The paper “Explorative Analysis of Long Time Series of Very High Resolution Spatial Rainfall” by Emma Dybro Thomassen et al. examines characteristics of extreme rainfall events studied using data from a C-band weather radar composite. The methods used by the authors are not novel, yet I found the analysis and results interesting and I believe they will be of interest for the readers of HESS as well. The paper is well structured and the analysis is sound. However, I found several issues that, although none of them are “major”, I would like to bring to the attention of the authors. Please see my comments and suggestions below.

[page lines]

[2 18] “…but none of these WGs describe the spatial dynamics of rainfall at a resolution suitable for urban hydrology” – I disagree! If you review the recent literature you will find at least 3 different rainfall generators that are generating gridded rainfall in a resolution that is finer than kilometer and 5 min. See, for example: Benoit et al. (2018) - Stochastic rainfall modelling at sub-kilometer scale, Peleg et al. (2017) - Partitioning the impacts of spatial and climatological rainfall variability in urban drainage modeling, and Peleg and Morin (2014) - Stochastic convective rain-field simulation using a high-resolution synoptically conditioned weather generator.

[2 24] “The downside of this is the absence of physically based variables to represent the spatio-temporal variation in rainfall and thereby enabling linking WGs to e.g. climate change models” – I think there is also some work done in this direction. See Paschalis et al. (2013) – “A stochastic model for high-resolution space-time precipitation Simulation”, and more recent - Peleg et al. (2017) – “An advanced stochastic weather generator for simulating 2-D high-resolution climate variables”.

[Case area] I am missing some climatological information here that will be needed for later understanding of your results. For example, is there any seasonality in the climate? Can you give some examples to what was consider as extreme events in the past (both intensity and duration)? What are the dominant synoptic systems in this region? I guess convective storms during summer and stratiform rainfall during winter? Moreover – high urbanization – where do I see that in Figure 1? “damage potential” – can you give some numbers to quantify the damage?

[3 22] “The post processed data have less than 5 % difference from annual ground truth” – what are the errors for the hourly and daily scales? The study deals with a much finer resolution than the annual scale. More important – how good are the rainfall estimates when extreme events are explored?

[Spatial selection of extreme events] – I understand you are searching for the best strategy to determine what events are ‘extreme’, but I am not completely understand the rational in your sampling strategies. SS1 – why this particular grid cell was chosen? How different your results be if the single grid cell was elsewhere? SS2 – Why not taking 5 connected grid cells? I am not saying that the strategy is wrong, but it needs to be better explained and discussed.

[5 3] “the five grid cells filled in Figure 2” – that means SS2, right? If so, please indicate this explicitly. Why only from SS2? What with the others sampling strategies? Please explain.

[5 7] “100 randomly selected grid cells” - Why not all grid cells? I guess because of computational time? Are the random selected grid cells represent the terrain? (e.g. all elevations are sampled)

[5 19] and [5 22] “stations” – you mean ‘grid cells’?

[Spatial variation] Please revise the text in this paragraph. It is not so clear how you define and compute the spatial variation.

[6 7-9] “The intensity threshold is set to 25% of the maximum 5-minute intensity for the given event with a minimum threshold of 7 mm h^{-1}. The areal threshold is set to a minimum coverage of 10 km^2” – How did you defined these thresholds? Are you following any physical reasoning? How different the results will be with different thresholds? For example, if 10 mm h^{-1} threshold was applied instead of the 7 mm h^{-1}? Some sensitivity analysis might be required here.

[6 9] “Event specific threshold” – what do you mean by that?

[Subsection 3.5.1-3.5.3] Can be shorten. There is no need in so many details, these methods are quite commonly used. I suggest summarizing all statistical methods in one paragraph and supply some key references for the readers.

[9 8-10] “In order to sample events which are meteorologically independent and use the knowledge about extreme events from rain gauge data, sampling method one, only considering one grid cell when sampling extreme events, is selected” – I found this sentence trivial. What I am missing from this paragraph is a clear statement of which sampling strategy to choose, and why. Some discussion might be useful.

[9 18] “convective vs. front events” – this is a good opportunity to discuss some of the differences between convective (more intense, smaller extent, shorter duration) and stratiform (or front) rainfall. It will make the reader better understand the clusters you are suggesting later on.

[Spatial correlation] Interesting. But again some discussion is missing. You can easily compare your results with other studies/climates, see for example: Villarini et al. (2008) – “Rainfall and sampling uncertainties: A rain gauge perspective”, Mandapaka and Qin (2013) – “Analysis and Characterization of Probability Distribution and Small-Scale Spatial Variability of Rainfall in Singapore Using a Dense Gauge Network”, and Peleg et al. (2013) – “Radar subpixel-scale rainfall variability and uncertainty: lessons learned from observations of a dense rain-gauge network”. Are there any differences of the spatial correlation when considering different rainfall types/clusters?

[Table 3] These results are not clear to me. See my comment on the spatial variability above.

[Principal component analysis] I would argue that the PCA results indicate that you cannot point on the most important variables to use for characterization of extreme events. Consider removing the PCA from the paper. I do not see the add value of it for the readers and I think the cluster analysis that comes later is much more important. Unless you want to discuss the cross-correlation or covariance between your variables. In this case you need to put some more effort in the discussion.

[Cluster analysis] Why 4 clusters and not, e.g. 3? I suggest adding some discussion about the climatic systems that force those clusters. For example, higher temperature for the convective

events, deeper atmospheric pressure for the intense convective – something in this direction. I think you need to demonstrate that there is some physical rational behind the clustering.

[12 18] "…only one grid cell…" - Then - why use the weather radar? Use a rain gauge instead. Using a single grid cell you are risking in missing the extreme events completely. Please clarify.

[12 21-22] "This study suggests that further development on a sampling strategy for sampling spatial extreme events is needed" – There are some studies that already goes a bit in a different directions. See recent studies by Lochbihler et al. (2017) – "The spatial extent of rainfall events and its relation to precipitation scaling" and Peleg et al. (2018) – "Intensification of Convective Rain Cells at Warmer Temperatures Observed from High-Resolution Weather Radar Data".

[Title] – 13 years of data is not "long time series" and 1-km is, today, not consider anymore as "very high resolution". I suggest the authors will aim to a simpler title, maybe including weather generator and Germany in the title?

[Figure 1] I suggest merging the two subplots. Can you please also add the locations of the three radars and the distances to the case area?

[Figures 6 and 9] Can go to the supplementary material.