Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC2
https://doi.org/10.5194/hess-2021-389-RC2, 2021

Review of HESS-2021-389

Anonymous Referee #2

Referee comment on "Hydrological response of a peri-urban catchment exploiting conventional and unconventional rainfall observations: the case study of Lambro Catchment" by Greta Cazzaniga et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2021-389-RC2, 2021

This manuscript presents an interesting study, analysing the performance of CML rainfall estimates in a relatively small and complex catchment. The authors use a new dataset from 50 CMLs and apply an own data processing as well as an adjusted IDW method for spatial interpolation. They analyse the CML-derived rainfall estimates in relation to reference data derived from gauges. They also set up a hydrological model for the target region, force it with CML and/or gauge data and analyse the simulated streamflow in comparison to observations at the catchment outlet. They derive interesting findings, e.g. that the low sensitivity to rainfall of some CMLs limits the performance of the spatial CML-derived rainfall estimates. There are, however, several points that need detailed clarification and maybe even adjustments concerning the processing and the analysis. The manuscript is well structured, mostly well written and the figures are of high quality. Both the writing and the figures will need some improvements, though. Hence, in summary, I recommend a major revision.

Major comments:
---

1. Lack of validation of new CML processing: The authors introduce a new CML data set of 15-minute min-max data. They apply and own processing which is described in an already published conference paper (Nebuloni et al, 2020b). The referenced paper is not very detailed in describing the method and in analysing the methods performance. It also uses a CML dataset from a different region. Hence, in this manuscript more details should be provided on the actual derivation of rain rates from the CML raw data and on analysing the resulting performance. The presented analysis only shows aggregated results, e.g. in Fig 5. This is not enough to understand how individual CMLs perform and if specific CMLs impact the spatial rainfall estimates in the surrounding negatively because of challenging signal fluctuations in the raw data. Showing individual time series of CML raw data together with derived rain rates and data from nearby gauges (or from a spatially interpolated gauge product) would help. Better would be a more quantitative analysis.

More information about the pure rainfall estimates at the individual CMLs is important because errors there will propagate through the whole analysis. The fact that insensitive CMLs perform badly during low rain rates might not be the only reason for degraded performance of CML-rainfall in the hydrological simulation.

2. Lack of validation of modified IDW method: The authors introduce an adjustment factor for the IDW interpolation method, namely an additional weight that decreases with increasing length of a CML. The reasoning is that the localisation of rainfall information is worse for longer CMLs because of inhomogeneity of rainfall along the path and because, for IDW, the rainfall information is represented by a virtual gauge in the centre of the CML path.
2a. While this is true, there are also drawbacks to this approach. Longer CMLs have higher sensitivity to rainfall (as shown in this manuscript) and hence can be more accurate at lower rain rates than short CMLs with the same frequency. Furthermore the orientation of the CMLs in relation to their direction towards an HRU centroid should have an effect. A long CML that points directly towards a HRU centroid should have a higher weight there because it observes rainfall closer to this centroid than a perpendicularly oriented CML would. This does not mean that the proposed modified IDW method does not provide an advantage. But this advantage should be shown empirically, at least by providing individual examples of the modified IDW vs the standard IDW.
2b. After thinking a bit more about the modified IDW, I want to add the following comment. If a CML's length is accounted for in IDW by assigning length-dependent weights, shouldn't there be a distance dependent adjustment of this length-dependent weight? The further away a target point (to be interpolated) is from a CML, the less the length of the CMLs plays a role. I can see the reasoning of the length-dependent weight in close vicinity of the HRU centroids because longer CMLs tend to observe rainfall also outside the HRU, but if the CML density inside the HRU is low or zero the length-dependent weight might not be optimal. A long close-by CML could be "overruled" by shorter CMLs which are further away. Please comment.

3. Overlap of calibration period with analysed rain events: From how I understand the section about model calibration (starting in L233) the 12 analysed rain events are within the calibration period or within the validation period and hence are not independent from the calibration procedure. Since it is not exactly clear how the calibration was performed, I cannot comment on how problematic this is. But the authors should comment on that.

4. Recalibration of hydrological model only done for CMLs: The idea to recalibrate the hydrological model for the CML-derived input is good and important because it is hard to beat the original forcing which was used for calibration. However, the CML-derived rainfall forcing then has an unfair advantage because its setup is calibrated only to the 12 days used for validation while the original setup was calibrated with a much longer period. Wouldn't it be more fair if there would be a recalibrated setup for the 12 days with the gauge-derived forcing for the analysis shown in Fig 12? One can only speculate if and how much this would improve the performance of the gauge-derived forcing. Please justify or reconsider your approach.

Specific comments:

L13: I find the formulation "may lead to benefit in hydrological modelling" a bit vague for the last sentence of the abstract. It should be made clearer what these benefits (better understanding of hydro processes, better streamflow forecasts, etc) are and why they could be expected from the finding that "CML-driven outputs performances are comparable with RG-driven ones" which only indicates equal performance of CMLs and gauges (which is still an important finding, given the dense rain gauge network)

L23: I do not understand the second half of the sentence starting at "..., even if..."

L32: It should be noted here that the temporal aggregation time of the rainfall data is also an important factor. The longer the aggregation time the less critical the rain gauge density is.

L35 and following: This short review of Radar QPE could mention some publications on recent progress with dual-pol radars, e.g. Zhang et al 2020 (https://doi.org/10.1175/JHM-D-19-0194.1), Chen et al 2021 (https://doi.org/10.1175/JHM-D-20-0299.1) or this overview from Zhang et al 2019 (https://doi.org/10.1007/s00376-019-8172-4).

L46: The usage of microwave links for rainfall estimation was proposed earlier, e.g. by Atlas and Ulbrich 1977 (https://doi.org/10.1175/1520-0450(1977)016%3C1322:PAAIRM%3E2.0.CO;2). Giuli et al maybe where the first to propose a a mesh of link for tomographic reconstruction.

Fig 2 and related text: How were the CML paths treated in the calculation if a CMLs crosses several HRUs and how was the distance to the HRU centroid calculated for CMLs? In general, how was the mean distance of the sensors to the HRU centroid calculated? Are all sensors taken into account or only the closest ones?

Table 1: Would be good to specify what the averaging period for the max. rain rates are because only then one can interpret them. I assume these are 10-minute maxima from the raw data, but it should be stated here.

L145: Why did you chose to fill missing RG values with spatially interpolated data? Why not leave the gaps and treat missing values accordingly when the rainfall at the HRU centroid is derived?

L152: The frequency separation of CMLs is small, typically around 1 GHz for CMLs operating below 40 GHz. I do not agree with the statement that this is "adding a certain

degree of redundancy when it comes to rainfall estimates". Or maybe I do not understand what is meant here. Please explain.

L155: Nebuloni et al., 2020b does not provide much detail of the proposed processing method and in particular no quantitative analysis of the resulting CML rain rates, e.g. scatter plots, Pearson correlation, RMSE, bias, false positive rate, etc.. Because the processing of the CML data can have a large impact on the derived QPE it would be important to provide more information on the CML-derived rainfall data here. The choice not the use the available RAINLINK methods should also be explained.

L159: Nebuloni et al. 2020b explains that no wet antenna compensation was applied (end of section 3). What was done for the analysis in this manuscript?

L162: I do not understand why the RSL time series is considered in Watt here to point out to the +-12% uncertainty range in relation to the measured value. All relevant calculations to derive rain rates from raw RSL data are carried out in dB and the uncertainty in dB directly translates to the uncertainty in rain rate.

L167: Can you explain how these numbers were calculated?

L178: How much did the derived k-R parameters differ from the ITU recommendations? And how relevant is this difference in comparison to the uncertainty from quantization and wet antenna effect?

L181: It is not clear if and how the 10-second data was used in this study and why it was required to estimate the average rainfall from the min-max data. Several studies from Israel and the Netherlands have done rainfall estimation from 15-minute min-max data, e.g. by using a calibrated weighting factor of the min and max values. Please elaborate on this.

L187: Why was this simpler approach chosen? The fairly dense CMLs network would presumably provide a good basis for algorithms that derive uneven distributes of rainfall along the CMLs by taking into account measurements from nearby CMLs. I am not saying that simple approaches provide inferior results, but it would be interesting to know the reasoning behind this choice.

L189: I cannot comment on the scientific soundness of the hydrological model, but I would like to understand the choice of this particular model. Was this model applied already in the region, is it applied frequently by the authors or regional water authorities, is it developed by the authors or collaborators?

L232 and following: I am surprised that many of the 12 selected rain events are within the calibration period, namely the year 2019. The other selected rain events are within the validation period of the parameter calibration in the year 2020. Doesn't that carry the risk of overfitting the model on these rain events, assuming that the selected ones are the most prominent ones in these years?

L242: What does "trial and error calibration" mean here? From the sentences above I thought I understood that all the listed parameters and their combinations have been used to run the model. What other "optimum combination" is there that is not covered by sweeping over the defined parameter range?

L245: Was there any information available for how the outflow of the dam was regulated? If not, and if the model is not able to account for that, how much sense does it makes to study events with high flow which might be severely affected by the dam outflow?

L272: How was the selection of the parameter beta done?

L287: How are "low rain rates" and high ones defined?

L288: Since HRU 8, 2, 9 and 4 do not show an overestimation of CML rainfall compared to the gauge-derived data, I do not agree with the conclusion that "CMLs tend to return higher estimates... during high rain rate events...".

Fig 5:: It is not clear to me why there is only one marker type, i.e. either high or low rain rates per HRU. I expected that the event rainfall accumulation is done two times for each HRU, once for events (or maybe even hours) where a certain threshold rainfall rate is exceeded and once for the events (hours) where it is not exceeded. Please clarify how the split into high and low rain rates was done. In case more data points will be added, the plot could be split up into two subplots, one for high and one for low rainfall rates.

Fig 6: I do not understand (but maybe that is my fault) why there are no negative values smaller than -1. If the CML rainfall at a HRU centroid yields 0 because of a false-negatives, then Delta_E should be -1 times the RG value, shouldn't it?

Fig 6: Maybe a 2D histogram, e.g. a hexbin plot, would be easier to interpret than the current figure. The distrbution of points can also be conveyed in a 2D histogram and log-log scales can also be used.

Fig 6: Besides the negative values of the mean of Delta_E for small rain rates, I find it worth mentioning that the spread of Delta_E in the positive range increases significantly for small rain rates. Is this the effect of false-positive CML rain rates?

Fig 8: Since the distinction between high and low rain rates is made in most other plots, it would also be valuable to show it here, e.g. by having two differently coloured box plots for each HRU.

L313: It is an interesting finding that the decreased performance can be attributed to the low sensitivity of individual CMLs. From Fig 3. we know the distribution of sensitivities, but not where the individual CMLs are located. A map of the CMLs coloured by sensitivity could help with the interpretation. But Fig 1 most probably would get too busy with more colours.

L315: It would be good to show these direct comparisons of CMLs and closest rain gauge so that the reader can judge himself or herself how the different CMLs behave and perform. Maybe these could be added to the Appendix. See also my main comments.

L341: It would be better to write "...Fig 11b shows an example for which the CML-driven simulation better represents..." because otherwise this reads like a general statement that CML-driven simulations perform better.

L351: "the major drawback of the present work is definitely that we did not rely on a large and real-time CML-based dataset" Why didn't you use longer periods of CML data?

L355: To have a fair competition between RG and CML, you would have to recalibrate the RG-driven setup also only using the 12 selected events. By calibrating the CML-driven simulation exactly to the event that you analyse you might given them a significant advantage. (See also my major comment)

L364 Discussion section: This section reads more like a summary of the results, not like a discussion. Potential causes and consequences of the results, as well as limitations of the chosen approach should be discussed here. I suggest to add some subsections to structure that.

L423 and following: Regarding the first three points mentioned here it shall be noted that continuously operating CML real-time data collection and processing systems exist or are in a final implementation phase in Sweden (https://www.smhi.se/en/services/professional-services/memo-microwave-based-environmental-monitoring/), Czech Republic (http://www.tel4rain.cz/) and Germany

(https://amt.copernicus.org/articles/9/991/2016/). Furthermore I do not agree with point 4. Why should "heavy data reduction" be require? Data storage, transfer and processing of TBs of data is not a problem with today's computer resources, in particular at met services or research institutions.

Technical corrections:
---

general: Check the usage of "the", which has to be removed in many cases throughout the manuscript, e.g. in "...which may be tackled through the integration of the conventional sensors" the second "the" has to be removed because this is not a reference to specific conventional sensors here.

L3: better write „of the spatial distribution of rainfall" instead of „of the rainfall spatial distribution"

L5: Remove comma after "network"

L10: should be "...it is shown..."

L12: Maybe better write "model" instead of "outputs"

L15 + L17: My subjective opinion is that the two references here are not needed to back up the common and general statements in this sentence.

L24: I would remove "the" before "disdrometers" (but I am not 100% sure)

L28: write "to ungauged sites" (or maybe I do no correctly understand what you want to express here)

L44: "... is rising a great interest nowadays" too colloquial

L61: Write "The authors..."

L61: Typo in "thoughout"

L66: "we figured out" is too colloquial

L85: I would say that the catchment cannot be the case study as in "The case study is the Lambro catchment, ...", I would rather write something like "The case study was carried out in the Lambor catchment, ...

L86: I count four and not three difference provinces here

L86: Write "The Lambro river...."

L109: Write "As shown..."

L111: should it read "adopted here" instead of "here adopted"?

L138: I do not understand what "data reduction" mean with regard to CML data processing.

L164: "descends" is probably the wrong verb here

Note that I did a less detailed check for technical errors from here on

L251: Better write "compensate" instead of "dump"

L271: Leave out the "is" in "the longer is the CML"

L288: "On the other side..." should be "On the other hand". However, typically you should write "On the one hand" somewhere earlier in the paragraph before writing "on the other hand". Maybe another formulation would thus fit better here.

L309: "It does not emerge..." is not correct. Maybe write something like "There is no general positive or negative trend visible...".

L314: better write "these two CMLs"

L320: "formulations" is not the correct term, better write "equations"

L350: "it can be easily guessed" is not a good formulation

L367: "event" instead of "event’s"

L373: should be "since they are covered"

L377: "sensor density" instead of "sensors’ density"

L408: "This is..."

L408: The logic of the argument in this sentence is not clear.