

The paper “Data-driven distinction between convective, frontal and mixed extreme rainfall events in radar data” by Emma Dybro Thomassen et al. is in general well written, and it deals with an interesting and relevant topic for HESS. However, there are major issues to be addressed, as detailed below.

Major comments:

The main difficulty with this study is that it is not clear what is the research question that triggered the analyses conducted. It is stated that: “This study aims to quantify and describe spatial rainfall as a function of temporal and spatial dynamics, rainfall types and seasonal variation”. At the preceding paragraph it is written: “Here we apply a broad range of spatio-temporal characteristics in order to develop an automatic classification scheme of different event types and provide a better understanding of actual precipitation processes”. These are two different objectives, neither of them mention the word “extreme” which seems to be a main focus. In addition, the Introduction section goes through urban hydrology, design storms, gauge vs. radar rainfall, but not much on topics related to either of the potential objectives mentioned above.

Assuming the main goal is event classification, which also better matches the paper’s title, then there are three missing components. The first is validation. I guess the authors could ask some experts to (subjectively) classify the 39 events according to “convective”, “frontal” and “others” and test the clustering results against the expert classification. Without validation what we get is cluster analysis with a potential interpretation, not more than that. Second, I would expect a much deeper analysis/discussion of the space-time properties of each type and how is it related to precipitation processes associated with the event type (which was part of the declared objective). Lastly, classifying events is still far from developing an automatic classification scheme, so it is better to remove this part from the goal definition.

Other comments:

Event definition is not clear and I have several related questions: 1) the threshold intensity is per pixel or averaged over the area (with size depends on the sampling strategy)? if the former then how does it generalize for the entire area, if the latter then it is a problem because 1 mm/h for 1 km² and 38x48km² is very different. 2) for what duration the 1 mm/h threshold is applied? I guess that for 5-min, but it is not written clearly. 3) what is the minimal dry duration required for event separation? the authors provided the threshold to separate between dry and wet segment, but can 5-min separate two events? usually a minimal dry duration is set for event separation. 4) why 1 mm/h was set as a threshold? it is described as a “drizzle” threshold, but I think this is not a very low intensity. What percent of 5-min rainfall intensity is above this threshold?

Radar rainfall estimation: I am not clear what Z-R relation was used. It is written that the relation could be described by $Z=256R^{1.42}$. Was this the relation used? Another point is that the error reported is on annual basis while the data analyzed are for 15 min, 1h, 24h. So it would be much better to report the error for 24h, using cross-validation procedures (or daily rain gauges that are not participate in the gauge adjustment).

P5,L28: "Extreme events from five independent grid cells are sampled with SS1". In what sense these are independent and how do you know this? surely some events cover more than one pixel out of the five.

P9,L20: "In order to use the knowledge about extreme events from rain gauge data and be able to compare the results obtained to studies using rain gauge data, SS1 is chosen as the sampling strategy for this study". I don't understand why is it important to compare the results to rain gauge? it does not seem a part of any of the potential objectives.

Event selection: in relation to the above point, I am not convinced of the advantage of SS1 sampling strategy. Surely, as shown later in the paper, there are un-sampled extreme events that did not pass over the central sampling pixel with the highest rain intensity. Why not use the entire area?

P10,L1: "It is believed that the most severe extreme events in the case area is sampled for all grid cells, even though the ranking could be different between the grid cells". Why it has to be "believed" and cannot just to be checked? I am not sure this belief is correct. Since only 3 events per year are selected, considering the e-folding correlation distance of 5 km reported later, it can certainly be the case that the largest event in one pixel would be ranked more than 3 in another pixel and would not be selected.

Rain cell properties are not really discussed. It seems like "overkill" to detect and track rain cells without later on relate these properties to storm dynamics.

Minor comments:

P2,L2: Another rain gauge strength: direct measurement (much more accurate at the point)

P2,L5: Deriving spatio-temporal properties of the storm: the problem is not with the rain gauge instrument itself but with the rain gauge network that is often too sparse to represent these properties. In principle, a very dense rain gauge network could provide the relevant information on these properties (e.g., the Walnut Gulch gauge network in Arizona).

P.4,L2: 1.4 in percent or as a fraction (i.e., 140%)?

P6,L16: why $\Delta t = 11h$?

Rain cell properties are not really discussed

daily (abstract) vs. 24 h

Relevance to urban hydrology is not explored at all, so do not mention in abstract and in introduction