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.

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.
    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.
    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.

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.
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.
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.
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!

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.
   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:
   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?
   d) 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)

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.

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”?

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.

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”.

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-

Line 187: pair instead of pairs

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

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-.

Lines 193-196(Comments of Figure 2): the first thing I noticed looking at this figure is that CML really enhance MCC_rain wrt radar (as expected).

Figure 5, caption: I guess it is “mix(red)“ instead of “dry(red)”.

Line 264: 10°C instead

Line 453: DOI is not correct