How uncertain are precipitation and peakflow estimates for the July 2021 flooding event?

by

Mohamed Saadi, Carina Furusho-Percot, Alexandre Belleflamme, Ju-Yu Chen, Silke Trömel, and Stefan Kollet

Submitted to *Natural Hazards and Earth System Sciences*

Manuscript ID: NHESS-2022-111

Answer to Anonymous Referee #2

1 Summary of main changes

We gratefully acknowledge the valuable suggestions made by the Anonymous Referee #2 and we would like to thank them for their time and evaluation. We addressed all comments in detail (Section 2). In summary, the major changes in the data and methods applied in the study are:

- We included new radar-based, quantitative precipitation estimates (QPE) that better account for the vertical gradients of radar variables (and hence of precipitation rates). Compared to state-of-the-art QPE products (Chen et al., 2021), these new products (with VPC in their names, for Vertical Profile Correction) exploit measurements of Micro Rain Radars (MRR) that helped characterize the precipitation rates below the height monitored by the C-band radars of the DWD (*Deutscher Wetterdienst,* German Weather Service). In addition, a vertical profile correction was applied to horizontal reflectivity Z and specific differential phase K_{DP} following an approach by Chen et al. (2020). These new products significantly improved the radar-based QPE with respect to estimates from rain gauges.
- We removed the QPE product based on specific attenuation at vertical polarization (A_V) and K_{DP} (RAVKDP in the original manuscript) as it yielded similar results to RAHKDP, the one based on specific attenuation at horizontal polarization (A_H). Hence, the number of radar-based QPE products is now RADOLAN + six other products (RZ, RZKDP and RAKDP, in addition to the version with corrected vertical profiles RZ-VPC, RZKDP-VPC, and RAKDP-VPC)
- We added a new simulation of ParFlowCLM with distributed Manning's coefficient assigned based on land cover.

The conclusions of the paper have slightly changed. Namely, the new products with vertical profile correction improved the estimates of event precipitation with respect to rain gauges. The point-scale evaluation and catchment-scale evaluation led to similar ranking of the different QPE products with respect to RADOLAN. Finally, the probabilities of exceeding the historical peakflow were highly sensitive to QPE for all catchments.

Below we provide a detailed reply to the comments of Referee #2.

2 Response to comments of Anonymous Referee #2

General comment: *This work aims to investigate the influence of using a set of different radar-based QPE and different hydrological models on the uncertainties in simulating the record-breaking July 2021 flood event in Germany. Given the lack of peak flow information (the flood partly destroyed the monitoring systems), the analysis is focused on the probability that the simulated peakflow exceeds the highest historically observed peakflow before the flood. This is a very interesting point of view, given the challenges offered by the prediction of a record breaking flood to both precipitation estimation and hydrological prediction. The work is appropriate for NHESS and its readership.*
The manuscript is broadly well written and well structured. However, there are some specific issues listed below that should be considered before acceptance.

Comment 1: *"Better identifying the main focus of the work. The July 2021 flood in Germany is not only a record-breaking flood. It is a flood that far exceeded previously observed records (the authors could report existing post flood estimates that shows how far the estimated July 2021 peak exceeded the previous records). Of course, existing methods and models for flood forecasting cannot predict these floods well because flood generation processes of large extremes differ from those of smaller, more frequently observed events. Therefore, research aiming precisely to this issue by considering these kind of megafloods is timely and helpful. However, this point is completely ignored in the abstract, and it is elaborated relatively late in the introduction."*

Authors' response: We agree that identifying the main focus of the work is essential. The first two sentences of the abstract were meant to convey this idea. Following the Referee's suggestion, we reinforced the main idea as follows:

"The disastrous July 2021 flooding events made us question the ability of current hydrometeorological tools in providing timely and reliable flood forecasts for unprecedented events. This is an urgent concern since extreme events are increasing due to global warming, and existing methods are usually limited to more frequently observed events with usual flood generation processes."

We would like to stress that our aim is not to provide an exhaustive analysis of the flooding event (such in Mohr et al., 2022), but to focus on how precipitation estimates are uncertain for this event, and how this uncertainty in precipitation estimates compares to that of hydrological models to impact peakflow estimates. Since there are no measurements for the event, we proposed to focus on the probability of exceeding the highest measured peakflow, which is itself a novel way of circumventing this problem. The first paragraph of Section 1.3 of the manuscript identifies the focus of the work and its novelties:

"This study investigated the influence of improved QPE and different representations of hydrological processes on the uncertainties in simulating extreme flooding events. The novelties of our study consist in: (1) using new QPE products from vertical-profile corrected, phase-based observables of C-band and X-band radars, (2) contrasting hydrological modeling approaches (conceptual vs. partial differential equations (PDE)-based model), and (3) proposing an evaluation framework of the hydrometeorological prediction chain for unprecedented extreme events with unavailable discharge measurements. Since no peakflow measurements are available (partly due to destroyed monitoring systems), our analysis focused on the probability that the simulated peakflow exceeds the highest historically observed peakflow. This is relevant because hydrological models are often evaluated based on their ability to detect the probability of flows exceeding catchment-specific, critical thresholds for flood warning applications (Anctil and Ramos, 2017)."

Comment 2: *"The point (L205-2010) made on the different results obtained based on considering raingauges and raingauge-based catchment-scale precipitation estimates is someway misleading. First, it totally ignores the uncertainty in the catchment-scale estimates based on raingauges (and here I urge the authors to consider techniques better than Thiessen for this). Second, this conclusion obviously depends on the set of raingauges considered. If the reference raingauges are those considered for estimating the catchment-scale precipitation, I doubt outcomes may be different. By the way, this conclusion is missed in the conclusion section."*

Authors' response: We acknowledge that our choice of Thiessen polygons to compute the catchment-scale precipitation is subjected to uncertainties and the density of the rain gauge network. For this reason, we included the daily estimates from REGNIE (Rauthe et al., 2013), which is a gridded, high-resolution product that accounts for several attributes of rain gauges in the interpolation process. Note that this product covers only 50% of the catchment area for the Rur at Monschau. In the new version of the manuscript, we added the following:

"Acknowledging the uncertainties that may arise from using Thiessen polygons to compute catchment-scale precipitation depths, we compared these to catchment-scale precipitation estimates from the daily gridded product REGNIE (1-km resolution), which accounts for the position, the height, the exposition and the slope of the gauge stations in the interpolation of the precipitation fields from rain gauges (Rauthe et al., 2013)."

Figure R1 (Figure 4 of the revised manuscript) shows that the estimates from rain gauges using Thiessen polygons are similar to REGNIE's, except for the Erft at Bliesheim, where Thiessen method underestimated the total precipitation depth for the 14 July 2021. We can conclude that the Thiessen polygons give reasonable results for our case study.

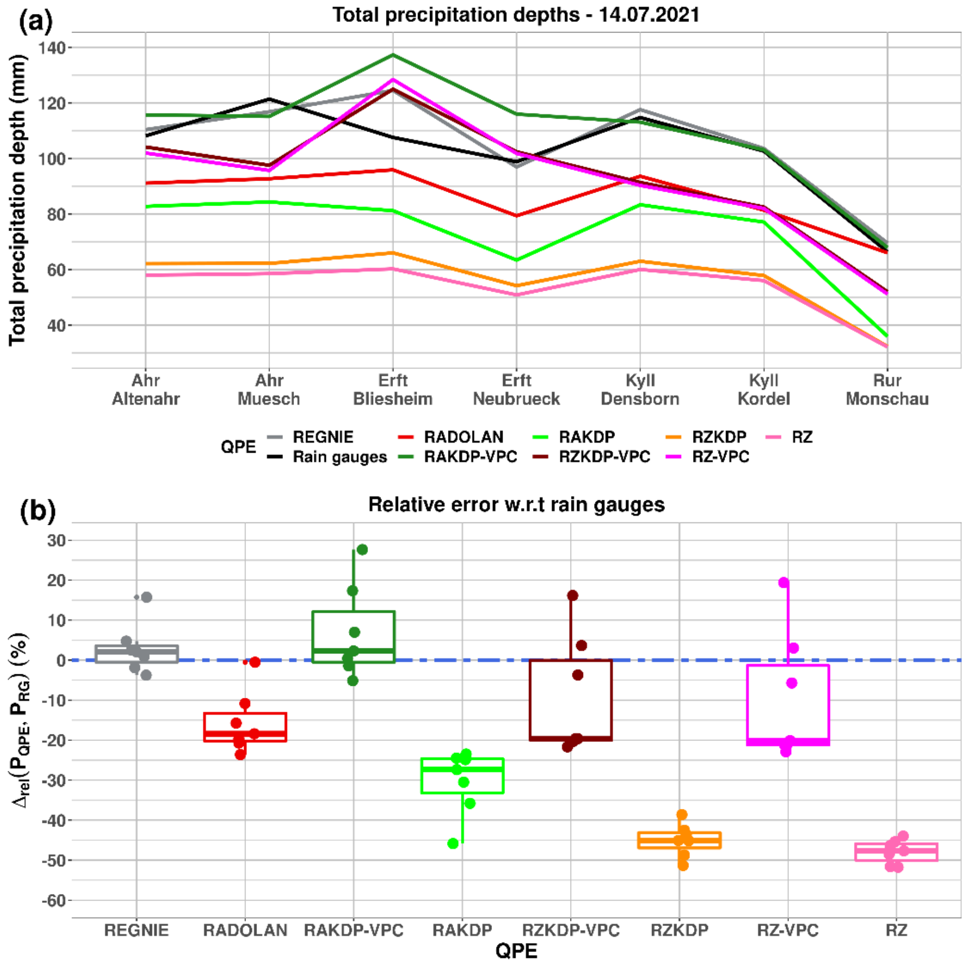

Figure R1 : (a) Total precipitation depths for the 14 July 2021 estimated by rain gauges, REGNIE and radar-based QPE products. (b) Relative errors in REGNIE and radar-based QPE with respect to estimates from rain gauges using Thiessen polygons of the total catchment-scale precipitation depth for the 14 July 2021.

For the second point considering the conclusion in L205-210 of the original manuscript, we agree with the Referee that this depends on the set of rain gauges considered, but also on the spatial variability of precipitation fields. If the network of rain gauges missed the spatial variability, then the catchment-scale evaluation can be strongly different from the point-scale evaluation. In the revised manuscript, considering new precipitation products with correction of vertical profiles, the conclusions at the point-scale and the catchment-scale were quite similar with respect to the ranking of the different radar-based QPE. Therefore, we changed the lines 205-210 of the original manuscript to:

"Conclusions about the agreement between QPE products and rain gauges are similar when we look at the catchment-scale evaluation. Specifically, QPE based on specific attenuation (A) with corrected vertical profiles for K_{DP} (RAKDP-VPC) outperformed RADOLAN in reproducing estimates from rain gauges (using Thiessen polygons) across the seven catchments (Fig. 4), and reduced relative error from a median of -18 % for RADOLAN to +2 %. With the exception of RAKDP-VPC, radar-based QPE products tended to underestimate catchment-scale precipitation with respect to rain gauges in most cases, confirming the point-scale results (see NMB scores in Fig. 3). However, this comparison underlines the fact that the assessment of QPE products is catchment-dependent. RAKDP-VPC outperformed RADOLAN (with respect to rain gauges) for the catchments drained by the Ahr and the Kyll, whereas they both agreed for the Rur at Monschau. For the catchments drained by the Erft, RAKDP-VPC overestimated precipitation depths with respect to rain gauges, whereas RADOLAN underestimated

the total precipitation depth. Finally, using the Thiessen polygon method led to similar catchment-scale precipitation depths compared to the regionalized REGNIE product, except for the Erft at Bliesheim where the Thiessen polygon method underestimated the total precipitation depth with respect to REGNIE."

In the revised manuscript, we stated that both the point-scale and the catchment-scale evaluations led to similar results, i.e. improved precipitation estimates thanks to better characterization of the vertical profile of radar variables:

"Better characterization of the vertical profiles of radar variables led to significant improvements of radar-based QPE for the extreme event of 14 July 2021 with respect to rain gauges. These improvements were confirmed at both the point scale and the catchment scale."

Comment 3: *"The point (L254-256) about the causes leading to the strong underestimation (For the 14 July 2021 event, this underestimation may be explained by intense collision-coalescence processes taking place close to the surface..) lacks any ground. I mean: it is likely that collision-coalescence processes may cause those underestimation, but this attribution needs a far better explanation."*

Authors' response: The use of Micro Rain Radar (MRR) observations showed that all radar variables increased towards the ground which clearly suggests the dominance of collision-coalescence processes below the melting layer for this event (Figure R2a-b). Although collision-coalescence processes alone do not change the precipitation flux within a column, the retrieved rain rate still increases toward the surface (Figure R2c). By examining the contributions of drizzle (D < 0.5 mm) and raindrops (2 mm < D < 4 mm) to the DSDs, the former shows a secondary peak at 1 km height followed by a rapid decrease downwards, while the number of raindrops constantly increases toward the ground below the ML (Figure R3). Accordingly, the increasing rain towards the surface can be explained by the transformation of water vapor into droplets above 1 km height and its transformation into rainwater via warm-rain processes below (Chen et al., Submitted). In addition, accounting for these measurements and correcting the vertical profiles of Z and K_{DP} reduced the errors of the radar-based QPE with respect to rain gauges, as can be seen in Figure R1.

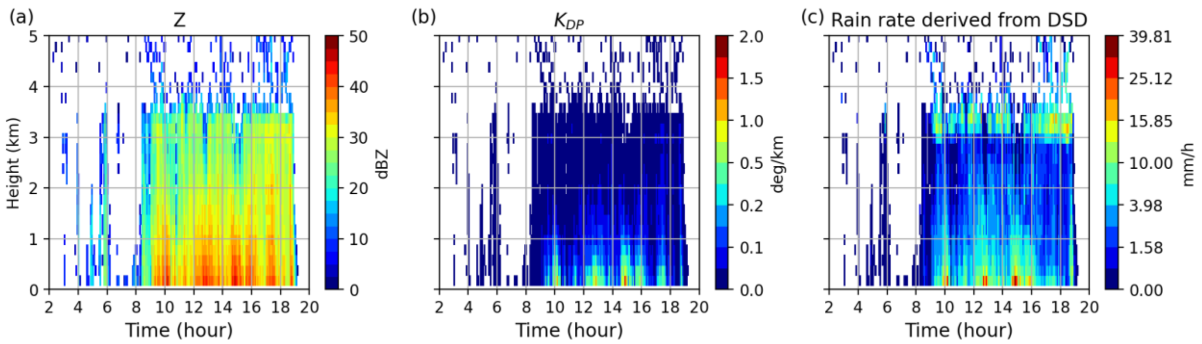

Figure R2: Retrieved profiles based on the MRR measurements from University of Bonn on 14 July 2021, including (a) reflectivity Z, (b) specific differential phase KDP, and (c) rain rates derived from the retrieved raindrop size distributions (DSD).

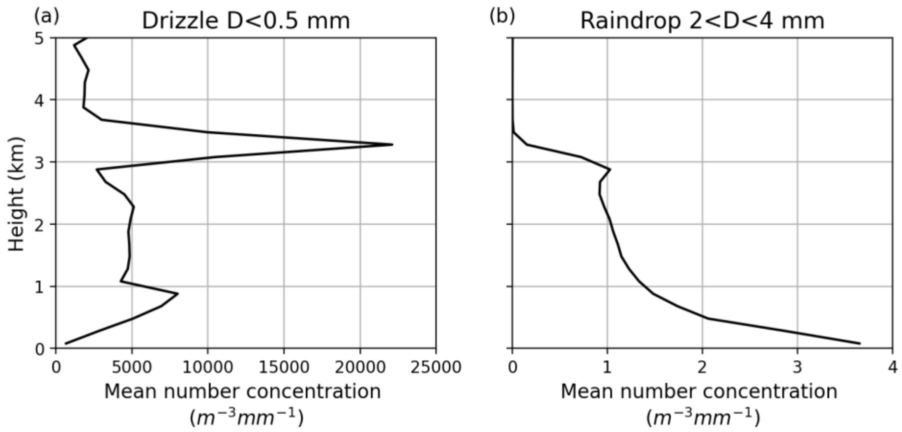

Figure R3: Mean number concentration profiles of (a) drizzle with diameter D < 0.5 mm, and (b) raindrops with 2 mm < D < 4 mm calculated from the DSDs retrieved from the observations of the MRR located at University of Bonn.

Comment 4: *"Information on how antecedent conditions were computed, and about the accuracy of these estimates, is missing, in spite of the critical role this information have on the sensitivity of the model to QPE error."*

Authors' response: Prior to 14 July 2021, both GR4H and ParFlowCLM were run continuously starting from 2006-2007 for all catchments. This allowed for exploiting all the record periods to yield the best estimate of model initial conditions prior to the event. We now mentioned this in the revised manuscript when we present how QPEs are evaluated using hydrological models. The following statement was added to Section 3.4:

"Second, we examined the effect of QPE on the frequency of exceeding the highest historically observed peakflow for each catchment (Table 1) by simulated peakflows. Both GR4H and ParFlowCLM were initialized using a long spin-up period starting from 2006 for GR4H and 2007 for ParFlowCLM. This allowed for exploiting the whole available record period of climatic forcing to yield the best estimates of antecedent soil moisture conditions. Then, each radar-based QPE was used as input to both models to obtain twelve peakflow simulations from GR4H and four peakflow simulations from ParFlowCLM. These peakflows are compared with the highest historically measured peakflow."

Comment 5: *"The parameter uncertainty of ParFlowCLM is strongly underestimated when focusing only on Manning values, as the authors did. At least they should do a better job considering uncertainty in the information about soil properties (lets only think to soil depth)."*

Authors' response: We agree with the Referee that the uncertainty of ParFlowCLM is underestimated without looking at other parameters, such as soil properties. We stated this in the Discussion section as one of the limitations of our study:

"The large uncertainty due to the Manning's coefficient is perhaps accentuated by the nature of the relationship between the coefficient and the discharge, but it is still here a lower bound since uncertainty to other parameters (hydraulic conductivity, van Genuchten parameters) was not included."

and

"Fourth, the accuracy of the parameter estimation in our study could be improved by investigating the uncertainty related to other distributed parameters (such as hydraulic conductivity; Poméon et al., 2020), or using hourly discharge streamflows for the GR4H calibration."

However, our objective was not to give an exhaustive quantification of the effect of parameter uncertainty on ParFlowCLM simulations. The large uncertainties caused by Manning's coefficient and QPE inputs illustrate how peakflow simulations are uncertain, let alone the contribution of other parameters. In addition, there are some computational limitations for us to do such an exercise. With the objective of having a regional scale model for flood forecasting, ParFlowCLM is currently implemented at the scale of Central Europe with $4*10^6$ grids and 15 soil layers, yielding a total of $6*10^7$ grids. We chose Manning coefficient as the peakflows are highly sensitive to this parameter and it is usually the focus in extreme flooding events studies (Lumbroso and Gaume, 2012).

In a very similar study with more focus on parameter uncertainty, Poméon et al. (2020) included uncertainty in the hydraulic conductivity on the simulations of ParFlow for flash floods events in several German catchments. However, they only adopted a uniformly distributed values of each parameter.

Comment 6: *"The use of English in the paper, while of a reasonably high standard, contains many idiosyncrasies, like the sentence: "The QPE impacted both GR4H and ParFlowCLM simulations", where 'Errors in the QPE impacted both…' is more likely."*

Authors' response: We corrected the sentence in question and checked for other idiosyncrasies in the revised manuscript.

Comment 7: *"References are missing lot of standard information."*

Authors' response: We completed the references list in the new version of the manuscript.

3 Cited References

Anctil, F., Ramos, M.-H., 2017. Verification Metrics for Hydrological Ensemble Forecasts, in: Duan, Q., Pappenberger, F., Thielen, J., Wood, A., Cloke, H.L., Schaake, J.C. (Eds.), Handbook of Hydrometeorological Ensemble Forecasting. Springer, Berlin, Heidelberg, pp. 1–30. https://doi.org/10.1007/978-3-642-40457-3_3-1

Chen, H., Cifelli, R., White, A., 2020. Improving Operational Radar Rainfall Estimates Using Profiler Observations Over Complex Terrain in Northern California. IEEE Trans. Geosci. Remote Sens. 58, 1821–1832. https://doi.org/10.1109/TGRS.2019.2949214

Chen, J.-Y., Trömel, S., Ryzhkov, A., Simmer, C., 2021. Assessing the Benefits of Specific Attenuation for Quantitative Precipitation Estimation with a C-Band Radar Network. J. Hydrometeorol. 22, 2617–2631. https://doi.org/10.1175/JHM-D-20-0299.1

Chen, J.-Y., Reinoso-Rondinel, R., Trömel, S., Simmer, C., and Ryzhkov, A., Submitted. A radar-based quantitative precipitation estimation algorithm to overcome the impact of vertical gradients of warm-rain precipitation: the flood in western Germany on 14 July 2021.

Lumbroso, D., Gaume, E., 2012. Reducing the uncertainty in indirect estimates of extreme flash flood discharges. J. Hydrol. 414–415, 16–30. https://doi.org/10.1016/j.jhydrol.2011.08.048

Mohr, S., Ehret, U., Kunz, M., Ludwig, P., Caldas-Alvarez, A., Daniell, J.E., Ehmele, F., Feldmann, H., Franca, M.J., Gattke, C., Hundhausen, M., Knippertz, P., Küpfer, K., Mühr, B., Pinto, J.G., Quinting, J., Schäfer, A.M., Scheibel, M., Seidel, F., Wisotzky, C., 2022. A multi-disciplinary analysis of the exceptional flood event of July 2021 in central Europe. Part 1: Event description and analysis. Nat. Hazards Earth Syst. Sci. Discuss. 1–44. https://doi.org/10.5194/nhess-2022-137

Poméon, T., Wagner, N., Furusho, C., Kollet, S., Reinoso-Rondinel, R., 2020a. Performance of a PDE-Based Hydrologic Model in a Flash Flood Modeling Framework in Sparsely-Gauged Catchments. Water 12, 2157. https://doi.org/10.3390/w12082157

Poméon, T., Wagner, N., Furusho, C., Kollet, S., Reinoso-Rondinel, R., 2020b. Performance of a PDE-Based Hydrologic Model in a Flash Flood Modeling Framework in Sparsely-Gauged Catchments. Water 12, 2157. https://doi.org/10.3390/w12082157

Rauthe, M., Steiner, H., Riediger, U., Mazurkiewicz, A., Gratzki, A., 2013. A Central European precipitation climatology – Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS). Meteorol. Z. 235–256. https://doi.org/10.1127/0941-2948/2013/0436