Reviewer 2

The presented study by Thompson et al. entitled "Identifying regions of Europe which may be less prepared for extreme rainfall" investigates, how well different regions across Europe are prepared regarding extreme precipitation events. For this purpose, they apply a metric on two observation-based gridded precipitation data (E-OBS, HadUK-Grid), which combines the statistical approach of the risk of a new rainfall extreme (absolute magnitude of an extreme) with the probability of a new extreme to happen (time since the last extreme occurred). The latter also includes a more social science related aspect in the form of "the memory of people" on extremes. The further back in the past the extreme event took place, the less "memory" exist. To my opinion this is a highly relevant aspect when it comes to climate communication and acceptance of adaptation measures.

The study is well-structured and easy to follow. Data and methods are well described, and the results are presented in a straightforward and easy to follow way. It falls within the scope of NHESS.

Thank you for reviewing our manuscript. We are pleased that you find our study well-structured and easy to follow.

However, I have some major concerns on some specific points, which to my opinion would improve the quality of the study significantly and better highlights the importance of this type of analysis, especially regarding applications, and which should be considered in a revised version. Please find attached a detailed list of comments and recommendations.

We have made changes based on both your comments and those of the other reviewers, and we hope you find the new version worthy of publication.

In response to your comments we have updated our manuscript, improving figures and clarity of the text. We have expanded the introduction and discussion. Additional supplementary figures have been added. The first shows the 'assumed preparedness' divided across a greater number of categories. Secondly, we have included a second dataset to verify our results – using the HYRAS dataset as suggested to compare results over Germany, finding highly consistent patterns.

We also identified an error in the calculation of the 99th and 99.5th percentile data. This error correction has led to a change in the risk metric and the threshold used to differentiate between categories in the preparedness framework. It has impacted the magnitude and leads to small changes in the spatial pattern. Some shift in location categories has occurred (as most obvious when comparing the incorrect and corrected Fig. 6). However, the conclusions of the paper are unaltered.

Below are point-by-point responses, with changes to the text in italics with line numbers – as in the tracked changes version – indicated.

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.

We expand the discussion to highlight the limitation of using this single measure of daily rainfall to encompass all rainfall extremes. We have also added further discussion of how flood impacts do not always co-locate with the rainfall – where large scale rain leads to catchment level flooding.

Line 372: *We assess extreme rainfall based on daily data and so both large-scale rainfall events and short-term convective rainfall are merged. However if, for example, hourly data was available it could be possible to assess convective rainfall – more likely to cause localised flash floods – independently. The two types of rainfall extremes will lead to differing flood impacts, which may create different levels of social memory.*

Line 354*: We assess at gridpoint scale, however, assessment over hydrological catchments could enable identification of non-spatially coherent flood hazard from extreme rainfall.*

The different types of extremes are now also mentioned in the introduction (see next comment and response).

(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).

We have added a paragraph to the introduction providing an overview of extreme daily rainfall events in across Europe.

Line 52: *Such extreme rainfall events are occurring with increasing frequency and intensity in our changing climate (Seneviratne et al., 2021). Such events can arise from many different atmospheric phenomena such as large scale extratropical cyclones and short-term, localised, convective storms. Impacts can be catastrophic, as seen in the example above, and it is not only Europe that is affected. In 2024 alone extreme rainfall globally contributed to >8,700 deaths and >US $550 billion in damages (Green et al., 2025). In this study, we focus on Europe - where extratropical cyclones and cut-off low pressure systems regularly cause extreme rainfall in winter and convective storms can lead to intense short-timescale rainfall. Different event types tend to impact specific regions and seasons, with similar events often repeating themselves. For example, large-scale low-pressure weather systems isolated from the jet stream, known locally as DANA (Depresión Aislada en Niveles Altos), regularly occur over the Iberian peninsula in September, whereas further east torrential, convective, rainfall later in the winter is more frequent Grimalt-Gelabert et al., 2021). In northern Europe winter heavy precipitation is primarily driven by Atlantic low-pressure systems, modulated by the North Atlantic Oscillation (Hurrell, 1995; Delworth et al., 2016). Projected future changes in the NAO suggest an increase in extreme winter rainfall in Northern Europe (Fuentes-Franco et al., 2023). A warmer atmosphere can hold more moisture, and thus extreme rainfall is expected to get worse throughout Europe with further warming (Fowler et al., 2024).*

Fowler, H.J., Blenkinsop, S., Green, A. and Davies, P.A., Precipitation extremes in 2023: Climate chronicles. Nature Reviews Earth & Environment, 5(4), pp.250-252. 2024.

Green, A. C., Fowler, H. J., Blenkinsop, S. & Davies, P. A. Precipitation extremes in 2024. Nat. Rev. Earth Environ. ***6****, 243–245. 2025.*

Grimalt-Gelabert, M., Alomar-Garau, G. and Martin-Vide, J., Synoptic causes of torrential rainfall in the Balearic Islands (1941–2010). Atmosphere, 12(8), p.1035. 2021.

Hurrell, J.W., Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. Science, 269(5224), pp.676-679. 1995.

Delworth, T.L., Zeng, F., Vecchi, G.A., Yang, X., Zhang, L. and Zhang, R., The North Atlantic Oscillation as a driver of rapid climate change in the Northern Hemisphere. Nature Geoscience, 9(7), pp.509-512. 2016.

Fuentes-Franco, R., Docquier, D., Koenigk, T. et al. Winter heavy precipitation events over Northern Europe modulated by a weaker NAO variability by the end of the 21st century. npj Clim Atmos Sci ***6****, 72 https://doi.org/10.1038/s41612-023-00396-1. 2023.*

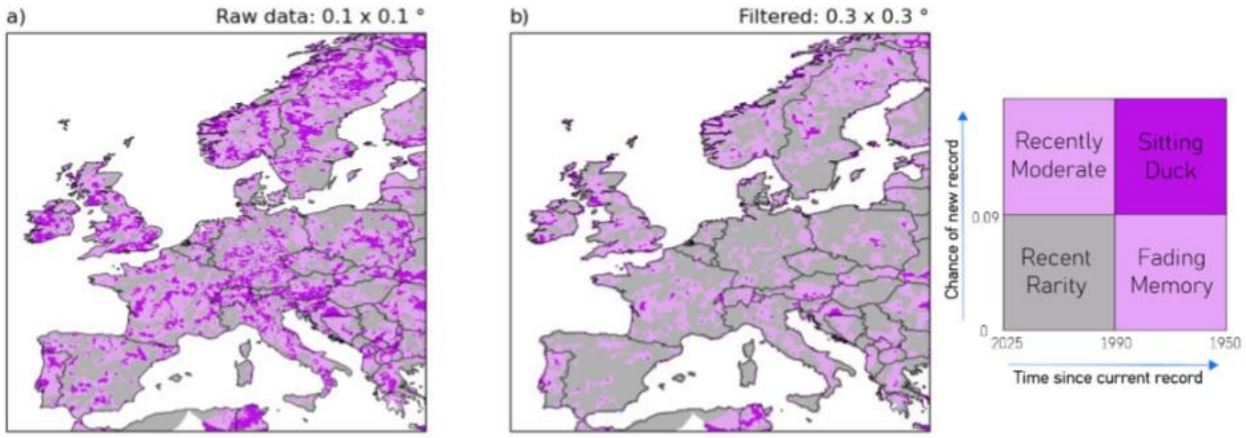
(3) Using gridded data as is:

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.

This is something we had considered, however we worry that not every location will prepared based on extreme events in nearby locations. As we comment on in the discussion (line 354) it may be that adaptation would occur over an administrative area – and therefore a nearby event which is not within the same administrative region may not have the same influence. Further understanding of the role of nearby events – whether by distance or within administrative regions – would aid development of this method as a tool for policy and risk management.

We have produced an additional figure considering 3x3° moving boxes (3x3 gridpoints). We cannot take the greatest chance of new record and longest time since current record independently for a 3x3 box – as then we may merge information from two separate events. Instead, for this we re-allocate the categories into 3 colour categories rather than 4, enabling a linear progression. Each gridpoint is allocated the lowest value (greatest assumed preparedness) of the 3x3 grid surrounding it, i.e. it is shifted to the point closest to the bottom left of the framework. We observe a drastic reduction in regions categorised as sitting ducks – suggesting if neighbours do adapt based on nearby events the preparedness is greatly increased.



(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.

We have made changes throughout the manuscript to ensure a clearer distinction. We do not remove the term record completely, but do not use it to mean the recorded observed data at any point.

The changes made include using 'record-breaking' , 'record-breaking event' , or 'record event' in place of 'record' where appropriate. Where grammatically possible, 'record' has been changed to 'greatest'. In the methods we explicitly state that the maximum value is the record. Example of the changes made include:

Line 27: *how recently the greatest daily rainfall occurred;*

Line 47: *local maximum values were set at many sites…*

Line 138: *To calculate the first metric, we identify the current maximum value (record event) in the timeseries.*

Line 181: *The observed greatest daily rainfall (mm).*

Line 208: *Figure 3: When did the current record event occur? Showing the year of greatest E-OBS daily rainfall (1950-2025).*

Line 8215: *We find 50% of the region has experienced greatest daily rainfall since 2000 (33% of the data period), and 26% within the last 10 years of the 75 year long dataset (13% of the data period).*

Line 365: *For instance, the map of record event (Fig.3) shows only the greatest event in the dataset.*

(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.

Indeed, the assumed preparedness presented in Fig.5 is a snapshot of the current state up to the end of 2025. As new events occur, and as past events fade further from memory, the assumed preparedness of a specific location will change. The case studies shown in Fig.6 illustrate this – with an increase in assumed preparedness shown after all three rainfall events assessed. In the supplementary figures we show how the assumed preparedness is altered when using different end dates – illustrating how new, record-breaking rainfall events will lead to increased assumed preparedness.

As memory is one component we are using to represent preparedness we do not consider different start dates – we use the full available dataset from 1950. Determining if there is a

point where past data is no longer relevant (where memory can no longer increase preparedness) would be a highly informative – but different – study.

(6) Cross-validation:

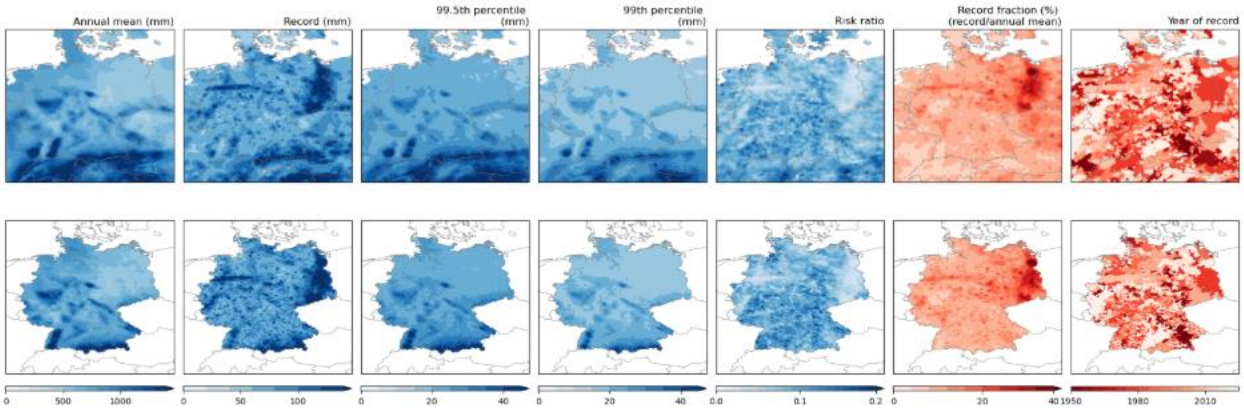
Main data source used in this study is E-OBS. The authors use the HadUK-Grid to cross check the method and results (for the UK) using this independent data set. Later on, the authors present three case studies of extreme events: Valencia (Spain) in 2024, Ahr flood (Germany) in 2021, and storm Boris (Austria, 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.

We have added another dataset to cross-check the results – this time for a region included within the case studies. As suggested, we use the HYRAS data – chosen as it is open source and available from pre-1950 to the present day. We have been unable to source other options which are available up to the present day (and thus include any of our case studies). We note, HYRAS is limited to Germany, whereas the Ahr flood impacted multiple countries. However, within the region we can compare the HYRAS dataset provides comparable results to E-OBS, adding further confidence in the findings presented in the main manuscript.

In preparing this figure, we identified a coding error in calculating 99th and 99.5th percentiles – so all figures have been updated accordingly.

For further studies investigating specific countries or regions we would recommend considering all available datasets, and ideally comparing results from multiple data sources.

New supplementary figure:



Supplementary Figure 3: Comparing the E-OBS 0.1° and HYRAS 1km datasets for the Germany. The top row is from E-OBS as in the main text, with HYRAS for the same time period (1950-2025) in the bottom.

Updated text, to incorporate additional data source:

Line 130: *To improve confidence in the results the methodology is applied to two national datasets: HadUK-Grid and HYRAS (Hollis et al., 2019; Rauthe et al., 2013). HadUK-Grid ~~which~~ provides gridded climate data for the UK based on observations from the UK's land network of weather stations. Daily rainfall data is available from 1891 onwards; we use the 5km gridded data from 1891-2025. HYRAS is a high-resolution gridded daily dataset covering Germany, available from 1940 onwards and on a 1 km spatial resolution (Rauthe et al., 2013). We use data from 1950-2025, the period available in both HYRAS and E-OBS.*

Line 200: *Similar maps produced using HadUK-Grid and HYRAS are included in Fig.S2 and Fig.S3. Consistency between datasets improves confidence in the results. The spatial patterns of E-OBS and the two national datasets are broadly in agreement and the patterns do not change when including data spanning a longer time period.*

Line 406: *Comparisons with other observational datasets, including those with global coverage, would be beneficial. Here, we show results from E-OBS and two national datasets for the UK and Germany*

Data availability: *The HYRAS dataset is available from the German Weather Service, DWD (Index of /climate_environment/CDC/grids_germany/daily/hyras_de/precipitation/).*

Additional reference: *Rauthe, M., Steiner, H., Riediger, U., Mazurkiewicz, A. and Gratzki, A., A Central European precipitation climatology–Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS). Meteorologische Zeitschrift, 22(3), pp.235-256. 2013*

(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.

Thank-you for highlighting this, suggesting that we currently downplay the importance of the approach. We have reframed sections of the discussion to better highlight the benefits and novelty of the method and results presented. We hope this helps to emphasis the value of the study.

We have also expanded on the discussion of the results, as outlined in the more specific comments below.

Line 333: *To enable better understanding of observed extremes, we present a map of assumed preparedness to daily rainfall extremes across Europe (Fig.5). This map can be used to identify regions which may have lower levels of preparedness, based on their past experiences and statistical likelihood of a new record-breaking event. Identified locations may benefit from focused measures, enabling scientifically supported distribution of resources.*

Line 344: *We show that for some example events, the method we present would have flagged the region as low assumed preparedness prior to the event, and so could have benefitted from focused adaptation or risk management plans. For both the Ahr floods and Storm Boris prior to the event the regions impacted show less levels of assumed preparedness with areas marked ‘sitting duck’ – impacts may have been reduced by local education and communication to reduce risks (Ommer et al., 2023; da Costa et al., 2026).*

Line 363: *We use just two simple metrics to assess preparedness, enabling the method to be easily replicated for different regions, datasets, or climatic variables. However, using such metrics leads to some assumptions and simplifications.*

Line 378: *Further research into how memory of an extreme weather event fades and over what timescales past experiences still matter is essential to enhance the value of the results presented further.*

Line 420: *In conclusion, we present a novel method to assess spatial variability in assumed societal preparedness to extremes using observational data alone. In this warming world we know record-breaking extreme rainfall events will become more common. Society needs to prepare for future extreme rainfall events, and the current preparedness to these events differs by location. Our method allows better understanding of the spatial distribution of such risks.*

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

Thank-you for pointing out this study - which is also very relevant to other work I am undertaking. We now cite this within the results section:

Line 38: *The Spanish Mediterranean is particularly susceptible to heavy rainfall, with impactful events often occurring in Autumn (Khodayer et al., 2022).*

Additional reference:

Khodayar, S., Pastor, F., Valiente, J.A., Benetó, P. and Ehmele, F., What causes a heavy precipitation period to become extreme? The exceptional October of 2018 in the Western Mediterranean. *Weather and Climate Extremes*, *38*, p.100493. 2022.

- 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/)

Indeed, there is evidence that Storm Boris was a classical Vb event. We have added further literature about earlier events and the association of Vb events with extreme rainfall:

Line 43: *The event followed the classical 'Vb' pattern which often leads to impactful extreme rainfall in this region, earlier examples include the 2002 Elbe flood and 2013 Danube and Elbe flood (Ulbrich et al., 2003; Messmer et al., 2015; Schröter et al., 2015; Hofstätter et al., 2016). There is evidence that such events will decrease in a warmer climate, though a shift in rainfall patterns could occur (Messmer et al., 2020).*

Additional references:

Ulbrich, U., Brücher, T., Fink, A.H., Leckebusch, G.C., Krüger, A. and Pinto, J.G., The central European floods of August 2002: Part 2-Synoptic causes and considerations with respect to climatic change. Weather, 58(11), pp.434-442. 2003.

Messmer, M., Gómez-Navarro, J. J., and Raible, C. C.: Climatology of Vb cyclones, physical mechanisms and their impact on extreme precipitation over Central Europe, Earth Syst. Dynam., 6, 541–553, https://doi.org/10.5194/esd-6-541-2015, 2015.

Hofstätter, M., B. Chimani, A. Lexer, and G. Blöschl, A new classification scheme of European cyclone tracks with relevance to precipitation, Water Resour. Res., 52, 7086–7104, doi:10.1002/2016WR019146. 2016.

Messmer, M., Raible, C. C., and Gómez-Navarro, J. J.: Impact of climate change on the climatology of Vb cyclones, Tel lus A: Dynamic Meteorology and Oceanography, 72, 1–18, https://doi.org/10.1080/16000870.2020.1724021, 2020.

Schröter, K., Kunz, M., Elmer, F., Mühr, B. and Merz, B., What made the June 2013 flood in Germany an exceptional event? A hydro-meteorological evaluation. Hydrology and Earth System Sciences, 19(1), pp.309-327. 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

The reference has been updated to the published version (both in the text and the reference list). It is also now cited earlier in the Ahr flood paragraph.

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

Change in return period would not automatically lead to a change in preparedness – but increases (or decreases) the likelihood of an event occurring. If return period is reducing then the likelihood of a high impact event increases, and this preparedness may be reduced unless action is taken. The text has been expanded to convey this:

Line 78: *The change in return period metric used by Masukwedza et al. (2025) assessed the extent to which risk is changing in each location. This is a metric of how rapidly the climate in that location is changing, therefore a reduction in return period without adaptation would indicate reduced preparedness. We do not include this metric when assessing assumed preparedness. As will be discussed further in later sections, the metrics we use do not take climate change into account.*

- 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?

In this study we are using data from 1950, but there are examples of earlier extreme events which are clearly still remembered within a community - for example the Holmforth flash floods in the UK in 1944 were remembered locally 80 years on (Holmfirth: Fatal 1944 flood anniversary marked with walking trail - BBC News). Through methods such as questionnaires and monitoring of oral local history such situations could be considered in a qualitative manner, however it would be tricky to quantify. We do comment on this in the discussion, and now add a comment on the length of the dataset used:

Line 357: *Some regions may have longer 'memory' than our dataset allow for, perhaps due to notable impacts to the community, and how that memory is transmitted. ~~(~~e.g. local oral history and education~~)~~ (Metcalfe et al., 2020).*

- 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)

The shortcomings of E-OBS are now more thoroughly discussed:

Line 126: *However, E-OBS also has limitations – Haylock et al (2008) showed that E-OBS values are up to one-third less than station data at the same location. Hofstra et al. (2009) show even greater differences for some extreme rainfall events. The spatial mean is very close and there are no alternative datasets providing the resolution and spatial and temporal coverage of E-OBS (Ehmele et al., 2020).*

Ehmele, F., Kautz, L.A., Feldmann, H. and Pinto, J.G., Long-term variance of heavy precipitation across central Europe using a large ensemble of regional climate model simulations. Earth System Dynamics, 11(2), pp.469-490. 2020.

Haylock, M.R., Hofstra, N., Klein Tank, A.M.G., Klok, E.J., Jones, P.D. and New, M. A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. Journal of Geophysical Research: Atmospheres, 113(D20). 2008.

Hofstra, N., Haylock, M., New, M. and Jones, P.D., Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature. Journal of Geophysical Research: Atmospheres, 114(D21). 2009.

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

Indeed it is at a gridpoint level, we now clarify this:

Line 137: *To calculate the first metric, for each gridpoint we identify the current maximum value (record event) in the timeseries. We then determine the year in which this event occurred.*

- 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.

There are not references for this exact metric, though it is based on the same idea as the ECMWF 'shift of tails' metric which we now cite. In developing the metric we assessed multiple options – initially take a more simplistic approach of simply the difference between 99th percentile and the record. Incorporating both 99th and 99.5th percentiles allows a better representation of the 'fatness' of the tail, without the need for a full distribution fit as would be required if using an EVT approach. We argue that the simplified representation enables a more reproducible method, particularly when applying to high resolution and/or long time period datasets.

Changes to the manuscript:

Line 140: *This metric is based on percentiles, and provides information about how extreme an event is, in a similar manner to the ECMWF 'shift of tails' metric (Tsonevsky, 2016).*

Added reference:

Tsonevsky, I., 2016: New EFI parameters for forecasting severe convection. ECMWF Newsletter, No. 144, ECMWF, Reading, United Kingdom, 29–32, https://www.researchgate.net/publication/292151482_New_EFI_parameters_for_forecasting_severe_convection.

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

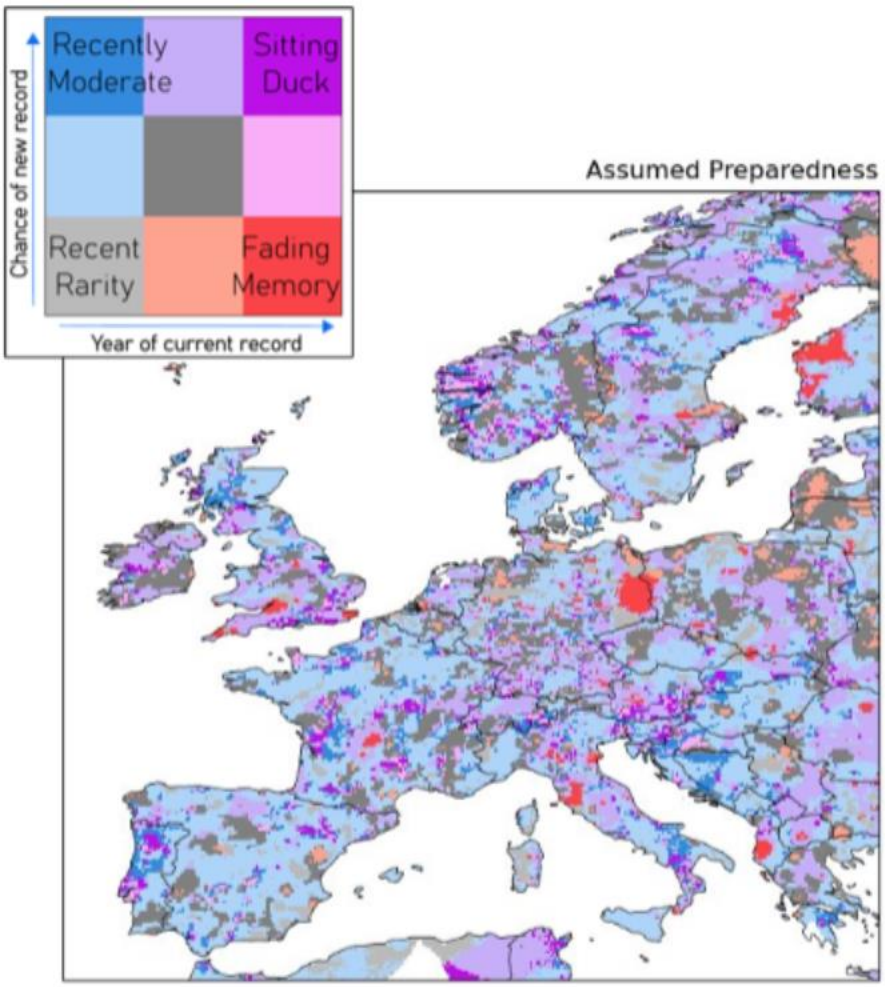
This has been altered – if climate change leads to shifts in the shape (i.e. longer tails) then the assumption is not valid. This is now stated.

Line 150: *In a changing climate this may not hold true – as the shape of the tail may vary temporally. We discuss the limitations of not incorporating climate change explicitly in later sections.*

- 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?

Indeed, the split in data could be set differently – or with more categories as suggested by another reviewer. We did compare different options – such as splitting into a 3x3 grid with 9 categories. This figure is now included as Fig.S5. For different users more nuanced presentation of results could be considered – with categories divided different, such as to identify the 10% most at-risk locations. Changes to add Fig.S5:

Line 160: *The data could also be divided into more categories to provide more nuanced results – in supplementary figures we include a version divided into 9 categories, with time since current record split at 1980 and 2000 and chance of new record split at 0.06 and 0.12.*



Supplementary Figure 5: Assumed preparedness across Europe. *Showing each location classified by the assumed preparedness framework, using E-OBS 1950-2025. Year of record divided into 3 at 1980 and 2000. Chance of new record divided at 0.06 and 0.12.*

In Fig.6 we show the exact position within the preparedness framework of 20 locations, from this we can visually assess how shifting the boundaries would alter the categorisation for those locations – and see how some locations sit close to boundaries.

- 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.

Changed to your suggestions:

Line 170: …*we can assess in which category a region may fall.*

- 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.

This has been added:

Line 187: *…observational uncertainty due to the sparseness of data or the interpolation from station data to the gridded product.*

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

This, and an additional reference referring to Storm Boris, have been added:

Line 188: *In contrast, in western Europe large scale organised weather features – such as cut-off low pressure systems - cause larger scale events (Riboldi et al., 2025; 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?

There are many possible reasons, factors may include orography, the proximity to Mediterranean as a moisture source, or other drivers. Further text has been added, but investigation is beyond the scope of this study:

Line 194: *In contrast northern France, Belgium and the Netherlands show much lower levels. This suggests that the meteorological conditions of these regions are not as likely to create such high-proportion events – perhaps due to local orography or the proximity to the Mediterranean as a moisture sources. In such region perhaps preparing for the same levels is unnecessary.*

- 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?

In Line 210, we change *'would'* for *'could'*, as climate variability may naturally lead to periods of more extreme events, and we now cite two articles to support this statement:

Fischer, E.M., Bador, M., Huser, R., Kendon, E.J., Robinson, A. and Sippel, S., Record-breaking extremes in a warming climate. Nature Reviews Earth & Environment, 6(7), pp.456-470. 2025.

de Vries, I., Schillinger, M., Fischer, E., Sippel, S. and Knutti, R.: Precipitation disaster hotspots depend on historical climate variability. Nature Communications, 2025.

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

We have reworded this, adding a comment on the possible role of changing station data input to E-OBS, along with an additional reference:

Line 212: *Climate change is increasing extreme rainfall (Seneviratne et al., 2021), and thus it is unsurprising that we see more records occurring in the later years – however the density of input stations into the datasets also shows temporal inhomogeneity which may have play some role (van der Schrier et al., 2025; Fig.3).*

van der Schrier, Gerard & National Center for Atmospheric Research Staff (Eds). Last modified 2025-12-11 "The Climate Data Guide: E-OBS: High-resolution gridded mean/max/min temperature, precipitation and sea level pressure for Europe & Northern Africa." Retrieved from https://climatedataguide.ucar.edu/climate-data/e-obs-high-resolution-gridded-meanmaxmin-temperature-precipitation-and-sea-level on 2026-09-22.

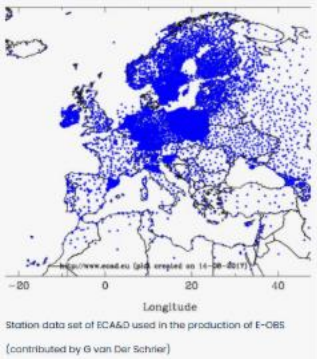
- L176f: "more records prior to 1985 in central Europe and UK". What are possible reasons for this? Is there a shift in seasonaliy? Is there a shift in the general precipitation regime? Is there a shift in cyclone tracks? Drier summers with less convection or less Vb cyclone could lead to less records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

We do not explore trends in the extremes, only the year of the greatest daily rainfall. We cannot infer more events generally in a period – only that the most extreme occurred then. The dark red indicates that there was at least one (and maybe only one) extreme which is the current record. For instance, the dark region in south-west UK is one shade, indicating the possibility that one single event leads to the shading. Text has been altered to clarify:

Line 221: *There are regions of dark red, indicating long-standing record events from prior to 1985, in central Europe and in southern UK. Early record events cannot be used to infer a decreasing trend in rainfall – as we show only the one individual greatest event.*

- 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?

This may be true for Poland, which has the greatest density of stations included. Below I include a map of station data used. This shows that generally Scandinavia is well covered, though Iberia is not. As there is not notable difference in patchiness between Scandinavia and Iberia it does not appear that it is due, solely, to station density.



- 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: You 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.

Indeed, we agree that percentiles will also suffer from data limitations – but argue that this is at least reduced when more data points are included within the calculations (e.g. record only one data point, whereas 99th includes 200). As you say, as we move towards higher percentiles the 'smoothness' will be expected to decrease – and thus 99.5th will be less smooth than 99th. See ~Line 212 for our discussion of this.

- 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.

We highlight in the discussion the need for localised interpretation of results, as some places will adapt based on nearby events whereas other places may not. Without local knowledge the level of preparedness cannot be assessed – hence the use of 'may be' throughout these results which we hope prevents overinterpretation.

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

The cities used are the largest 20 within the region, based on urban population. This has been added to the text:

Line 272: *The assumed preparedness of individual locations can be identified; we show the year of record event and risk measure for 20 major cities across Europe, determined by greatest urban populations (Fig.6).*

- 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.

This figure has been replotted with larger markers, and with lines added joining the 'before' and 'after' locations of each location.

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.

The case studies are to demonstrate that the preparedness does change directly after an extreme event, i.e. that we do see the logical consequence. We aim to demonstrate that the method does provide the logical result – now highlighted in the text.

Line 281: *As for the case studies, this indicates that the method gives the expected results.*

- 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)

The discussion of Fig.S3 has been expanded:

Line 309: *In Fig.S3 we also show the preparedness maps for the whole of Europe using subsetted data, to show how other regions change through time. Without any change to the boundaries of the 'assumed preparedness' categories it would be expected that more regions would become grey, indicating a recent rarity, as more recent (post-1990) years are included. We see this is indeed the case, with many of the regions highlighted in Fig.S3d/e moving to categories of higher 'assumed preparedness'. As for the case studies, this indicates that the method gives the expected results.*

.- 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.

This section has been reworded to highlight the influence of river catchments and steep terrain in the region, leading to the lack of coherence in rainfall and impact locations:

Line 315: *In the case of the Valencia event (Fig.7b) the greatest rainfall is not over the region where greatest impacts were documented but instead downstream, within the same catchment. There is less change in classification after the event (Fig.7e/h). Further inland there are regions where 'recently moderate' (blue) changes towards 'recent rarity' (grey) - the recent event has reduced the chance of a new (even greater) record, however, it should be noted that we do not consider the role of climate change. In a warming world such extreme rainfall events are occurring more frequently. The discrepancy between greatest rainfall and greatest impacts will be influenced by the steep terrain and culverted waterways, which led to rapid surface run-off. It may also be that extreme rainfall over Valencia itself occurred over timescales not captured by the daily timeframe of E-OBS. Close to Valencia, records were broken on 1-, 3- and 6-hourly timescales – though not daily (C3S and WMO, 2025).*

- 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.

This has been altered.

- 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.

We have added in these further factors, highlighting the differing timescales:

Line 370: *We do not account for changes in a catchment, either long term (e.g. urbanisation or bridge building) or shorter term (e.g. dam management, river conditions, and seasonal pre-conditioning), which can influence the timescales of rainfall which will lead to impacts (Mohr et al., 2023).*

- 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.

Indeed this is important. We have added to the text:

Line 384: *We need better understanding of how places adapt when they are not directly impacted but are close to regions impacted, and of the psychology of acceptance that adaptation is required when it has not yet been necessary.*

- L310: 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).

We have added to the discussion to highlight the challenges of fitting a suitable distribution.

Line 390: *Furthermore, the extremes may not fit within the full distribution, requiring a different shaped function to be well represented (Hundecha et al., 2009). Whilst using only observational data we have only a small subset of extreme events and so fitting a distribution create increased uncertainties. The simplified method we use is expected to give comparable results, and it has the added benefit of being applicable to datasets covering larger regions or higher spatial and temporal resolutions where fitting a distribution to every grid point could prove unfeasible.*

In this study we choose to use only observational-derived data. However, in Line 397 we highlight the potential use of large ensembles to better constrain the distribution, and particularly the tails. We cited global models for a range of applications – and we now also cite Ehmele et al. (2022).

Additional references:

Ehmele, F., Kautz, L.-A., Feldmann, H., He, Y., Kadlec, M., Kelemen, F. D., Lentink, H. S., Ludwig, P., Manful, D., and Pinto, J. G.: Adaptation and application of the large LAERTES-EU regional climate model ensemble for modeling hydrological extremes: a pilot study for the Rhine basin, Nat. Hazards Earth Syst. Sci., 22, 677–692, https://doi.org/10.5194/nhess-22-677-2022, 2022.

Hundecha, Y., Pahlow, M. and Schumann, A., Modeling of daily precipitation at multiple locations using a mixture of distributions to characterize the extremes. Water resources research, 45(12), 2009.

- 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?

Detrending the timeseries prior to analysis could lead to events in the initial few years becoming the greatest values – potentially moving events from 'recent rarity' (with high assumed preparedness) to 'fading memory' or 'sitting duck' (with lower assumed preparedness). Therefore, detrending would not be applicable to the method.

Furthermore, we assess preparedness as independent to the changing climate. If we were to further develop the method to incorporate climate change incorporating the trend may be a valuable step to include.

- 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.

We now show HYRAS data for Germany, which is at a 1km resolution – thus have two alternative datasets at two different resolutions. Both produce results highly comparable to E-OBS, therefore we have chosen not to include the 12km HadUK-grid version.