Reviewer #1 (Nicolas Velasquez)

The authors present a comprehensive comparison of multiple rainfall products against a collection of four manual rain gauges. The comparison helps to identify the quality of the sources potentially useful for early warning implementations. The authors focus their work on comparing point data with pixel data, which has been widely accepted by the community. However, point-based comparisons assume point observations as the ground truth, and they are, but for a limited area of space. On the other end, a comparison with more significance will integrate streamflow observations with rainfall accumulations at the watershed scale. At the end of the day, the rainfall totals and localization controls most of the watershed response during a flood event, not the rainfall that happened at a point where we were technically able to measure it.

> Clarification: not "against a collection of four manual rain gauges", but 112 gauges (see line 310 and Fig. 4). Four selected stations were only chosen to present more detailed analyses in Section 4.5 (Table 6 and Figure 10).
>
> As for the verification of precipitation based on catchment outflow: our aim was to assess the reliability of the precipitation available online by comparing it with the most reliable data, i.e. primarily from manual rain gauges.

Rainfall comparisons are always important for the community, and therefore, this work could be of significance. However, it needs to address some major issues. Following, I present my general and specific comments.

General comments

The abstract does not provide a description of the techniques used by the authors to make the comparisons. Also, it does not clarify why it is important to have a precise estimation of the rainfall. Finally, they don't relate their results to a significant outcome.

> We have changed the abstract:
>
> "**Abstract**. A huge and dangerous flood occurred in September 2024 in the upper and middle Odra river basin, including mountainous areas in south-western Poland. The event provided an opportunity to investigate the feasibility of reliable estimation of high-resolution precipitation field, which is crucial for effective flood protection. Different measurement techniques were analysed: rain gauge data, weather radar-based, satellite-based, non-conventional (CML-based) and multi-source estimates. Apart from real-time and near real-time data, later available reanalyses based on satellite information (IMERG, PDIR-Now) and numerical mesoscale model simulations (ERA5, WRF) were also examined. Reference data used to verify the reliability of the different techniques for measurement and estimation of precipitation included observations from manual rain gauges and multi-source estimates from the RainGRS system developed at IMGW for daily and hourly accumulations, respectively. Statistical analyses and visual comparisons were carried out. Among the data available in real time the best results were found for rain gauge measurements, radar data adjusted to rain gauges, and RainGRS estimates. Fairly good reliability was achieved by non-conventional CML-based measurements. In terms of offline reanalyses, mesoscale model simulations also demonstrated reasonably good agreement with reference precipitation, while poorer results were obtained by all satellite-based estimates except the IMERG."

In section 2.1, the authors mention a collection of studies done before on flooding at the Odra River. However, they don't mention the main conclusions of these studies and how they relate to their current work. This must be addressed.

> We have completed section 2.1 with the following paragraph inserted at its end:

“The above studies indicate that the upper Odra River basin is highly vulnerable to flooding caused by intense precipitation in the mountainous part of the basin. This is also influenced by the shape of the river network, which favours the cumulation of floods from individual tributaries. The flood risk there occurs almost annually during the summer.”

The main comparison happens at four manual gauges located over the mountains in the south- west area of the watershed. Nevertheless, mountains tend to generate large disturbances over the rainfall patterns, probably biasing the results. Also, I am missing a figure in Section 2 presenting the rainfall accumulation of the event. Despite magnitude inaccuracies, a total accumulation figure will illustrate the area of interest for this event. It will also tell if the use of these gauges is worth it or not. You can make that figure by accumulating the rainfall from Figure 5.

As we wrote above, the main reference data were not from 4, but from 112 manual rain gauges (see lines 310-311).

In terms of rainfall accumulation of the event you are right: we have made a new figure in section 2.2 with the 4-day rain gauge accumulation from the whole flood period and have inserted it in line 191:

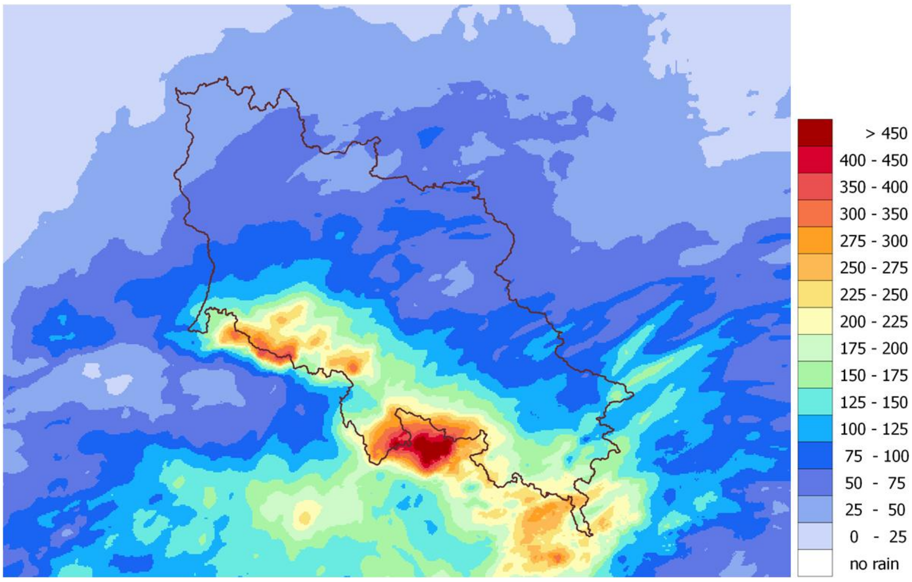

Figure 2: Field of precipitation accumulation during the flood of 13-16 September 2024 (four days) for the upper and middle Odra River basin in Poland, obtained from the RainGRS Clim reanalysis (GRS Clim).

It is troublesome to compare 1km or 0.5km rainfall products with point data. More in a mountainous region where convective rainfall was probably dominant. Point data represents a relatively insignificant portion of a $1km^2$ area, and therefore, it may be highly biased. Instead, the watershed should be used to compare rainfall totals. Also, it is at the watershed scale when this comparison matters for flood forecasting and assessment purposes.

Our aim was to assess the reliability of online precipitation by comparing it with the most reliable data, i.e. measurements by manual rain gauges. In the case of the floods described in our article, we mainly observed widespread precipitation, which generally creates smaller problems when comparing with grid data.

It is hard to conclude on the performance of a rainfall product based on the comparison of a single event. Conditions such as the season and the covered area are likely to induce a change in the performance. Additionally, the gauged records cover only a fraction of the watershed. The authors should consider expanding their analysis to more cases in the watershed.

We completely agree with the Reviewer that the detailed comparative analysis of the different precipitation measurement techniques does not allow us to draw general conclusions about their effectiveness. This would require an analysis of a really large number of events, at different seasons, in areas of different character and orography, with different types of precipitation, etc. Our aim was only to examine the effectiveness of high-resolution estimation of the precipitation field during an extreme precipitation event that happened in a specific basin. We believe it is worth showing how the techniques used worked in this specific situation.

Moreover, we do not have the possibility to carry out analogous analyses for other events. During previous major floods (2010, 1997), IMGW did not have such an extensive database of online measurements.

The English needs some editing.

We have revised the text in terms of language.

Specific comments

Lines 45-47: The authors talk about manual rain gauge measurements. At this point, they should be talking about automation. Manuals are fine for citizen science and community-based projects. However, these gauges have significant limitations in the temporal scale, especially when it comes to flood forecasting in mountainous regions.

The reviewer is of course right that the temporal resolution of manually measured data is low. However, they are still the most reliable precipitation measurements available, but given their limitations in terms of time scale, we used multi-source estimates based on data from automatic rain gauges, radar and satellites as an auxiliary reference for 1-h analyses (Tables 3 and 5). These data proved to be the most reliable when verifying daily accumulations based on manual rain gauges (Table 2).

Lines 48-50: There is an extensive literature on this topic. However, the authors don't mention it.

We have added literature references in line 51:

Hohmann, C., Kirchengast, G., O, S., Rieger, W., and Foelsche, U.: Small catchment runoff sensitivity to station density and spatial interpolation: hydrological modeling of heavy rainfall using a dense rain gauge network, Water, 13, 1381, doi:10.3390/w13101381, 2021.

Loritz, R., Hrachowitz, M., Neuper, M., and Zehe, E.: The role and value of distributed precipitation data in hydrological models, Hydrology and Earth System Sciences, 25, 147–167, doi:10.5194/hess-25-147-2021, 2021.

Lines 67-73: Same, I am missing literature on this matter here.

We have added a link in line 72 to the already existing reference: Ośródka and Szturc (2022) and this new reference:

Méri, L., Gaál, L., Bartok, J., Gažák, M., Gera, M., Jurašek, M., and Kelemen, M.: Improved radar composites and enhanced value of meteorological radar data using different quality indices, Sustainability, 13, 5285, https://doi.org/10.3390/su13095285, 2021.

Lines 93-104: Better referencing. However, the authors failed to mention some of the most important operational products on this matter, such as the Multi-Radar Multi-Sensor (MRMS) product provided by NOAA.

We have added a reference to MRMS in line 104:

"NOAA provides the Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation product automatically generated through integration of data from radar networks, surface and satellite observations, numerical weather prediction (NWP) models, and climatology (Zhang et al., 2016)."

Zhang, J., Howard, K., Langston, C., Kaney, B., Qi, Y., Tang, L., Grams, H., Wang, Y., Cocks, S., Martinaitis, S., Arthur, A., Cooper, K., Brogden, J., and Kitzmiller, D.: Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. Bulletin of the American Meteorological Society, 97, 621–638. https://doi.org/10.1175/BAMS-D-14-00174.1, 2016.

Lines 127-130: It is still not clear to me how a manual rain gauge is able to measure the details of a storm. For my understanding, a manual rain gauge refers to a gauge that requires a human to read it (like an accumulation bucket). Maybe I am wrong, but I know that this can also confuse other people who work in this area.

The manual rain gauge obviously does not measure the details of the temporal precipitation distribution during a storm, but the daily accumulation are the most accurate of all techniques, so we used the GAU Manual data to verify daily accumulations. Of course, these measurements do not allow analysis of rainfall on shorter time scales, so we used multi-source GRS estimates to verify hourly accumulations.

Lines 137-139: From my experience, precipitation fields of 1km every 10 minutes or every hour can still be too coarse for flash flood forecasting in mountainous regions. In many of these rivers, the response time is in the order of minutes, and the variability of the convective systems is in the order of m^2 . The authors should contrast their statement with the literature on floods in mountainous regions. In your case, we are talking of a large watershed (44,000 km^2) and therefore, the scale and the time frame are ok. However, you should state that in your introduction; otherwise, readers may get confused.

We have revised the paragraph in lines 137-142:

"The main objective of this work is to examine the real possibilities of precise estimation of a precipitation field with a high spatial resolution of about 1 km and a high temporal resolution of at least 10 min, or one hour during intense precipitation events that cause floods in upper Odra River catchment area in September 2024. All available real-time and offline measurements and estimates were verified to determine their applicability and to quantify their reliability."

Line 167: The reference has an issue.

We have corrected the bibliographic description in References in lines 934-938:

"Ligenza, P., Tokarczyk, T. and Adynkiewicz-Piragas, M. (Eds.): Przebieg i skutki wybranych powodzi w dorzeczu Odry od XIX wieku do czasów współczesnych (The course and impacts of selected floods in the Oder river basin from the XIXth century to the present day), Instytut Meteorologii i Gospodarki Wodnej − Państwowy Instytut Badawczy, Warszawa, 2021, 132 pp., ISBN: 978-83-64979-45-3, 2021 (in Polish)."

and a reference in the text to this item in line 167: „*Przebieg*..., 2021" to: „Ligenza et al., 2021".

Line 176: "and other works by this author" change that and add the proper citations. They could be important for some readers.

We have changed the citation „Kundzewicz, 2014" into:

„e.g. Kundzewicz, 2014"

Lines 180-181: revise the phrase that is between these lines. It is hard to understand.

We have changed the sequence into:

> "Research suggests that climate change affects the frequency and severity of floods, leading to an increased risk of flooding (e.g. Kundzewicz et al., 2023)."

Line 187: Change "per" for "in"

We have changed.

Lines 192-193: Add the reference after the 400 mm.

We have removed the reference to Kimutai et al. (2024) because we cited it not because of the amount of rainfall, but because of the use of a 4-day total.

Table 1: The authors are mixing units of km and degrees. They should transform the degrees according to the watershed projection. They only provide the transformation for ERA5. It is hard to make a fair comparison if they don't use the same units. You show this information in line 261. Present it on your table.

Thanks for this comment! We have supplemented the information in Table 1 with the pixel sizes in km:

- rows SAT and H61B: "Roughly 5-6 km for Poland" → "Roughly 3.5 km x 6.0 km*"
- row IMERG: "0.1º x 0.1º" → "Roughly 7 km x 11 km* (0.1º x 0.1º)"
- row PDIR-Now: "0.04º x 0.04º" → "Roughly 2.8 km x 4.5 km* (0.04º x 0.04º)"
- row ERA5: "0.25º x 0.25º" → "Roughly 18 km x 28 km* (0.25º x 0.25º)"
- Footnote: "* Roughly 18 km x 26 km." → "* In the area of the study basin."

Figure 2: Should present the radar beam radius (50, 100, or 150 km?). This is also important when contrasting rainfall estimates from radar data. Also, if possible, the figure should present the blind spots for the radars due to the mountains.

We have added circles with a radius of 150 km around the radars to Fig. 2, because our experience with radar-based precipitation shows that up to such a distance the data retain the highest reliability, if, of course, there are no terrain obstacles in the path of the beam. There are mountains in the south-eastern part of the study catchment, but these are surrounded by as many as eight radars. Thus, there are no places in the study area where the lowest radar beam is too high above ground. It can therefore be assumed that there are no places in the area for which high quality radar data would not be available.

Figures 3 and 4 look redundant. Their information can be condensed in Figure 2.

Following the Reviewer's suggestion, we have combined the former Figs. 3 and 4: in Fig. 3 there are all measurements in near real time: automatic (telemetric) rain gauges, weather radars, and CMLs, and in Fig. 4 there are only manual gauges (we have only slightly changed the station symbols).

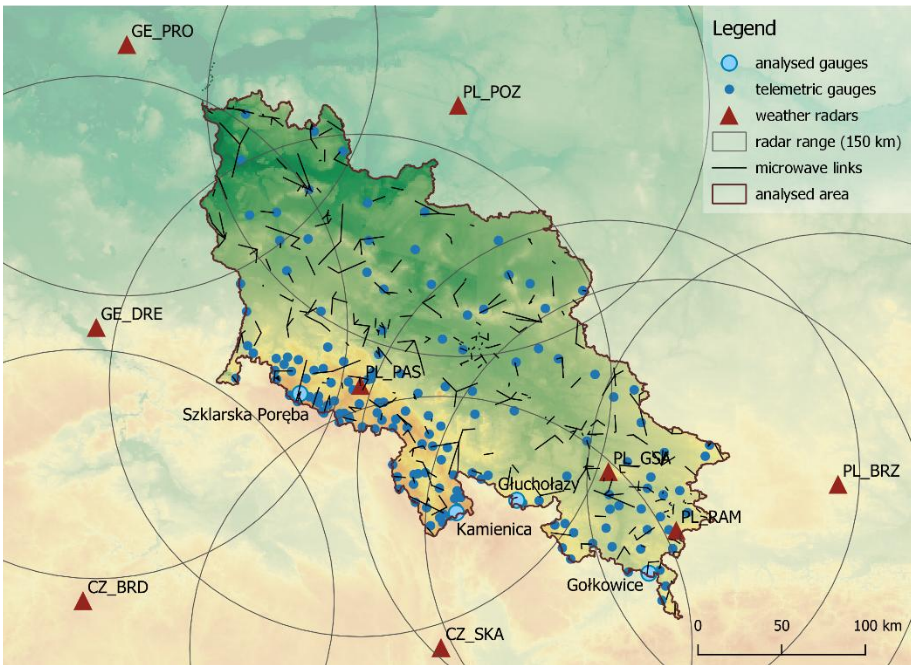

Figure 3: Locations of measurement stations in the upper and middle Odra River basin: telemetric rain gauges (blue dots), weather radars (brown triangles) with 150-km range (brown circles), commercial microwave links (black lines), and four manual rain gauges selected for more detailed analysis (larger blue dots).

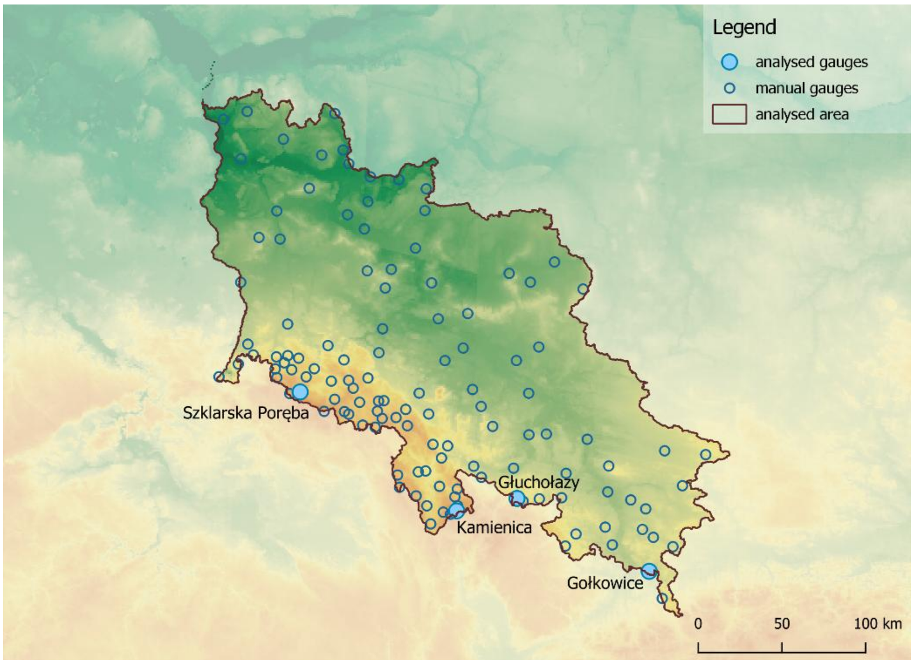

Figure 4: Locations of manual rain gauges (blue circles) and four ones selected for more detailed analysis (larger blue dots) in the upper and middle Odra River basin.

Lines 504-505: If I understand well, the comparison corresponds to days where only 50mm or more was observed by the manual gauges. If this is correct, explain it better.

We have changed this sentence to:

"The results of the statistical analysis based on daily accumulations from manual rain gauge measurements (GAU manual) for days with recorded rainfall of 50 mm or more are presented in Table 4."

Lines 530-532: Why 5mm? why 200 pixels?

These values were chosen empirically in such a way as not to take 1-hour totals with low precipitation into the statistics, as they introduce randomness into the results, especially as regards the CC.

Lines 534-536: The argument lacks scientific support.

We have removed the last sentence from this paragraph (lines 534-536) and added a new one at line 532:

„Thresholds of 5-mm for hourly accumulations and 200 pixels for the area where such precipitation occurred (approximately 0.5% of the entire catchment) were introduced to exclude data with low precipitation from the statistics."

Tables 2 to 5: These tables look repetitive and hard to read. A reader probably won't take an important message from them. Instead, a collection of figures (one next to the other) presenting the comparisons done under the different conditions should be more illustrative. Ideas for such a collection include, but are not limited to:

- Scatter plots comparing the correlation vs the RMSE for each case. One case next to the other.
- Scatter plots comparing the metric for the daily case (x axis) and the other cases (y axis).

We made graphs of CC depending on RMSE and inserted them as Figs. 8, 9, 10, and 12. They illustrate the results from Tables 2, 3, 4 and 5, respectively.

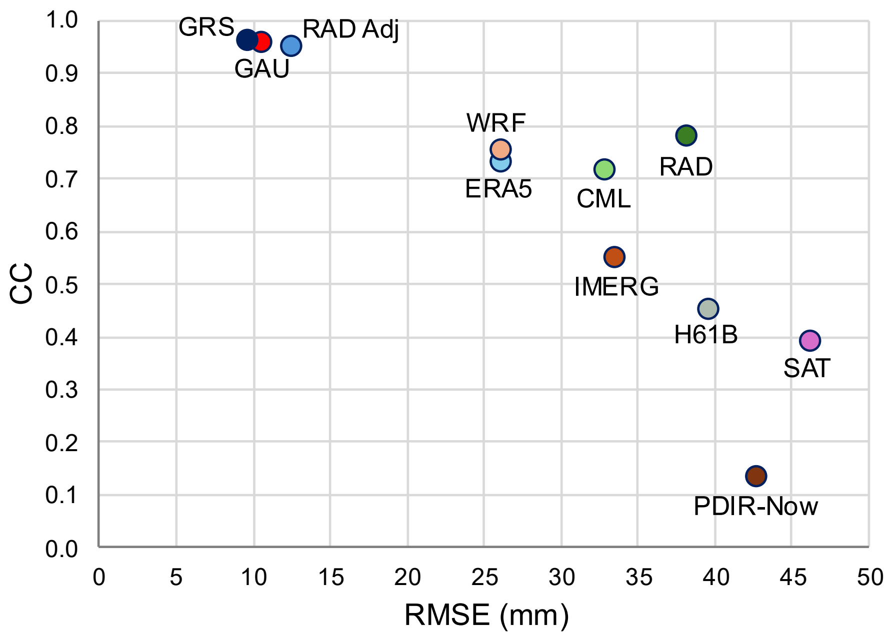

Figure 8: Scatter plot comparing CC vs RMSE for each measurement and estimation technique for daily precipitation accumulations from 13-16 September 2024, against data from manual rain gauges (GAU Manual) as reference.

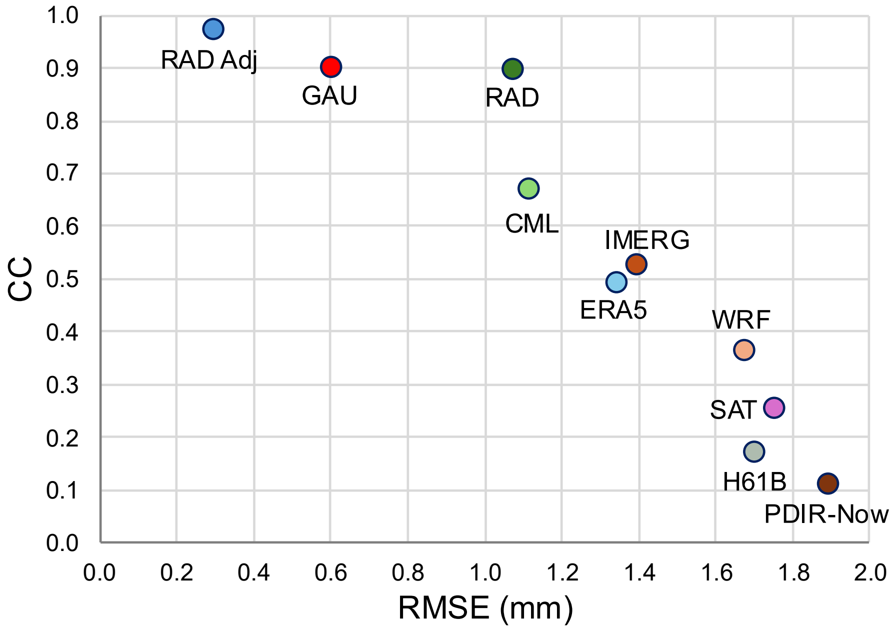

Figure 9: Scatter plot comparing CC vs RMSE for each measurement and estimation technique for hourly precipitation accumulations from 13-16 September 2024, against the RainGRS estimates (GRS) as reference.

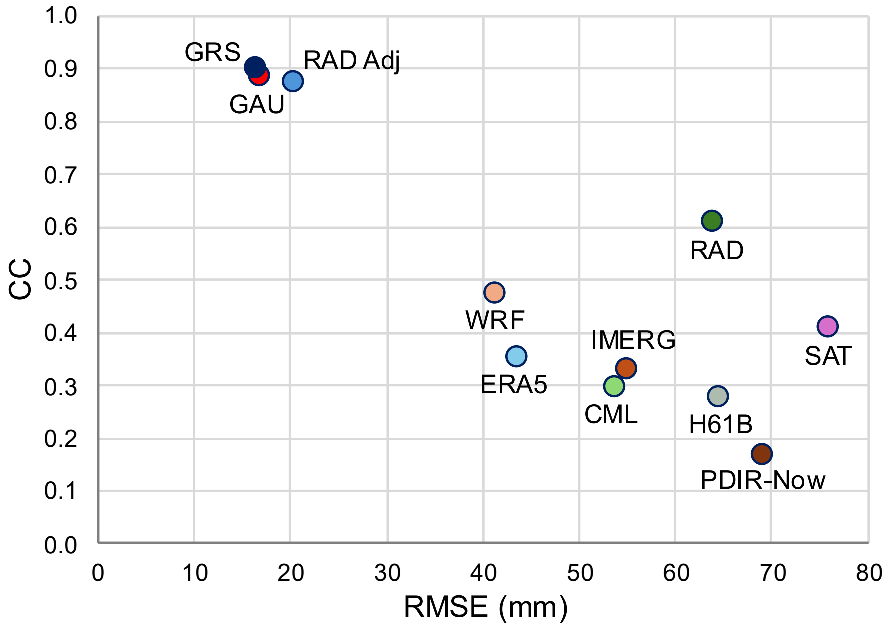

Figure 10: Scatter plot comparing CC vs RMSE for each measurement and estimation technique for daily precipitation accumulations from 13-16 September 2024 against data from manual rain gauges (GAU Manual) as a reference with a threshold of 50 mm.

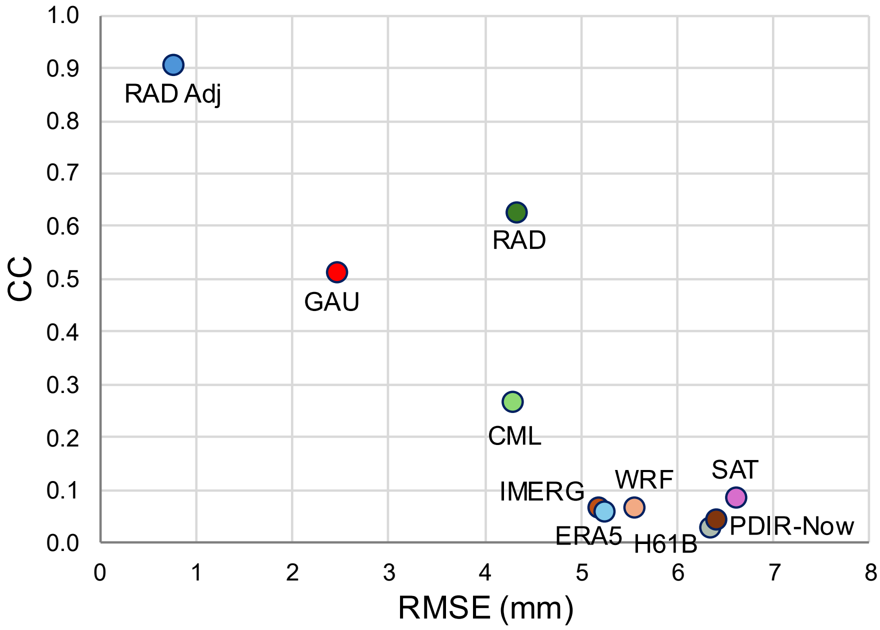

Figure 12: Scatter plot comparing CC vs RMSE for each measurement and estimation technique for hourly precipitation accumulations from 13-16 September 2024 against the RainGRS estimates (GRS) as a reference with a threshold of 5 mm.

Table 6: More boring tables that most likely nobody will read. This large collection of numbers in the middle of a paper is not telling a story. Instead, it makes the manuscript look like a consulting report. Consider other ways of showing the differences between the hietograms. The table can go to an appendix section.

We have moved Table 6 to Appendix A.

Lines 627-628: This is an expected outcome.

Yes, this is a rather expected observation, but we think it is worth emphasising here.

Section 4.6: Can be significantly improved if the authors include a schematic summarizing their results.

We have introduced diagrams 8, 9, 10, and 12, which intuitively illustrate the results obtained. We therefore believe that it is no longer necessary to do an additional schematic summarizing of the results.