation process so that it is easier for the reader to comprehend, see major points 2 and 3.

Lines 179-182, comment on second row of the figure. I would say “snow mostly below 0°C” while “rainfall is mainly above 0°C”.

Yes. We suggest changing this Figure to include the RT and CR estimates instead.

Line 186: “using monthly time series”. To me “monthly time series” means that you have used a time series long one month and extract information from the time series as a whole. Please explain-

We suggest deleting this as it does not convey any important information.

Line 187: pair instead of pairs

We agree.

Figure 2: the gray cluster is hardly visible in the two bottom panels. It would be better to use a darker tone of gray.

We suggest changing the color map to one of matplotlibs sequential colormaps

Lines 186-192 (Comments of Figure 2): I see MCC_rain of CR is usually better when it's above 0.4, maybe worth to state it-.
Yes, we further suggest adding a confusion matrix with several metrics so that these findings are more explicit.

Lines 193-196(Comments of Figure 2): the first thing I noticed looking at this figure is that CML really enhances MCC_rain wrt radar (as expected).
That is true. We suggest adding this to the discussion:

- "In terms of rainfall classification, the CR method performs just as well or outperforms the RT method for all temperatures above -2 degrees (Table 3). This could be due to the fact that the CMLs are located on the ground, which situates them closer to the disdrometers compared to the radar beam, or due to the radar beam being blocked by mountains."

Figure 5, caption: I guess it is "mix(red)" instead of "dry(red)".
We suggest simplifying the aggregation process so that mixed precipitation is not directly a part of our results. See comment above.

Line 264: 10°C instead
Ok

Line 453: DOI is not correct
Ok