

## Final response

This is our final response after the discussion phase and it addresses both referee comments. During the open discussion phase we already provided a quick, but comprehensive, response to referee 2 (Andreas Scheidegger). It was meant to provide feedback from our side to encourage further contributions. In the following, we will first give a summary of our assessment of the major issues, as they were reported by the two referees. We will then discuss these major issues and propose the related changes and additions for a revision of our manuscript. Finally we will list all individual comments of the referees and our corresponding responses, referring to our general answers where appropriate.

### Summary and assessment of major referee comments

We thank both referees for their critical assessment of our manuscript. According to their comments they acknowledged that the manuscript is well written and that our analysis is scientifically correct. Both referees have pointed out major issues, though. According to the individual referee comments, these major issues are:

1. **Lack of scientific significance:** In contrast to referee 1, who rates our manuscript to have “good scientific significance”, Andreas Scheidegger’s main objection is “poor scientific significance” and he therefore recommends a rejection. While we clearly see that we have to improve on the justification of our research in the manuscript, we do not agree with this assessment. Our detailed response and the proposed changes for a revised manuscript can be found in the next section.
2. **Legitimacy of comparing to method based on rolling standard deviation:** The comparison with the rolling standard deviation (RSTD) method of Schleiss and Berne 2010 [1] is criticized by both referees, albeit for different reasons. Referee 1 states that an intercomparison with another neural network architecture, namely LSTM, is essential. Unfortunately, a clear reason for rejecting the comparison to the RSTD method is not visible from their argumentation. Andreas Scheidegger is concerned about the simplicity of our chosen reference method and that this diminishes scientific quality. Our main argument for using the method is that it is still state of the art. Thus, using it as a reference enables other CML researchers to put the performance gain into perspective. We explain this argumentation and our proposed changes for the manuscript below.

### Summarizing answers to major issues and proposed changes for a revision:

**To 1. Lack of scientific significance:**

In our answer to Andreas Scheidegger's general comments on our work (See AC1 of the discussion) we already discussed the importance of rain event detection for the CML-rainfall community (backed up by supporting statements from other community members that deal with the same kind of CML data). In summary, we acknowledge that the relevance may be low from a machine learning perspective, but this is not a computer science paper and it was not submitted to a computer science journal. As explained in AC1 the relevance is given by the application, which is the improvement of quantitative precipitation estimation with commercial microwave links. To appear shallow in one discipline is a common hurdle for interdisciplinary research items, although interdisciplinary research is wanted by the scientific community. Our suggestions for improving our manuscript in order to highlight the relevance for the application are the following:

- i. We will revise the introduction by adding a more detailed review of previously proposed methods for rain event detection, separating them into application on the different data acquisition types of min/max and instantaneous sampling. Methods developed for one kind of data acquisition are in general incompatible with the other kind. Whenever this information is publicly available, we will review the problems previous methods had with false detections and missed events. While many previous studies were event based evaluations, our study is free of any preselection which is a necessary analysis for potential operational applications.
- ii. To show the significant impact of rain event detection on estimated rain rates, we will add an analysis about the rainfall overestimation through falsely classified rain events and underestimation through missed events in absolute numbers. Using the same processing scheme as in Graf et al. 2019 we will discuss the improvement over the previous method in terms of absolute hourly and monthly rainfall amount. A preliminary evaluation is shown in the plot below from which we can derive a reduction of the overestimation through falsely classified rain events by 27% of the monthly overestimation through the Q80 method. Additionally the CNN reduces the underestimation through missed events by 15%. A larger evaluation and description of the derivation of the final rain rates will be added to the revised manuscript.

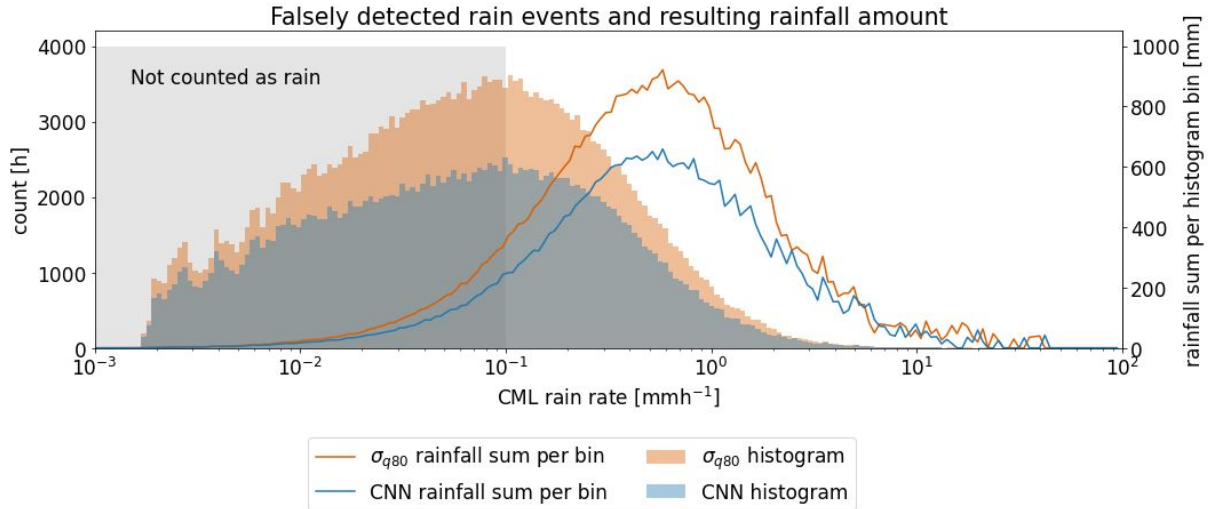


Figure 1: Histogram of false positive hours (FP) in April 2018 (reference rain rate below 0.1 mm). At a threshold of 0.8 the CNN reduces the number of falsely detected rain events and therefore the total falsely generated rainfall amount by approximately 27%. At the same time the CNN still misses less events than the standard deviation method (not shown in this plot).

## To 2. The comparison to the method of Schleiss and Berne 2010 [1].

In our opinion, the comparison is justified as follows:

First, the method should still be considered state of the art due to the works of Fencil et al. 2020 [3], Graf et al. 2019 [4], and Kim and Kwon 2018 [5] using the rolling standard deviation according to [1] for rain event detection. Additionally De Vos et al. 2019 [6] are using correlation to nearby links and Ostrometzky and Messer 2018 [7] propose a simple rolling minimum to set the baseline level. Both methods are of similar or lower complexity than the RSTD method which is still applied to CML attenuation data due to its robustness (in the sense that only few parameters have to be tuned) and easy applicability. In our work, we show a significant improvement over the RSTD, while the resulting model is just as easy to apply using our pre-trained model, which we also share as open source software.

Second, although very simple, the RSTD method is not performing poorly. In a Lab setup rain events could be purely detected using the rolling standard deviation due to the justified assumption that fluctuations are bounded during dry periods and they exceed the boundary even for small rain events. Unfortunately, this does not always apply to real world data, where strong signal fluctuations occur due to e.g. multipath effects. The challenge is to separate those artifacts from real rainfall fluctuations, and the fact that Andreas Scheidegger is not surprised that the CNN can outperform the

RSTD solidifies our assumption that we picked the right model for the large scale real world application.

We believe that this justifies our choice of comparing to the RSTD method from [1]. We propose to revise our introduction by citing the works that underline the state of the art status of using the RSTD method and we will add more examples of real world CML attenuation time-series that underline the challenges which can not be solved by this method, thus generating the knowledge (or performance) gap, that our method seeks to fill.

The question which artificial neural network architecture should be used is therefore of secondary relevance. Previously, LSTM was used for min/max sampled attenuation data and a small amount of CMLs [2]. The main point of our study is showing the potential of an artificial neural network approach for instantaneously sampled data from a large amount of sensors likely to occur in an operational setup and certify the robustness for further application. This was not done before. Again, we believe that showing this potential by comparing to the state of the art is justified, if not necessary.

## Direct answers to referee comments:

Note: The complete referee comments are copied here using italic font, our response uses normal font.

### Anonymous Referee #1

*In general, this paper demonstrates the use of a 1D convolution neural network for the task of Wet-Dry classification using commercial microwave links (CMLs). The scientific significance of this paper is in presenting the potential of the suggested method for the specific application: the use of 1D CNN for wet-dry classification with commercial microwave links. But, without any theoretical justification for the use of 1D CNN, it must be compared empirically with other algorithms/methods. The results are shown in this work only compare 1D CNN with a model-driven method [1], however, the suggested method must be compared with another data-driven algorithm, previously suggested (and cited by the author) - the use of LSTM for wet-dry classification [2]. This comparison is important also since the LSTM can capture long sequence, while the CNN only see a fix window size of the attenuation time series. Additionally, the authors didn't use the CML's parameters (e.g. link length, frequency, and polarization ) as an additional input to the neural network, which may make this method more sensitive to differences in those parameters.*

We thank the referee for their critical assessment of the comparison to the RSTD method which is used to show the potential of our proposed method. Our justification for doing this comparison is given in paragraph 2 above. The use of LSTM is discussed in 1

*Specific comments required for the paper to be acceptable for publication:*

*1. Comparing the results of the LSTM and CNN on the same data set is essential.*

To 1.: While it is true that LSTM is a common neural network architecture which might also be applicable to the task we miss a fact based justification for 'must be compared with another data-driven algorithm'. As written above, the use of a rolling standard deviation can be considered state of the art. The method makes heavy use of long term CML statistics and can be called a statistical method. We therefore do not understand what the word 'model-based' should mean in this context or why the comparison to the RSTD method should be unfair or of less scientific relevance. Apart from that, our justification for using CNNs is the generally accepted fact that they are good in

recognizing patterns independent of their location within a longer sequence of the time-series and this is also what we wrote in the manuscript (see lines 8, 67 and 158). As we already stated in paragraph 3 of AC1, LSTM is a common ANN architecture for time series analysis. This does not mean that CNNs, which have the benefit of very fast parallel processing, are not applicable to time-series data. Indeed, our results prove that CNNs are a valid processing tool for one-dimensional data.

*2. Study the effect of different window sizes on the performance of the proposed method.*

We evaluated the effect of longer window sizes and did not find a significant improvement. But we agree that information about suitable window sizes is important. Hence, in the revised version, we will shortly describe the different setup with a 5 hour window length and add the results to table 2.

*3. Study the effect of different CML's parameters (e.g. link length, frequency, and polarization).*

We evaluated the effect of adding the CML parameters and did not find a significant improvement. According to the k-R power-law, CML frequency and length influence the amplitude of the rainfall-induced fluctuations. E.g. short CMLs with a comparably low frequency are less sensitive to rainfall. We expected that this information about the CML sensitivity to rainfall would help the classification performance. We did not see a relevant improvement, though. One possible explanation is that knowing the CML parameters does not help to detect wet periods close to or below the individual CML's detection limit, since there just might not be a detectable signal. Furthermore, according to our experience with CML time series, the occurrence of signal fluctuations during dry periods, for which distinguishing between wet and dry is most challenging, does not depend on CML parameters.

Initially, we decided to only describe the less complex and equally well performing setup without the CML parameters for the sake of brevity.

We will now include our results from the CNN with the CML parameters in the revised manuscript, by shortly describing the different setup and adding the results to table 2.

## Andreas Scheidegger:

### *\* General*

*The manuscript describes the application of a one-dimensional convolutional neuronal network (CNN) to classify wet and dry periods based on microwave link attenuation data. The CNN is compared against a very simple classification scheme that is only based on the standard deviation of the signal. Not surprisingly, the CNN performed better.*

*The manuscript is well written and the underlying work seems solid. Still, in my opinion, this paper lacks ambition and innovation to deserve a publication in AMT. As the authors mention, various ANN's and other machine learning techniques have previously been applied in different settings on MWL data. Also the wet/dry classification problem does not appear particularly challenging from a machine learning perspective. Furthermore, for time-series data recurrent neuronal network architectures (e.g.LSTM) seem a more obvious choice (which could be combined with convolution layers if needed).*

This is, in our opinion, the main issue which we discuss in our summarizing answer 1. above and in the majority of AC1. In summary, rain event detection is a necessary processing step to set the baseline signal level, required to derive the specific attenuation  $k$  which is then used to derive the rain rate  $R$  via the k-R relation. The community state of the art are methods like the RSTD and the problems with false detections can be shown from recent publications. Although less relevant from a machine learning point of view, the relevance for the application is high.

*A more interesting question would be to investigate if we can train a ANN to predict the rainfall intensities directly, and so avoid all submodels for baseline estimation, wetantenna correction, and so on. Such a model could also make use of additional information, like MWL properties, frequency, maybe temperature, ...*

This is answered in AC1 paragraph 4.

*A good transferability of the trained model to a region with different climate is key for an application where no reference data are available (such as in the mentioned Burkina Faso). This could have been partly emulated by training the model in one region and then validating it in a region with different climate. Or by training the model in winter and validating it in summer.*

This is answered in AC1 paragraph 5.

*I'm sure the current work offers the authors a solid foundation for more ambitious investigations.*

No answer required.

*L 90: "...it has to be proven that artificial neural networks allow for high-performance, fast and robust processing of large data sets..." - I think this is already proven by countless other application.*

This statement has to be read in the context of CML data sets. The purpose is not to show that CNNs can achieve this in general, which is indeed proven on countless examples. To show that the behaviour of a large CML data set can be predicted by training only on a comparably small subset is still to be proven for the application and the CNN is our method of choice to do so. To our knowledge no previous study achieved or even investigated this degree of generalization. In fact, most previous studies use a setup that uses long time statistics of all individual CMLs for a low amount of CMLs and many of them deal with min/max sampled data, which is different from our instantaneous measurements. We will adjust the relevant paragraph in the manuscript accordingly, to better explain that.

*L 145: Besides the attenuation data for the hour to classify, the Network was also feed with the two hours of "old" data. Did this improve the classification? If yes, it would indicate that attenuation data have some kin of memory effect (antenna wetting?).*

Adding the two hours of "old" data has a positive effect. Antenna wetting is kind of a memory effect since it will increase the baseline level during the rain events, but it will also keep the baseline level increased after the rain event. With the drying of the antenna, the baseline level then decreases slowly after the event. Since this drying process is quite continuous it does not lead to strong rain-like attenuation fluctuations of the signal level. Hence, we think, the wet antenna effect is not the reason for the improvement with the added 'old' data. Our reasoning for using 'old' data was that this data provides more context for the CNN, i.e. there is a lot more information on how a CML time series generally 'behaves'. Also as a human it is a lot easier to distinguish rain events from dry fluctuations when the available time period is larger. Humans are, compared to CNNs, unfortunately very slow in doing so. In addition to the longer



window size, as requested by reviewer 1, we will include the performance when not providing any old data to show this improvement.

*L185: Where did you add the dropout layers? How many?*

Dropout was used between the fully connected layers, i.e. two times. We will indicate the exact location in figure 2.

*L216: Are TP, FP, FN, and TN defined?*

We introduced this notation in table 1. We will add a direct reference to the table at the first occurrence of TP, FP, FN, and TN. We will also add a set theoretic definition in addition to the table.

*L230: What is the advantage of the MCC compared to the ROC?*

The Matthews Correlation Coefficient is a single number, which can be used to optimize the threshold for the CNN or the thresholds of the rolling standard deviation. The ROC consists of the two numbers FPR and TPR. Using them for optimizing a threshold is not straight forward and would require a cost function that can be minimized. The advantage of ROC is that the performance of classifiers with a variable threshold can be compared independent of a fixed threshold value by considering the ROC curve. We will clarify these different purposes in the respective parts of the method section.

## **References:**

[1] Schleiss, M. and Berne, A.: *Identification of Dry and Rainy Periods Using Telecom-munication Microwave Links*, IEEE Geoscience and Remote Sensing Letters, 7, 611–615, doi.org/10.1109/LGRS.2010.2043052, 2010.

[2] Habi, H. V. and H. Messer, 2018: *Wet-Dry Classification Using LSTM and Commercial Microwave Links*, IEEE 10th Sensor Array and Multichannel Signal Processing Workshop(SAM), doi.org/10.1109/SAM.2018.8448679

[3] Fencil, M., Dohnal, M., Valtr, P., Grabner, M., and Bareš, V.: *Atmospheric observations with E-band microwave links – challenges and opportunities*, Atmos. Meas. Tech. Discuss., <https://doi.org/10.5194/amt-2020-28>, in review, 2020

[4] Graf, M., Chwala, C., Polz, J., and Kunstmann, H.: *Rainfall estimation from a German-wide commercial microwave link network: Optimized processing and validation for one year of data*, Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2019-423>, in review, 2019.

[5] Kim, M.-S. and B.H. Kwon, 2018: *Rainfall Detection and Rainfall Rate Estimation Using Microwave Attenuation*, Atmosphere, 9, 287, [doi.org/10.3390/atmos9080287](https://doi.org/10.3390/atmos9080287)

[6] de Vos, L.W., A. Overeem, H. Leijnse, and R. Uijlenhoet, 2019: *Rainfall Estimation Accuracy of a Nationwide Instantaneously Sampling Commercial Microwave Link Network: Error Dependency on Known Characteristics* J. Atmos. Oceanic Technol., 36, 1267–1283, [doi.org/10.1175/JTECH-D-18-0197.1](https://doi.org/10.1175/JTECH-D-18-0197.1)

[7] Ostrometzky, J., & Messer, H. (2018). *Dynamic determination of the baseline level in microwave links for rain monitoring from minimum attenuation values*. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11(1), 24–33. <https://doi.org/10.1109/JSTARS.2017.2752902>