REVIEW OF THE PAPER "Improved rain event detection in Commercial Microwave Link time series via combination with MSG SEVIRI data", AMT 2023-175

General comment

The authors compare different algorithms for detecting dry/wet time slots from opportunistic measurements collected by Commercial Microwave Links and, at the same time, they assess whether the wet/dry classifiers can be improved using satellite data collected by MSG SEVIRI. Dry/Wet slot classification is an important step in the processing chain of CML data to extract quantitative precipitation estimates. This is a very specific aspect of the topic of opportunistic sensing of precipitation by CMLs, which I believe fits the scope of AMT. Moreover, I think it has some novel aspects as it is one of the first papers demonstrating the effectivity of data fusion between opportunistic sensors and products from Earth observation satellites.

First the authors check whether SEVIRI products are reliable comparing them with RG-adjusted radar data, chosen as a reference (sec. 4.1). Then, they analyze the performance of several wet/dry classifiers based on CML data only, on SEVIRI data only and on a combination of both. Finally, they assess the impact of different wet/dry classifiers on the performance of CML as quantitative rainfall sensors. The logic sounds correct. However, I think the datasets used, the methods and the results are not well explained. There are several aspects to be clarified: hence, in my view, this contribution needs a major revision before being rated as acceptable for publication.

Specific comments

I am kindly asking the authors to address the following specific comments:

- Sec. 3.2 (about methods): I think the wet/dry classification rationale by each of the 3 different types of sensor is not well explained and some information is missing in my view (or maybe I missed it).
    - I suggest to add an itemized list in Sec. 3.2 and explain wet/dry rationales for radar, CML and SEVIRI.
        - Radar data: I missed the way how raw radar data (I guess rainfall intensity) are processed and combined together to build the reference wet/dry time series. Are radar data thresholded above 1 mm/h and then mapped over each CML path? How do you flag a CML as wet/dry from the weighted average of overlapped radar pixels? Is the 1 mm/h threshold applied on the weighted average instead? How did you put together radar and CML time scales (5 min and 1min)? Please clarify.
        - SEVIRI data: there are as many wet/dry time series as are the pNN labels (SEVIRI outputs). For instance, the time series p10 at a certain pixel is generated assuming the pixel wet when it is flagged as p10. Is it correct?
        - CMLs: two wet/dry time series are produced for each link, corresponding to two different methods (RS and CNN). Is that correct?
    - I suggest to add a table with time step and spatial resolution of each sensor. I didn't get which is the radar pixel size. SEVIRI's pixel size depends on the elevation and azimuth of the observed point on the Earth (i.e. Germany). It is not stated which is the range of SEVIRI pixel sizes for Germany. Finally, we can say that the spatial resolution of CML equals their length if no path reduction factor is used.
    - At least in the text, please provide absolute numbers of the populations involved: we are talking about 10,000, 1,000,000 or even longer time series? And what is the total dataset size (number of sensors x number of samples)? Finally, would be good to have some min-max range for positives , i.e. wet link occurrences in the observation period.

- Introduction, lines 61-67: even though they have been proposed for satellite links, it would be good to take a look at the following methods (as dry/wet identification in terrestrial and satellite links are very similar):
    - L. Barthes et al., "Rainfall measurement from the opportunistic use of an earth–space link in the ku band," Atmosph. Meas. Techniq., vol. 6, no. 8, pp. 2181–2193, 2013.
    - C. H. Arslan, et al., "Satellite-link attenuation measurement technique for estimating rainfall accumulation," IEEE Trans. Geosci. Remote Sens., vol. 56, no. 2, pp. 681–693, 2017
    - F. Giannetti et al., "The nefocast system for detection and estimation of rainfall fields by the opportunistic use of broadcast satellite signals," IEEE Aerosp. Electron. Syst. Mag., vol. 34, no. 6, pp. 16–27, 2019
    - B. He et al., "Use of the c-band microwave link to distinguish between rainy and dry periods," Adv. Meteor., vol. 2019, 2019
    - M. Xian et al., "Rainfall monitoring based on machine learning by earthspace link in the ku band," IEEE J. Sel. Topics Applied Earth Observ. Rem. Sens., vol. 13, pp. 3656–3668, 2020.
    - R. Giro et al., "Real-time rainfall estimation using satellite signals: Development and assessment of a new procedure," IEEE Trans. Instrum. Meas., vol. 71, pp. 1–10, 2022.
    - C. Gianoglio, et al., "Rain discrimination with machine learning classifiers for opportunistic rain detection system using satellite micro-wave links," Sensors, vol. 23, no. 3, p. 1202, 2023.
- Sec. 3.3: Performance indicators
    - is it really beneficial to introduce all those indicators? In principle, sensitivity (i.e. TPR according to the authors) and specificity (not considered by the authors) should be almost all we need. MCC can be useful as it is a global indicator combining FP and FN rejection, but it is not as straightforward as the previous two. It's not easy to state how good is an MCC value equal to 0.60-0.62 and how much an increase of MCC by 0.09 and 0.13 (I took these numbers from the abstract) is indeed valuable. Indeed, it is not obvious to make ratings of methods based on the MCC values in Figs. 5 and 7. Finally, please note that the importance or having high sensitivity rather than high specificity methods or vice-versa, depends on how wet/dry classification is used within the CML processing chain. Is it used to calculate the baseline? In this case, not misclassifying wet slots as dry is critical, i.e. sensitivity is the key indicator. I think the way you look at wet/dry classification as a part of data processing will drive the choice of the most significant performance indicator. Some discussion and a better justification of the indicators used is needed, instead of listing formulas and writing definitions.
    - ACC is shown in Fig. 3 (purple bars, I see small differences among bars) and in Fig. 9, first column. PCC is shown in Fig. 3 and Fig. 9 as well. Unless these two highlight different aspects of the confusion matrix than MCC, I think it's simpler and better to show only MCC throughout.
    - The PPV indicator sounds a bit ambiguous to me in the framework of CML data, because it gets low (i.e. poor performance) if either there are a lot of FPs with respect to TPs, but it is also low for a given number of FPs if there are few TPs on the whole. FPs are often produced by factors other than rain, hence they have not in a strict relation with the number of TPs.
    - An interesting investigation would be to assess where errors (FP and FN occurrences) actually are. Are they at the start of an event rather than at the end? Or there are sequences of FPs far from events? Have the authors done such kind of analysis?
- Figures 3,4,5,7: it's not a good idea to put "night" or "day" as labels on the y-axes of those figures. You should put the quantity displayed in the graph instead. Night and day should be placed into a text box in a free space over the figure or as descriptors after subfigure identifiers (e.g. a) , b), etc.) above each graph. Same for the x-axis: for instance, in Fig.3, the authors used different labels for the

x-axis of subfigures a) and b): actually, they are the same axes. And they put TPR (y-axis) on the x-axis. The figures are indeed complex and they aggregate too much information spread in too many dimensions (night/day, percentage of rain occurrence as from SEVIRI products, rain intensity class, statistical indicator). Moreover, those small pictures on the top right of Fig. 3 are not well described: I guess here the TPR is weighted with the accumulated precipitation, that's why it is different from the blue histogram of Fig. 3b. Not straightforward really. It definitely looks too much and too difficult to track. Please simplify, dropping less meaningful dimensions. For instance, in Fig. 3a, TPR shows minor differences between night and day, as the authors state on page 10. Hence, one of the graphs of Fig 3a) can be dropped in my view. Also, I can't see any significant difference or trend between the two different SEVIRI products in Fig 3a.

- Figure: 3: the authors show TPR of SEVIRI-based wet/dry classifier assuming radar-based classifier is the truth.
    - Is this comparison carried out mapping radar and SEVIRI pixels onto CML paths or it is just that the SEVIRI grid has been mapped onto the radar grid?
    - I didn't get the trend of TPR as a function of pNN, where NN is the probability of precipitation in percent (as the authors stated on page. 5), being pNN a SEVIRI product. So I expect TPR to increase with pNN, as SEVIRI wet/dry time series with high pNN have less wet slots than the ones with small pNN, hence a lower probability to incur in FNs. Why is it the opposite? On the other hand, the trend of PPV with pNN in Fig. 3b looks to agree with the feeling that a SEVIRI sample with high pNN is really a wet sample.
- Fig. 4: I have several comments here.
    - on Page 11, Line: 257 the author state: "The distribution of rain intensities and total precipitation amount of eight rain event detection methods is shown in Fig.4". It's way too generic. The explanation in the figure caption does not really help either (it should be improved as well). Finally, the word "count" in the x-axis label of the figure has nothing to do with rain intensities (following the logic of Fig. 3, count is the quantity displayed on the y-axis). So, what did you actually plot in Fig. 4? Is it a comparison between rain intensity/depth estimates across CML paths done by CML (2 methods)/SEVIRI (6 products) vs radar measurements? Or is it the percent difference between CML/SEVIRI wet counts and radar wet counts averaged over the population of a certain intensity class further divided by day and night? By the way, when you described SEVIRI data in Sec. 2.3, you didn't state clearly if precipitation intensity/depth are among SEVIRI products or retrievable from SEVIRI products., so I guess you are talking about counts. Please clarify and correspondingly edit your manuscript and it is really hard to get out of this based on what is written in explanations. Sec. 4.2.1 and in the figure caption.
    - In their explanation of the figure on page 11, the author state that "The SEVIRI-based ADB data sets behaved very similarly to the two TSB data sets RS and CNN". When I see it, TSB data work better than most of SEVIRI products except for the high-intensity class.
    - "a tendency of radar data to underestimate heavy rain intensities" (page 11 to justify CML/SEVIRI overestimate at large intensity) is a dangerous statement in my view. So far, the authors considered radar data as the truth and now they state that they maybe not good in estimating heavy rainfall. The author should provide a convincing evidence based on literature focused on German radar data, rather than citing a paper (Schleiss 2020) that worked over data from other radar networks.
- Results: in Fig. 9., the authors assessed the effectivity of a classifier based on mixing together TSB and ADB methods by performance indicators derived from the confusion matrix. The authors stated in the abstract "The separation of the attenuation time series in rainy and dry periods (rain event detection) is the most important step in this processing and largely determines the quality of the resulting rainfall estimates." So, a reliable classifier will end up in improving CML-based rainfall

estimates. This is not demonstrated, however. It would be really good to see how Pc10all decreases the error on rainfall intensity estimates with respect to Pc10 and CNN through a scatterplot of rainfall intensity/depth as the ones in Fig. 9.

Technical corrections

- Page 1, pp. 38-39: “Nevertheless, gauge-adjusted radar data is considered to be the best possible data basis”, quite a strong statement if you ask me. I would be happy to state that RG calibrated radar data are considered reliable for estimating precipitation over large areas.
- Page 3, line 65: a full stop is missing
- Page 3, line 72: NWC SAF acronym not explained and written as one single word on following line 84.
- Page 3, line 74: “has carried out analyzes” is misspelled
- Page 5, line 125: “in Thoss” rather than “at Thoss”
- Page 5 line 126: a full stop is missing
- Page 7, line 189: “In step 1, we choose a method (either CML-TSB or SEVIRI-ADB) that shows a good average performance as a first guess”
- Page 7, line 193: Section rather than Chapter
- Page 10, Line 230: “To assess the quality of SEVIRI data it was 230 directly compared to radar data”, this statement sounds a bit awkward: a comma is missing at least, or state it better.
- Page 10, Line 237: “over different rain intensity”, would state “over different rain intensity values”
- Page 10, Lines 232-33: “TPR shows the percentage of wet RADKLIM-YW time steps per intensity class and precipitation amount, represented by SEVIRI data for different thresholds”. Badly written. I am afraid this has to do with the comment I did above about the complexity of figures. I would state it in a simpler way: “Fig. 3a shows the TPR (in percent) of SEVIRI wet/dry classification at day (top) and night (bottom) divided per intensity class and per probability precipitation”.
- Page 10, Line 240, “P30 showed the opposite”. Cannot get it. p30 just shows that the p30 population of SEVIRI wet slot is closer to the population of RADAR wet slots then the p01 population (as PPV_p30> PPV_p01).
- Page 13, lines 290-292: I guess dew effect is temperature dependent. I am not asking the authors to do such an analysis, but maybe processing data based on local temperature classes would help to clarify this point.
- Page 17, caption of Fig.8: “The rain intensities for FP and TP are estimated by the CML, while the rain intensities for FN are taken from the reference”. I cannot get it. Why didn’t you take rainfall intensities from the reference (i.e. radar) all the say? Please clarify.