Review on Thompson et al. – Identifying regions of Europe which may be less prepared for extreme rainfall

General comments (GC):

(1) Extreme events:
While considering daily precipitation data (for good reason, as other long-term, gridded, continental-scale data is not available), both data and method include two different types of extremes related to heavy precipitation: river floods caused by widespread and long-lasting synoptic precipitation, and local flash floods mainly caused by short-term convective precipitation. These two types of extremes likely have different impact on the "social memory", that the study tries to cover by the "year since last record" approach. I admit that this is hard to separate (maybe by seasonal differentiated analyses), but at least I would expect a paragraph in the discussion section dealing with this topic.

(2) Introduction and motivation:
Following (1), I miss in the introduction and motivation some state-of-the-art discussion that combines the meteorological/climatological drivers of such extreme events, how they vary across the continent, what temporal natural variability is expected or already observed, and of course how these features are expected to change in the future. Although the focus of the study is neither on climate change nor the theoretical physics behind extremes, it will help the reader to connect roots, impacts, awareness/preparedness and need of action regarding extreme events. Within that context, the authors could not only mention the three case studies as examples, but list some more including references to demonstrate that such events are not singular but occur frequently, which motivates the necessity for preparation/adaptation (see also specific comments below).

(3) Using gridded data as is:
The authors themselves do mention in the discussion section, that local preparedness and also the general awareness of extremes is likely influenced by events happening near-by in the surrounding areas. This could be covered by using something like a distance weighting or additionally considering moving boxes, e. g., of 3x3 or similar. I wonder why the authors did not consider doing this kind of analysis and would be curious to see the effect on the preparedness maps.

(4) Wording:
While reading, I found it difficult to distinguish between the two meanings of the word "record", as it is used for both cases: (a) record meaning a "recorded" observed data series and (b) record in terms of an absolute maximum value occurred within a data series. I recommend to do a clear separation, e. g., always speaking of a data series regarding (a) and the maximum value regarding (b), or similar differentiation.

(5) Categorization:
I am also struggling on the usage of the "year since last record". This metric is strongly dependent (as also presented by the authors) by the considered time period and so it could be by chance, that a grid point is categorized into "recently moderate" or "recently rarity" just by the given data set. I wonder if there is a more objective quantification possible that covers the intention behind but is more or less independent of the start and end year of a given data series. I could image of a kind of bootstrapping choosing a period

not until the end of the data set but somewhere in between and moving start/end of the investigation period forward/backward. Alternatively, I suggest a more detailed discussion on the method and its applicability itself focusing on benefits and limitations such as the mentioned. Otherwise, the presented preparedness will always be “just a snapshot” of the current state at the time the analysis took place, i. e., the time the data series end.

(6) Cross-validation:
Main data source used in this study is E-OBS. The authors use the HadUK-Grid to cross-check the method and results (for the UK) using this independent data set. Later on, the authors present three case studies of extreme events: Valencia (Spain) in 2024, Ahr flood (Germany) in 2021, and storm Boris (Austria, Czech Republic, Poland) in 2024. I would have expected that instead of using a different data set for the UK for testing and cross-checking, to use data sets that also cover these case studies. For the Ahr flood, there is the HYRAS data from German Weather Service (DWD; https://doi.org/10.1127/0941-2948/2013/0436) with daily precipitation 1951-present-day at 1km resolution freely available; for the Valencia case there is for example the IBERIA01 data set (1971-2010; https://essd.copernicus.org/articles/11/1947/2019/) at 0.1° resolution; there seems to be a high-resolution continuously updated daily data set available at AEMET Spain as well; for storm Boris, there might be two data sets of interest, SPARTACUS for Austria (https://link.springer.com/article/10.1007/s00704-017-2093-x) 1961-present at 1km resolution or the G2DC-PL+ data for Poland (1951-2019, https://essd.copernicus.org/articles/13/1273/2021/) at 2km resolution. There might be other data sets for Czech Republic as well. I would recommend to include at least one or two of these data sets for the cross-checking of the results for the specific case studies as well.

(7) Discussion and conclusions
This section includes a lot of points where the method has limitations and shortcomings and only very few comments on the added value and the usability. In principle, such a discussion is fair, but in my opinion the focus too much on the limitations, which left the reader asking “why did you do this at all?”. I strongly recommend to include or highlight some more and real benefits of the presented approach, what is the novelty to existing techniques, etc.

Furthermore, as to some extend listed in the specific comments below, the discussion of the results in general is often very superficial and could be more profound and detailed covering some more aspects.

Specific comments (SC):

- L35ff: There is another study by Khodayar et al., 2022 (https://doi.org/10.1016/j.wace.2022.100493) focusing on comparable extreme precipitation events in the Western Mediterranean including drivers and other factors

- L40ff: This sounds like a classical Vb cyclone association, which would be typical for heavy precipitation and floods in that region. Was it really such a Vb-cyclone? You could mention this together with more examples like the 2002 and 2013 floods along the Danube and Elbe rivers, e. g., Schröter et al, 2025 (https://hess.copernicus.org/articles/19/309/2015/)

- L44: Talking about the Ahr flood 2021, also Mohr et al, 2023 should be mentioned here as well (it comes later in the text)

- L45: It is Mohr et al, 2023: https://doi.org/10.5194/nhess-23-525-2023 The reference used in the text is the preprint version

- L59ff: What is the actual impact of the change in return periods on the preparedness?

- L72ff: It is not unusual that “proper” measurements at a certain location cover only a short period, which could be shorter that the “memory” of the (elderly) people. In such cases, there would be a local awareness of extremes even though it is not covered by the scientific data. How would such situations affect the metric used in this study? Generally, is it even possible to consider such situations?

- L100: “E-OBS has been shown to be superior to alternative ... datasets … such as … ERA5.” This might be true, however, several studies have also shown shortcomings of E-OBS to raw station data or other gridded datasets, which should be mentioned as well. For example: Hofstra et al., 2009 (https://doi.org/10.1029/2009JD011799), Haylock et al., 2008 (https://doi.org/10.1029/2008JD010201), or Ehmele et al., 2020 (https://doi.org/10.5194/esd-11-469-2020)

- L106f: This is done separately for each grid point, right? Should be mentioned somewhere

- L109ff: Metric 2 and the used percentiles: Are there references for this metric and usage of specifically these percentiles? If it is your decision, please motivate it more detailed.

- L116f: Is it a valid assumption? What about changes in shape or general shifts of the precipitation PDFs with climate change?

- L122ff: You split the data into two groups at a risk threshold of 0.2, which is roughly a 50:50 division, mentioning that this is arbitrary. How would a different choice than affect the results presented later on?

- L132f: “...if a region would have been identified as sitting duck…”. I would leave this statement more open. “we can assess in which category a region may fall”, as the maps you produce not only allows for this single category but for all.

- L148: “...or observational uncertainty due to sparseness of data.” There are some more factors leading to uncertainties such as the interpolation from point stations to a grid etc.

- L149: “...cut-off low pressure systems…”. True as, for example, for the events presented by Khodayar et al., 2022.

- L154f: “...France, Belgium, etc .. regions are not likely to create such high-proportion events…”. Why is this the case? Less orography, so less forcing for ascend? Less convection due to less thermodynamic drivers?

- L168: “ in a static climate … the record to be equally likely to occur at any point…”. Is there any proof (e. g., from literature) for this rather hard statement?

- L169: “...more records in later years…”. Could this be an effect of more station data with shorter time series entering E-OBS?

- L176f: “more records prior to 1985 in central Europe and UK”. What are possible reasons for this? Is there a shift in seasonaliy? Is there a shift in the general precipitation regime? Is there a shift in cyclone tracks? Drier summers with less convection or less Vb cyclone could lead to less records in central and eastern Europe, while a general shift of cyclone tracks, for example, to the north could lead to less extremes in the UK.

- Fig. 4d + 5: The risk metric seem visually more “patchy” in central Europe than in Scandinavia or Iberia, so the level of detail seems higher. Could this be artificial due to the higher station density in central Europe?

- L188f: You state that the percentile maps are smooth compared to the record map due to more data points included in the calculation. Yes, BUT: Your are talking about the “records” so the absolute maximum values occurred. Due to the limited amount of available data (short data series), the distance between two percentile is not linear even when using very fine steps. So the absolute value very often is well above the 99.99…% percentile, which also results in a more scattered pattern when plotting.

- L214ff: The results for the preparedness are partly really patchy. Therefore, I think the interpretation given in this lines is too detailed and the results should not be over-interpreted. Furthermore, as you correctly state in the discussion section (see also GC3 above) the local results may be affected by “extremes nearby”, so a distance weighting or smoothing could help preventing an over-interpretation.

- Fig. 6: Is there an objective criterion for the choice of cities? It seems rather random.

- Fig. 7: Even on the screen, the white markers are hard to see and the marker style is very hard to distinguish, especially in Fig. 6j, it can only be guessed where the white star and white dot is. For the color, I suggest using black. You could also label all the dots in all figures, e. g., with B1, B2 for storm Boris, etc.

Furthermore, I struggle with the post-event analyses. It seems rather a logical consequence by definition of the used metrics, that the preparedness changes directly after an extreme event. So, I wonder what the actual implications of these results are and what you want to demonstrate.

- L248f: I miss some more discussion the results of Fig. S3. What does this mean for the presented method and results? This also relates to GC5 regarding the used time period (snapshot).

- L255ff: Here, you are mixing convective extremes and synoptic precipitation extremes. Convection has a more local and direct impact (flash floods), while the location of the synoptic precipitation leading to floods along major rivers likely is not necessarily co-located with the place of flooding.

- L260: You separate the case study Ahr and Valencia by a new paragraph, so I would also separate storm Boris here from the Valencia event starting a new paragraph.

- L298f: Beside then mentioned factors, also general constructions (urbanization) within former flood plains or bridges building dams from clogged debris have an effect on the impact as shown, for example, in the Ahr case by Mohr et al, 2023 or Roggenkamp and Hergert, 2022.

- L305: “sitting ducks as have been lucky so far”. In my opinion, this is a crucial point in terms of psychology when it comes to acceptance and take action of adaptation or preparation measures.

- L310ff: Fitting a distribution could help but must also be done carefully with some assumptions such as stationarity of the rainfall distribution. Also extremes might not be best captured by fitting one function to the complete distribution. Especially for precipitation, the pdf is largely skew and the heavy tail likely need a different-shaped function as the rest. Furthermore, when fitting a distribution to observations, the percentile estimate remains a “first guess”. One solution could be the use of large model data ensembles as for example shown by Ehmele et al., 2022 (https://doi.org/10.5194/nhess-22-677-2022).

- L320ff: Could a detrending of the time series prior to the analysis help to reduce the influence of possible temporal changes within a data series?

- Fig. S2: in the caption you state that the risk ratio of the 12km HadUK is similar to the 5km version (not shown). To convince the reader and as evidence that your approach is independently applicable to various data sets with different resolutions however producing comparable results, you should present these in an additional supplemental figure.