Referee 1:

Major comments:

1. (Abstract) L15-19: Throughout the manuscript, supporting materials for urban hydrology and mitigations of attenuation are not presented. Revise this part and reflect what has been presented.

Response: We will add more supporting material and more in-depth discussion about the link between radar rainfall measurements and their use in urban hydrology. The discussion will be based on the following list of papers:

- Aronica, G., Freni, G., Oliveri, E., 2005. Uncertainty analysis of the influence of rainfall time resolution in the modelling of urban drainage systems. Hydrol. Process. 19, 1055–1071. https://doi.org/10.1002/hyp.5645
- Bruni, G., Reinoso, R., van de Giesen, N.C., Clemens, F.H.L.R., ten Veldhuis, J.A.E., 2015. On the sensitivity of urban hydrodynamic modelling to rainfall spatial and temporal resolution. Hydrol. Earth Syst. Sci. 19, 691–709. https://doi.org/10.5194/hess-19-691-2015
- Courty, L.G., Rico-Ramirez, M.Á., Pedrozo-Acuña, A., 2018. The Significance of the Spatial Variability of Rainfall on the Numerical Simulation of Urban Floods. Water 10, 207. https://doi.org/10.3390/w10020207
- Cristiano, E., ten Veldhius, M., van de Giesen, N., 2017. Spatial and temporal variability of rainfall and their effects on hydrological response in urban areas – a review. Hydrol. Earth Syst. Sci. 21, 3859–3878. https://doi.org/10.5194/hess-21-3859-2017
- He, X., Sonnenborg, T.O., Refsgaard, J.C., Vejen, F., Jensen, K.H., 2013. Evaluation of the value of radar QPE data and rain gauge data for hydrological modeling, Water Resources Research, 49 (9), pp. 5989-6005, https://doi.org/10.1002/wrcr.20471
- Löwe, R., Thorndahl, S., Mikkelsen, P.S., Rasmussen, M.R. and Madsen, H (2014), Probabilistic online runoff forecasting for urban catchments using inputs from rain gauges as well as statically and dynamically adjusted weather radar, Journal of Hydrology, Volume 512, 397-407, http://dx.doi.org/10.1016/j.jhydrol.2014.03.027
- Niemi, T.J., Warsta, L., Taka, M., Hickman, B., Pulkkinen, S., Krebs, G., Moisseev, D.N., Koivusalo, H., Kokkonen, T., 2017. Applicability of open rainfall data to event-scale urban rainfall-runoff modelling. J. Hydrol. 547, 143–155. https://doi.org/10.1016/j.jhydrol.2017.01.056
- Ochoa-Rodriguez, S., Wang, L.-P., Gires, A., Pina, R.D., Reinoso-Rondinel, R., Bruni, G., Ichiba, A., Gaitan, S., Cristiano, E., van Assel, J., Kroll, S., Murlà-Tuyls, D., Tisserand, B., Schertzer, D., Tchiguirinskaia, I., Onof, C., Willems, P., ten Veldhuis, M.-C., 2015. Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: A multi-catchment investigation. J. Hydrol. 531, Part 2, 389–407. https://doi.org/10.1016/j.jhydrol.2015.05.035
- Rafieeinasab, A., Norouzi, A., Kim, S., Habibi, H., Nazari, B., Seo, D.-J., Lee, H., Cosgrove, B., Cui, Z., 2015. Toward high-resolution flash flood prediction in large urban areas – Analysis of sensitivity to spatiotemporal resolution of rainfall input and hydrologic modeling. J. Hydrol. 531, Part 2, 370–388. https://doi.org/10.1016/j.jhydrol.2015.08.045
- Rico-Ramirez, M.A., Liguori, S., Schellart, A.N.A., 2015. Quantifying radar-rainfall uncertainties in urban drainage flow modelling. J. Hydrol. 528, 17–28. https://doi.org/10.1016/j.jhydrol.2015.05.057

- Thorndahl, S. Nielsen, J.E. and Jensen, D.G. (2016) Urban pluvial flood prediction: evaluating radar rainfall nowcasts and numerical weather prediction models as inputs. Water Science and Technology 74 (11) pp: 2599-2610 http://dx.doi.org/doi:10.2166/wst.2016.474
- Wright, D.B., Smith, J.A., Baeck, M.L., 2014. Flood frequency analysis using radar rainfall fields and stochastic storm transposition. Water Resour. Res. 50, 1592–1615. https://doi.org/10.1002/2013WR014224
- Yoon, S.-S., Lee, B., 2017. Effects of Using High-Density Rain Gauge Networks and Weather Radar Data on Urban Hydrological Analyses. Water 9, 931. https://doi.org/10.3390/w9120931
- Zhou, Z., Smith, J.A., Yang, L., Baeck, M.L., Chaney, M., Ten Veldhuis, M.-C., Deng, H., Liu, S., 2017. The complexities of urban flood response: Flood frequency analyses for the Charlotte Metropolitan Region. Water Resour. Res. 53, 7401–7425. https://doi.org/10.1002/2016WR019997

2a. The link with hydrology or urban flooding/forecast:
One of the objectives of this study is to better understand the link between rainfall and urban flooding (L7-9) or/and the use of radar in hydrology and flood forecasting (L84-85). However, very few discussions were presented in this aspect. Add either more supporting materials for flooding parts (link with the presented work) or clarify better the objective of the presented work.

Response: More in-depth discussion about these issues will be added during revision (both in the introduction and the results). See response to major comment 1 for more details.

2b. Hydrological model (L171, L205, L397, L472, L490) has been mentioned in several sections without reference cited and the statements are rather generally made, which requires improvement in either writing or strengthening the explanation with more supporting materials (particularly for the statement made in the conclusion).

Response: This will be made more explicit during revision.

3. Better clarification and more supporting materials are required in results and conclusions (see the minor comments 16-37).

Response: The results and conclusion sections will be rewritten taking into account all referee comments. In particular, we will try to make a better and more clear distinction between representativeness errors (areal vs point) and overall accuracy of the radar products (as suggested by referee 3) and to better highlight the link to hydrological response.

Minor Comments:

1. L10-L11: Clarify better “the top 50 events”, “overall agreement”, “the peaks” of what.

Response: The sentences will be clarified during revision.

2. L44: need clarification of “accuracy” (of what).

Response: This will be clarified during revision. In the context of this paper, accuracy primarily relates to bias and root mean square error.

3. L46-47: This term “higher-level” composite is less objective and vague. Rephrase it.

Response: OK

4. L59-60: “, the longest...15-20 years at best.” Is it the case for world-wide or those countries presented in the manuscript?

Response: To the best of our knowledge, we believe that this is the case worldwide.

5. L76-78: “Often...the results” This is not clearly written in the context. Specify better. Also, adding more backgrounds/references to support strong needs in multinational assessment and comparisons will be necessary. At least, in Europe, there has been an effort made with BALTRAD products (Michelson et al. 2018, referenced already in the manuscript but in later chapter) and with the OPERA products (e.g., Saltikoff et al. 2019, Park et al 2019), which can be referred in the introduction.

Response: More background information about BALTRAD and more details about past international efforts for assessing and exchanging radar data will be added during revision (with references).

6. Table 2: Clarify the data resolution original vs. used for the comparison, e.g., in the text Line 128, Danish data has been interpolated to 1 min. In Table 3, is the comparison done at 5 min and at 1 min?

Response: The comparisons for Denmark were done at the 5 min resolution to match the resolution of the radar and gauge data. Although 1-min gauge data could be used in theory (using advection interpolation), this is not recommended here as this would add additional uncertainty due to interpolation. Also, the sampling uncertainty in rain gauge measurements at such short timescales would be very large.

7. L153-154: reference missing for the operational product.

Response: The following reference will be added during revision:

- Koistinen, J. and Pohjola, H., 2014. Estimation of Ground-Level Reflectivity Factor in Operational Weather Radar Networks Using VPR-Based Correction Ensembles, J. Appl. Meteor. Climatol. 53, 2394–2411, https://doi.org/10.1175/JAMC-D-13-0343.1

8. L164: “Polar radar measurements”. Describe better, it seems a jargon, meaning radar measurement done at polar grid.

Response: Yes, the measurements are made over a polar grid and projected afterwards. The sentence will be clarified during revision.

9. L170: After applying HIPRAD, the temporal/spatial resolution of the data remains the same as shown in Table 2?

Response: Yes, the output has the same spatial and temporal resolution.

10. L178, “Aalborg” add country name and indicate the coverage of this radar in Fig1.

Response: Sure, no problem.

11. L188: what is “tas BALTRAD”?

Response: this is the official name of the product we used. But since this is confusing and probably not helpful for the reader, the name will be shortened to “BALTRAD” during revision.

12. L206-208: Add reference

Response: OK

13. L290: “the HIPRAD” here, isn’t it BALTRAD?

Response: No, these are two different products.

14. L249: “the highest available temporal” This term is used several times later, but isn’t it the same as gauge sampling resolution (shown in table 1)? Is there any reason for such term? If so, explain better.

Response: The highest available temporal resolution refers to the highest common time resolution at which both radar and gauge data are available. This will be clarified during revision.

15. L 249: “Top event” → Event 1 (fig. 2), where are these gauges located in Fig 1?

Response: The location of the gauges will be highlighted in the figure and their distances to the closest radar will be mentioned in the text.

16. L253-254: Some results presented were already gauge adjusted and one (Finland) not. It is not clear to compare these numbers from literature examples (which is not clearly mentioned either if they were also derived before the adjustment or after?). Is it necessary?

Response: Yes, we believe that such comparisons are useful. At the same time, we agree with the referee that this can be rather misleading if done improperly. We will clarify this during revision.

17. L258: “The third rainfall peak” indicate here figure 4 (perhaps better with 4a indicating Denmark).

Response: OK

18. L264-265: “the relatively large peak intensity biases of 2.17, 2.09, 1.98 and 1.73 for Denmark, Finland, the Netherlands and Sweden...confirms this hypothesis” if the hypothesis refers the previous sentence, the bias for Netherlands should be larger than that of Finland because the peak intensity is higher for NL than for Finland (L256), isn’t it?

Response: The sentence will be changed to: “Clearly, the error structure between radar and gauges appears to fluctuate over time, with large deviations from the mean multiplicative bias in times of high rainfall intensities. As a result, the peak intensity biases (i.e., 2.17, 2.09, 1.98 and 1.73 for Denmark, Finland, the Netherlands and Sweden respectively) are systematically larger than the average bias over the whole event (i.e., 1.66, 1.37, 1.55 and 1.69).

19. L272 “at these scales” and L275 “such small scales”. What does it mean? Is it related to storm scale? Or do you mean that the comparison was done with the instantaneous and point estimates (that affects representativeness error)?

Response: It means that the measurements are compared at high temporal resolutions (i.e., 5, 10 or 15 minutes depending on the radar product). At these scales, sampling effects can have a rather large impact on traditional error metrics such as bias and rmse. We will clarify this during revision.

20. L283: This is redundantly written (merge with L280-282)

Response: The lines will be merged.

21. L300-301: Are these numbers MB after the ARFs reduction applied? is it also shown in Table 3?

Response: Yes, the numbers on L300-301 are before/after ARF. Table 3 gives the value of the bias that can be explained by the ARF.

22. L302-302: Is the statement made before applying the ARFs? Clarify better. After ARFs, Swedish result shows the best, doesn’t it?

Response: Yes, the Swedish product has the highest ARF. But if we take into account the part due to measurement support, its average multiplicative is the lowest of all. We understand that the way this is currently formulated in the paper might be confusing and will clarify this during revision (following up on a similar comment made by referee 3).

23. L306-307: This does not support any argument and redundantly written in L300. Rephrase or remove it.

Response: OK

24. L324, L405: “deeper analysis” Avoid “deeper” (somewhat subjective word) and revise the sentence.

Response: OK

25. L325: “temporal aggregation time scale” -> aggregation time scale (isn’t it the same as shown in Figures 8-10?)

Response: Yes, the formulation was not ideal. The sentence will be changed during revision.

26. L338-339: “Furthermore, the quality....an important role”. Add supporting explanation.

Response: We will provide more details on the radar and gauge data. See previous comments.

27. L359: It is not clear in Table 3 that the Danish products are the best in terms of RRMSE and CC. Revise this part.

Response: OK

28. L363-364: “However, a closer analysis....only 0.2”, what does it mean?

Response: It means that we performed a correlation analysis and found a value of 0.2 between PIB and peak intensity (in mm/h). The formulation of the sentence will be improved during revision and replaced by: “However, the rank correlation coefficient between the PIB and peak intensity is only 0.20. Therefore, intensity is likely not the dominant factor at play here.”

29. L375-376: Clarify what is “viewpoints”. Apart from the statement, how the attenuation and VPR correction applied to the group 2 data (Yes for Danish C band data, not explicitly indicated for the Swedish) were performed?

Response: More details about the geometry of the rain gauge network and the distances of the gauges to the radars will be added during revision (see response to major comment 2, referee 2).

30. L379: “a coarser scale” in time or/and space?

Response: In time.

31. L397-399: add reference. Is there any example run for the presented event?

Response: No, there’s no example run for this event. But we will provide references here to other papers cited in the introduction.

32. L418: “the same order...than for...” -> the same order...as for

Response: OK

33. L421-L422: This statement needs better supporting explanation, e.g., what dual-polarization capabilities was used in the processing of the data?

Response: This should now be clear thanks to the new information about the individual radar products. See response to major comment 3 of referee 2 for more details.

34. L469-470: “Bias correction...on peak intensity bias”. Is this conclusion derived from all the presented cases for four countries? There are some explanations for the Dutch product (L348-349), but not easy to find for the others. For Finland, the presented examples are not even bias corrected, so it is not clear what the authors mean.

Response: This was mostly referring to the Dutch and Swedish products and will be rephrased.

35. L471-472: Throughout the manuscript, “the importance of high-resolution radar observations in hydrological study” is hardly demonstrated/literature-reviewed with respect to the high-resolution radar products, which makes such conclusive statements weak. Add more solid outputs or references.

Response: More references will be added in the introduction during revision (see major comment 1). In addition, we will try to clarify the importance of the observed biases in terms of predicting hydrologic response.

36. L488-489: Add references.

Response: OK, no problem.

37. L489-490: Add references or strengthen supporting material for the referred rainfall uncertainties in hydrological models (e.g., some examples among any of the events 50 events*4 countries as a part of discussion or more explanation in L397-399).

Response: OK. We will use some of the references mentioned in the response to comment 1