Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2019-412-AC1, 2020

Atmospheric Measurement Techniques Discussions
Open Access

Interactive comment on "Rain event detection in commercial microwave link attenuation data using convolutional neural networks" *by* Julius Polz et al.

Julius Polz et al.

julius.polz@kit.edu

Received and published: 13 February 2020

Authors response to general comments by Andreas Scheidegger

Dear Andreas Scheidegger,
Thank you for reviewing our manuscript and for providing your criticism which made us reflect our analysis and in particular how we justify its relevance. Here, we want to briefly respond to the individual general points of your review. Please note that our suggestion for specific changes, adjustments and extension for the revised manuscript will be provided in our final response, after the open discussion period.

1. Importance and relevance of the rain event detection in TRSL time series data:

A: Based on our experience: We acquire data of 4000 CMLs with one minute resolution in real-time every minute and we work towards providing real-time rainfall estimation from that data. From the experience that we gathered doing this over the last years, we can state that detecting rain events, or more specifically, separating TRSL fluctuations during dry times from those during rainy times, is key to produce reliable rainfall estimates. If CML rainfall estimates shall be operationally applied, missed events (false negative) and false rain event (false positive) must be minimized as good as possible.

B: Based on the experience of the CML rainfall community: We have presented the work that is summarized in the current manuscript in June 2019 at the *symposium on the hydrometeorological usage of data from commercial microwave link networks* which was attended by many colleagues that actively work with CML data for rainfall estimation. As far as we can recall now, the dedicated analysis of rain event detection and our approach with CNNs was appreciated by our colleagues, except for the mentioning that LSTMs are maybe better suited for time series classification (for our response to that, see point 3 in this response letter). To our knowledge only two other research groups worldwide currently acquire a CML data set with a temporal resolution and absolute size similar to that of our CML data set. At SMHI data from several hundred CMLs is acquired at subminute-scale and processed to rainfall estimates (*link to project website*). At CVUT (Czech Technical University in Prague) a similar project called *tel4rain* is ongoing. Reflecting on your review, we have contacted both groups to discuss if research for the processing step of rain event detection in TRSL time series is still required and relevant. Both groups confirmed that they consider improvements in rain event detection in TRSL time series still very relevant, in particular when working towards operational usage of the CML derived rainfall estimates. We very much appreciate the detailed feedback on this question from Martin Fencl (martin.fencl@cvut.cz) from CVUT, who agreed to appear here with his contact details. Please note that we only reached out to discuss this specific issue and not the scientific details of our approach, using CNNs.

Again, we thank you for drawing our attention to the fact that we failed to justify the relevance of our results to the wider hydrometeorological community. This paper exists to validate a method that is about to be used within our community and to set a standard for benchmarking CML processing, which is very likely to keep the localization of rain events as an isolated processing step (see 4.). We will do our best to revise the manuscript accordingly using further argumentation as written below. We also want to clarify that this is not an effort to minimize work that we have to invest in revising the paper, but an effort to justify the topic and the relevance of the paper. We are happy to receive further constructive suggestions (like in 5.) on how to improve our manuscript.

2. Ambition and innovation:

First of all, we have to apologize for a mistake, which may be important in this discussion. In our review of previously used methods for the task of wet/dry classification, we wrote that Kim and Kwon 2018 [1] made use of LSTM networks. It turned out that this is not the case and that they also used a rolling standard deviation as their main criteria to separate wet and dry periods, similar to the method we compare to. This reduces the previous attempts to use deep learning for the task with no peer reviewed studies using more than one CML, that we know of. We believe that our work is novel in the following aspects:

A: This is the first time a large CML data set is used in combination with a data driven method. Previous attempts ranged from using a single CML up to the 34 CMLs used by Habi and Messer 2018 [2]. It is important to say that the CML data used by Habi and Messer is different from our data set since they use 15 minute min/max values for attenuation and information about different types of errors while we use instantaneous measurements at a one minute resolution. Therefore, not counting their approach to do the same data processing as we do, this paper is the first of its kind that uses more than one sensor. Many other groups also use instantaneously measured CML signal levels, but they might only have access to a handful of CMLs. Therefore, we believe that it is important for the scientific community to be able to use a model trained on our large network. With the initial submission we already made our trained model available via zenodo and github.

B: Though indeed very simple, methods like the rolling standard deviation, rolling median or correlation to neighbouring CMLs are still state of the art due to their known stability and easy applicability, although being less performant. A recent example is the analysis of country-wide CML data in the Netherlands by de Vos et al 2019 [3]. They use a rain event detection based on correlation between neighbouring CMLs. As their Fig. 4 shows, their approach works in general, but the amount of points along the y-axis clearly shows that false positives (falsely detected rain events) are an issue. Our work is the first evaluation of a data driven classifier, where the robustness of the method is a central point of the evaluation and proven on a large and diverse data set, facilitating the application for other researchers and operational users.

In conclusion, we understand that our review of previously used methods is lacking detail and does not manage to convey the still persisting challenges in rain event detection. We will better describe the state of the art and improve our explanations on how we advance it with our work.

3. CNN vs. LSTM

It is true that recurrent neural networks (RNNs) are a dedicated tool for time-series data, especially for time-series prediction or sequence to sequence learning. Convolutional neural networks (CNNs) on the other hand are mostly used for classification tasks on images. While RNNs do not apply very well to image data, there is no reason why CNNs would not apply in one dimension. If CNNs perform well in a classification task, like the one we present in our paper, they have one major advantage: They are much faster in training and inference (which was approximately a x40 speed-up in a preliminary test using a typical LSTM architecture) and also more stable during training. For real-time data processing with a short temporal resolution, how it is envisaged for CMLs, this is very important. Therefore, we want to emphasize that our results show that CNNs are a valid processing tool for one-dimensional data. On top of that, we believe that computational resources should be saved, unless absolutely necessary.

4. ANN for rain rate estimation

This is indeed an interesting topic. In fact, we already did such an evaluation and for now arrived with the following conclusions: The results are good, but not overwhelming which is probably due to the fact, that in the hourly RADOLAN reference the uncertainty of the absolute rainfall amount is much higher than the uncertainty of the temporal localization of an event. Learning absolute rain rates from this reference is therefore not our goal, since we want to use the attenuation-rain-rate (k-R) power law, which is very insensitive to DSD variations. This way, we avoid making absolute CML rainfall amounts mostly a "radar-adjusted" precipitation product. Another reason is that the wet antenna attenuation is still a big unknown, which should be investigated in the future and we do not want to mix this processing step with the detection or

the k-R relation, since there is an unknown risk of being "right for the wrong reason" when comparing to the weather radar. With the optimized event localization through the CNN and the near linear relationship between attenuation and rain rate, there is a promising chance of new insights into the wet antenna effect.

5. Training and/or validation for different climates

We appreciate this suggestion and we will do experiments for using CML subsets from different regions of Germany. Since CML data might be compromised during winter time, due to wet snow and ice covers on the antennas, we are, however, not sure if we can separate training and validation between winter and summer to simulate a large climatological difference. Regarding CML data from other climatic regions, we unfortunately do not yet have enough CML data for Burkina Faso and in general we are lacking reference data there. We just started with a project (AgRAIN) to improve both issues, but it will take some time to have an effect.

References

[1] 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

[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] 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

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2019-412, 2019.