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: “Combining commercial microwave links and weather radar for classification of dry snow and rainfall” by Erlend Øydvin et al. (2024)

This paper presents an interesting novel method to classify dry snow and rainfall by merging two precipitation measurement techniques. Previous studies has shown the difficulty to estimate precipitation phase, yet is a relatively important process in the hydrological cycle. The authors perform an elaborate analysis of their novel method and compare this with two more established methods. In general, this is a well-written manuscript. However, some parts would benefit from a revision, in order to improve the overall quality of the manuscript. Below I have stated more general and specific comments, which I hope the authors consider to be constructive.

**General:**

- Results:
  Parts of the result section are hard to understand due to the combination of complex figures in combination with it being unclear to which subplot is being referred. It would already be helpful if all subplots are labelled with a,b,c… (which might be a requirement by AMT, check their website) in combination with adding in-text

references to the specific subplots. Moreover, for the reader, especially those who are less familiar with CMLs, it would be perhaps worth considering to add a basic time-series (like Fig. 6) to the start of the results section, in order to show the reader how the methods workout in reality and help get an idea about the differences between the CR and RT method. In fact, such a figure already illustrates a lot and prepares the reader for the following more in-depth analysis. Additionally, for some figures it is not clear how you used the data to create the figure (e.g., Fig. 2, see specific comments).

We agree that the results part is too complex as it is now. We suggest simplifying the aggregation method so that the same method is used for all figures and removing some of the figures (see summarizing comment on the top). We also suggest labeling all subplots.

- Road surface conditions:
  I am struggling to see the added value of Sect. 3.5 in which the road friction is related to the classification methods (except the discussion of Fig. 6 which helps understanding the methods). I understand that the new method could also be used to estimate road conditions and I do agree that it is a nice illustration of the potential of these methods, but I think discussing road friction would require a more elaborate analysis. As you write in the discussion, road conditions are not simply a result of dry snow or rainfall, but also previous conditions (e.g., was there already a snow pack, has there been any snow/ice removal, road temperature preceding rain event) while also other precipitation phases should be considered (e.g., what about freezing rain?). Also until Sect 2.6, there hasn't been any reference of the road conditions, so it came as a bit of a surprise. Based on the introduction,I would have expected that your new method would have been applied to something like a hydrological model.

We agree that including the effect of wet snow on road surface conditions is a bit speculative without providing a larger analysis, which is tricky due to the already mentioned complexities of freezing roads. We therefore suggest removing Fig. 6 and adding simpler time series plots, that do not include road friction and road conditions, as suggested by the reviewer above.

- Discussion:
  I like that the discussion is relatively concise, but it does not include many references to previous studies on similar topics. I encourage you to include some references to previous studies, so that the reader would be better able to place your results into a wider perspective. For example, how do your findings compare with other studies that use similar methods (mostly for the RT method)? See also the specific comments.

While it would be very good to add this information, we think there are some challenges with comparing the RT estimates to other studies. We suggest to add the following text to the discussion to clarify this:

- "While other studies such as Gjertsen and Ødegaard (2005), Casellas et al. (2021), and Saltikoff et al. (2015) have evaluated the performance of temperature-based precipitation phase classification methods, these studies typically vary in methodology, terrain complexity, radar technology, and the instruments used to estimate the ground truth. Furthermore, although the disdrometers used in this study provide a large dataset, their inherent uncertainties, particularly the large number of rainy events recorded below zero degrees and the lack of a mixed class, introduce errors that are specific to this study. These factors make it challenging to directly compare the RT estimates in this study to those from similar studies."

**Specific comments:**

L9: perhaps include the location of your study.
We suggest rephrasing the line to read:

- "Both methods were evaluated using ground measurements from disdrometers within 8 km of a CML, analyzing data from Norway using 550 CMLs in December 2021 and 435 CMLs in June 2022."

L12: There is no mention of the application to road conditions.
We suggest removing the application to road conditions as it is not important in answering our main research question related to the RT and CR method. See other comments.

L54-58: You introduce the dewpoint temperature here, but would it be an idea to include the explanation as to why both are important (L19-33) here? This will help the reader relate the dewpoint temperature to the underlying processes. Or perhaps even referring back to the previously mentioned importance of profiles would already help the reader.
We suggest changing the lines 54-56 to read:

- "Although some studies do not observe any benefit of including humidity (Leroux et al., 2023), humidity is thought to be an important parameter since the atmospheric moisture level affects the melting and evaporating precipitation, influencing whether precipitation reaches the ground as solid or liquid (Kuhn, 1987). One common measure combining humidity and temperature is the dew point temperature, which is the temperature at which air becomes saturated with water vapor at the current water content (Lawrence, 2005). The dew point temperature works by indicating the temperature at which condensation begins, reflecting the moisture content in the atmosphere and thus affecting the rate of evaporation and melting of precipitation, which influences whether it reaches the ground as solid or liquid. This measure can provide important insights into the atmospheric conditions and aids in classifying precipitation types (Feiccabrino, 2020; Harder and Pomeroy, 2013, 2014)."

L73-85: I suggest to include some more studies where they show the influence of various precipitation types on the CML signal intensity in more detail. See for example: Hansryd, J., Li, Y., Chen, J., & Ligander, P. (2010). Long term path atenuation measurement of the 71–76 GHz band in a 70/80 GHz microwave link. Proceedings of the Fourth European Conference on Antennas and Propagation van Leth, T. C., Overeem, A., Leijnse, H., & Uijlenhoet, R. (2018). A measurement campaign to assess sources of error in microwave link rainfall estimation. Atmospheric Measurement Techniques, 11(8), 4645-4669. htps://doi.org/10.5194/amt-11-4645-2018
Good suggestion, we propose adding the following to L80:

- " Only a few studies have focused on how CMLs are affected by colder climates. Hansryd et al. (2010) reported that dry snow caused minimal signal attenuation, while wet snow caused higher signal attenuation. van Leth et al. (2018) observed that during an event with a mixture of rain and snow, the CMLs experiences a strong signal attenuation which persisted for about 10 minutes after the snowfall even"

L79-80: Based on this line, I would have expected the paper to focus on applying/elaborating on the methods of Cherkassky and Ostrometzky. I suggest to rephrase this.
We suggest removing this line and merge the following paragraph with the paragraph above.

L83: reports  reported, in order to use the same tense as the previous references (very minor comment)
We agree.

L90: I would suggest to already mention here that you cannot share the location of the CMLs due to data security reasons.
We suggest adding the following to L90:

- "According to the agreement with Ericsson, the exact location of the CMLs are secret and cannot be shared."

L103-104 (and L118-120): I struggle to understand why you do this. As I understand you extend the wet periods in order to make sure that the rain event has also completely passed the disdrometer, is that correct? Yet in your analysis, there is no comparison of rain events but you compare individual time steps and average/summed periods, so I don't see how this period extension exactly works. Or am I missing something here?
This is a valid remark, in short the idea is that by extending the wet periods, there is a higher chance that the predicted wet period will coincide with the true wet period observed by the disdrometer. However, following the review we suggest simplifying this aggregation by instead setting 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. See major comment 2a by reviewer 2 and summarizing comment on the top.

L107: Could you explain briefly on what concept the method of Leijnse et al. (2008) is based? Not every reader will know how their method works.
Since we suggest not doing any CML precipitation estimates in the revised article, we suggest removing this sentence.

L110-114: Could you provide a reference to this radar product for readers who would be interested to know more about this product?
There are unfortunately not any official references to the radar product.

L113-115: Has the seacluter and other large peaks been removed by you or was that already done in the data you downloaded?
It was already removed.

L121-136: How well do these disdrometers work? Are there any known biases or uncertainties? Especially because in Sect. 4.2 you refer to their potential uncertainty. This could also be discussed in the discussion.
There is, as far as we are aware, no official experiment indicating how well the disdrometers employed by the road authorities work. As far as we know, this is the first official work investigating the disdrometer performance.

L132: Why do you use 8km? Is there a specific reason for this? Is this based on previous studies, or based on your own data? Or is this a common value in CML studies?

This is just picking a number that gives a reasonably large number of CML-disdrometer pairs, without going too far from the CMLs.

L138: Are there any common biases or uncertainties in this data? If so, please mention to help the reader.
As this is downscaled ERA5 model data there are indeed uncertainties in this dataset. We suggest adding the following sentence to L138:

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

L147: Marks et al. use 0 C, but that is for the USA right? Why would that be applicable to Norway? As you write in your introduction, these threshold values can have a relatively large range, and thus can have a relatively large influence on your results.
You are correct that the threshold would vary based on different locations. It can also change during the precipitation events. Thus, while the RT algorithm could be tuned to match the observed precipitation type, we think that doing this would provide an unfair advantage of the RT method over the CR method. Moreover, the purpose of the RT method is mainly to provide a reliable reference method so that the errors of the CR method can be better understood. We have therefore used a threshold from the literature.

L155: Perhaps show the equation for MCC. I think not all readers will be familiar with it.
We agree, and following the review, we suggest adding the full confusion matrix as well as several other metrics. See summarizing comment on the top.

Fig 2.: The caption is hard to understand here, because you refer to RT and the disdrometers both as references. Also, it is not clear what you mean with "the MCC's are computed for each CML-disdrometer pairs using 1 month of data". What do you mean with 1 month of data? Just the December data? It almost seems like there is an additional step between the data described in Sect. 2 and this figure. (also minor comment: I think it should be pair instead of pairs)
We suggest rephrasing the Figure caption to read

- "Scatter density plots **(a, b, c, d, e)** comparing the accuracy, precision, recall, F1 and MCC score for the CR and RT method for each CML-disdrometer pair. Average temperature of each cell **(f, g, h, i, j)"**

Fig. 3: Is this the average of the MCC's for all CML-disdrometer pairs as a function of Td?
Yes, however, we suggest replacing this figure by a confusion matrix to address the review, see summarizing comment.

L198-210: Here the addition of in-text references to subplots a,b,etc. would help to guide the reader.
We suggest removing this Figure as, following reviewer 2, it could be replaced by a confusion matrix.

L211-221: I got confused here because of the text in the parentheses. Why are both the RT and the CR method referred to as CML rain/snow/dry? Wouldn't it be an idea to just refer to them as rain/snow/dry? Or is there a specific reason you included CML?

The difference between the left and right column is that in the left column, the CML rainfall rate was derived from wet periods as classified by weather radar, while in the right we used a CML based wet-dry detection method. We suggest instead using only CML wet detection (keep the right column) as using both radar and CML to identify wet periods is not necessary.

L214: to be fully correct the radar observes precipitation instead of rain (very minor comment)
We agree that this could be stated more clearly.

Fig 5. I understand that you cannot share the exact location of the CMLs because of data security reasons, but is there the possibility to describe the landscape a bit? For example, are there any large elevation differences (i.e., mountiains/fiords) in this area that could for example create orographic precipitation or cause beam blockage of the radar? Additionally, I suggest to add titles to the columns, so that it is immediately clear which method is shown in which column.
Adding the location of the CMLs and indicating the climatological differences would indeed make the article more engaging. We therefore suggest changing the map by focusing on a larger area and distort the CML coordinates slightly so that the climatological differences are visible, but the CML coordinates are secret.

Discussion: Are there any previous studies in which the RT method (or something similar) has been compared with disdrometers (or other precipitation phase observations). If so, I would recommend to include a short discussion on this. How do your results compare to those studies? Are they similar or is it the RT method more difficult in Norway because of for example the elevation differences? Such a discussion would allow to put the performance of the CR method into a wider context.
This is an important remark. We have addressed this remark above in the major comments.

L264: the circle of degree Celsius should be in superscript (very minor comment)
We agree.

L266: How often does a disdrometer misclassify precipitation phase? I would advise you to include a brief discussion based on previous literature regarding uncertainty when using disdrometers to estimate precipitation phase.
We suggest adding the following to the discussion

- "Disdrometers do not provide 100% accurate records of precipitation type. There is notable variability in agreement between identical disdrometer models, which agree about 90% of the time, and this variability tends to increase when different models are compared (Pickering et al., 2021; Friedrich et al., 2013; Bloemink, 2005)"

Fig 6: In the bottom subplot, I suggest to use a different shading color for the diff<-5 mm, because it overlaps with the shadings above. Or perhaps you can even leave this out, as you do not discuss this in the text if I'm correct.
We suggest removing this plot and replace it with simpler plots showing CML time series.

L270-285: Based on this section, it seems that only wet snow could cause misclassifications. I would suggest to make clear that any form of wet precipitation causes the CML signal to drop. Additionally could it also be that spatial temperature and humidity differences can cause that at the disdrometer location dry snow is falling while somewhere else the precipitation has started melting? I can imagine that this happens when temperatures are in the transition region between rain and snow.

To clarify this we suggest adding the following to the discussion

- "We note that although this study focused on rain and snow, any type of wet precipitation can cause the CML signal level to drop, leading to potential misclassifications. Another source of uncertainty is the spatial distance between the disdrometer and the CML midpoint, where temperature variations due to elevation and spatial differences can affect precipitation classification."

L288-289: Are there any previous studies that show differences between radar and CML/disdrometer measurements? If so, I suggest to discuss these here.

See comment to the 3rd general comment.