Responses to comments posted by Referee #2

We thank the second reviewer for their thoughtful comments and suggestions. Their input has been very helpful in enhancing the clarity and quality of our manuscript. The comments are provided below, with our replies shown in blue.

Some methodological concerns raised by reviewers were already addressed in the Discussion section of the original manuscript. To improve clarity and avoid misunderstanding, the revised manuscript presents the Discussion as a separate section with subsections, to which we refer readers at the relevant points.

Overall Assessment

This manuscript presents a novel information-theoretic framework to evaluate uncertainty in rainfall retrieval from commercial microwave links (CMLs) and examines the impact of processing steps and external variables. The study is well-designed, data-rich, and methodologically innovative. However, several issues—particularly in methodological clarity and interpretation of results—require substantial revision. I recommend Major Revision with the following detailed comments.

Major Concerns and Recommendations

1. Methodological Clarifications Needed: Information-theoretic measures (e.g., entropy, conditional entropy) lack detailed derivation, hindering reproducibility. For example, I suggest the authors expand on discretization steps (e.g., binning strategies for rainfall intensity) and clarify how joint probability distributions are constructed from multivariate data.

We clarified the construction and evaluation of the models specifically in the section Methods (L192 - 205 in the track-changes manuscript) and also expanded Eq. (3) (L161 - 162).

We also added references to the equations in the methodology of each analysis for better linkage with the chapter of information theory (L227 - 232 and L243 - 244).

Information theory requires binning to be chosen. We used uniform binning when possible and non-uniform binning for skewed data like rainfall intensity. This subjective step balances preservation of the distribution with the fineness of discretization, as discussed. We elaborate on the binning in the subsection 5.3.3 Binning in Discussion (L543 - 555). We add reference to the subsection in Discussion to the methodology (L223 - 226 and L241).

2. Deeper Interpretation of Results: Mechanisms behind uncertainty reduction (e.g., synoptic types) are not thoroughly explained. Connecting synoptic types to physical processes (e.g., convective vs. stratiform rain attenuation) may bring more insights.

Additionally, you can add spatial plots to illustrate variable-specific impacts on rainfall estimates.

Thank you for the inspiring comment. We added a brief passage in L402 - 408 about synoptic types and their effect on rainfall characteristics and drop size distribution, thus impacting CML rainfall estimates.

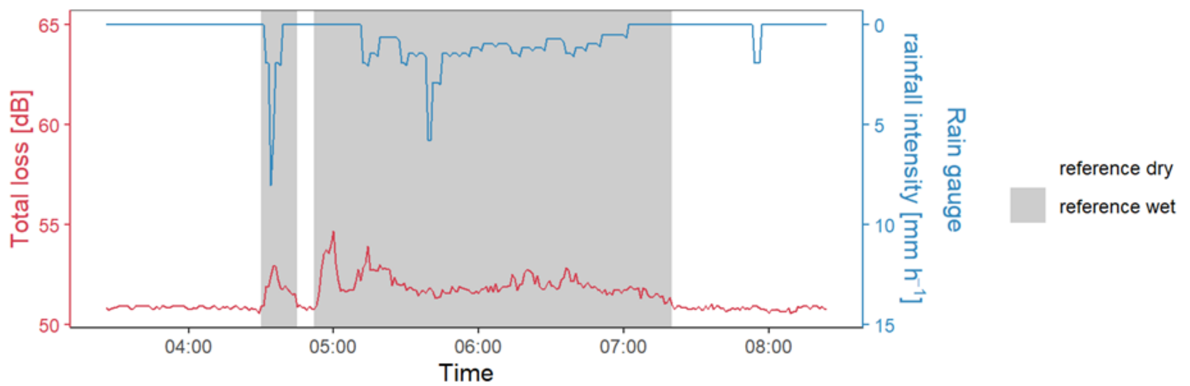
3. I agree with Referee 1's comment. At present, I do not understand why it is necessary to manually label dry or wet data when reference rainfall data is available. However, if this manual labeling is similar to using a simple constant threshold to divide dry and wet data, the so-called reference classification results will inevitably have significant errors. Please explain this issue.

First of all, we would like to clarify that dry-wet analysis is meant as a demonstration study illustrating the feasibility and the potential of a non-parametric model based on information theory. As such, it has limited representativeness. We have revised accordingly the introduction section on L77 - 78 where the analysis is first introduced. In the track-changes manuscript, we make this also clear in the abstract (L20 - 22) and conclusions (L583 - 586). In this context and given the available data (radar data were available only for selected events due to access restrictions), the semi-automatic classification involving visual inspection of CML data was a pragmatic solution.

The nearest rain gauge is located 3 km away from the CML, which is inadequate for capturing rainfall at a 1 min temporal resolution. We have therefore used this rain gauge for initial classification at hourly scale and refined the resolution by inspecting CML time series.

For the purpose of this response, we did an additional analysis testing the agreement of the rain gauge with our semi-automatic identification at the CML location. It shows overall agreement 97.1 %. However, focusing on the wet timesteps, the rain gauge indicates wet only in 50.8 % of wet timesteps at the CML location. The dry timesteps match at 98.6 %. Thus, manual fine-tuning is necessary to obtain a precise reference.

The plot below shows an example time series of the CML total loss and rainfall intensity from the closest tipping bucket rain gauge, with grey shading indicating our semi-automatically identified wet periods.



We add justification of the identification of the wet-dry periods to section 4.1.2. Detection of wet and dry timesteps – wet-dry classification analysis (L316 - 322).

4. Why did this study choose information theory rather than other probabilistic methods? I believe it is necessary to add a comparative discussion in the introduction or methods section to emphasize why information theory is more suitable for this problem.

We added the comparison with other approaches to the Information theory section (L123 - 130).

5. This paper does not explore interactions between predictors. For example, there may be dependencies between CML length and frequency, temperature, and weather type. I suggest including mutual information analysis of variable pairs to identify synergistic effects.

Thank you for this suggestion. This is an interesting topic, nevertheless, outside the objective of this study. We state this now explicitly in the discussion (L514). That said, we emphasize that potential statistical dependencies among predictors are inherently captured in the conditional entropy through the evaluation of their empirical joint probability distributions. Consequently, whether the predictors exhibit independent, redundant, or synergistic effects does not affect the interpretation of their contribution to reducing uncertainty about the target in this study.

For example, longer CMLs often operate at lower frequencies. However, even if these two predictors share some information, this does not necessarily imply that combination of CML length and frequency improves the reduction of uncertainty about the target rainfall intensity. In our dataset, the CML length and frequency have mutual information 0.96 bits, but together they do not reduce any of the uncertainty about the target, as conditional entropy (2.27 bits) equals the unconditional entropy of the target weather radar rainfall intensity).

6. The binning strategy (e.g., for rainfall intensity, attenuation) is empirically selected but has not been rigorously tested. In particular, the last binning value exhibits a sudden change compared to the preceding values (see Tables 1 and 2). I recommend conducting a sensitivity analysis on other classification schemes (e.g., equal-frequency bins, dynamic bin widths). Additionally, the authors should rationalize the bin boundaries for high rainfall intensities (e.g., why 50–160 mm/hour is a single bin).

The measures of information theory are highly influenced by high probability events/states. This is a very attractive characteristic of the approach. It is therefore acceptable to create a bin with greater width for high rainfall intensities as such rainfalls have very little probability in central European weather climate.

Testing of different binning strategies is a challenging task (and even more for multi-dimensional data) including compromising over- and undersmoothing of the data, as concluded in e.g. Álvarez Chaves et al. (2024). They highlighted that appropriate binning strategy requires optimisation techniques or combinatorial approaches. The variance of the distribution function increases with decreasing bin size, whereas its bias increases with increasing bin size (Uda, 2020). We exclude this topic from our scope as it does not directly contribute to our objectives. However, we elaborate on the binning strategy in the subsection 5.3.3 Binning in Discussion (L543 - 555) and we also provide additional changes in the methods (L223 - 226).

Álvarez Chaves, M., Gupta, H. V., Ehret, U., and Guthke, A.: On the Accurate Estimation of Information-Theoretic Quantities from Multi-Dimensional Sample Data. Entropy, 26(5), 387. https://doi.org/10.3390/e26050387, 2024.

Uda, S.: Application of information theory in systems biology, Biophysical Reviews, 12, 377–384, https://doi.org/10.1007/s12551-020-00665-w, 2020.

7. As the author mentioned, the ratio of dry period data to wet period data is approximately 97:3. I do not believe that the ROC curve is the best evaluation metric in such an unbalanced sample. Why not try the F1-score or the Matthews correlation coefficient (MCC)?

Thank you for this idea. You are right that the classification in the confusion matrix and its depiction in a ROC plot can be misleading as ROC can be optimistic for imbalanced data as it treats false positives and false negatives equally. We have chosen the receiver operating characteristics as it is easily understood with basic knowledge of statistics. However, we followed your advice and also expressed the Matthews correlation coefficient (MCC) for the comparison with other classification approaches. The MCC is introduced in the methodology (L268 - 273) and its results are shown in Fig. 6 (L498) and in the text (L484 and 491 - 493).

8. The author states that the rolling standard deviation (RSD) method proposed by Schleiss et al. is the state-of-the-art method for dry-wet classification. I disagree. The RSD method can be considered the most commonly used method, but it cannot be considered the state-of-the-art method. Recently, many machine learning methods have been applied to dry-wet classification, and they exhibit better performance than the RSD method.

Thank you for this comment. You are right that, although the rolling standard deviation (RSD) method by Schleiss and Berne (2010) is widely used, it does not represent the latest advances. We agree that “state-of-the-art” is not an appropriate term. We therefore follow your suggestion and name their approach as “commonly used benchmark”. The method is used frequently; therefore, we use it to make the comparisons. We also added references to other studies that used the method by Schleiss and Berne (2010) in the text (L283 - 284).