Reviewer 1:

Review of the manuscript “Identifying regions of Europe which may be less prepared for extreme rainfall”

The authors present a quantification of the assumed preparedness to precipitation extremes, based on gridded precipitation observation over Europe. The metric employed is two dimensional, incorporating both the year of the current record - to quantify the memory-related preparedness - and the chance of a new record. The method is straightforward, clearly explained, efficient at highlighting regions of potential unpreparedness based solely on precipitation data, and easily applicable. The paper is overall well written. I suggest publication after addressing the following comments.

Thank you for reviewing our manuscript. We are pleased that you see merits in the study. 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. An additional supplementary figure has been added, showing the 'assumed preparedness' divided across a greater number of categories.

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.

General comments

- Chances of a new record:

While I understand that for the sake of simplicity the authors chose a risk metric solely based on percentile and observed maxima, I recommend improving the robustness of the “Chance of a new record” metric, considering equivalences with extreme value theory, particularly the shape parameter.

In simpler terms, the risk metric evaluates how close Pmax is to P99, relative to the distance between P99.5 and P99, i.e., is the record Pmax “easy” to break. This assesses whether breaking the record is “easy” or “difficult.” With EVT, quantiles beyond the observed data can be estimated, and the “easiness” of breaking a record can be interpreted through the shape parameter of the distribution. For instance, a heavy-tailed distribution implies a higher likelihood of record-breaking events, as illustrated (without the frame of EVT) in your supplementary Figure 1. The manuscript would significantly benefit from framing the analysis within the EVT context and adopting its terminology.

We have chosen not to apply EVT for this study. We do this to aid replicability when working with large datasets, as commented on in the discussion. Furthermore, the extremes may require a different fit to the full distribution – but we are limited to ~75 years of data whilst using observational datasets. As we comment on in the discussion the methods presented could be applied to large ensembles of climate model simulations, in these cases – with many more years of data available and thus more extreme events to fit to – EVT may be better suited.

We have added reference to a similar metric used to assess extremes statistically by ECMWF, the shift of tails (SOT) metric also uses percentiles compared to the full distribution to determine 'extremeness' (Tsonevsky, 2016).

Changes to the text:

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.*

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.

Additionally, it would be helpful to clarify that the occurrence of an extreme during year Y does not decrease the probability of occurrence of an extreme in year Y+1 (i.i.d. assumption). Some phrasing could be misleading "less risk of a record" (l.252). Please ensure the reader understands that a record is difficult to break because the observed maximum is already far into the upper tail of the distribution.

Further clarification that it is not change in the probability of an extreme, only change in probability of a new record-breaking extremes, has been added in several places within the manuscript. We have added 'record-breaking' or 'new record' for each:

Line 313: … *indicating lower probability of a record-breaking event* …

Line 277: *Sofia exhibits the highest likelihood of a record-breaking event, …*

Line 303: …*region would be highlighted due to its higher chance of a new record,…*

Line 327: *Storm Boris set new records, leading to a reduction in the chance of a new record and an increase in preparedness after the recent extreme.*

Line 318: *…the recent event has reduced the chance of a new (even greater) record…*

- Downstream impacts of heavy precipitation:

The impact may extend beyond the location where the precipitation occurs, potentially affecting downstream fluvial regions. Please comment on the added value a similar analysis based on catchments, with maybe some examples.

Indeed this is true – as we highlight for the Valencia floods where greatest impacts were not co-located with greatest rainfall:

Line 353: *Geographical knowledge is key – as in some places impacts from extreme rainfall will not be spatially coherent with the location of greatest rainfall. In some regions adaptation may be driven by the greatest event within a larger (e.g. political) region rather than just locally.*

This has now been expanded, linking to the point the adaptation may occur over administrative regions rather than individual locations:

Line 354: *We assess at gridpoint scale, however, assessment over hydrological catchments could enable identification of non-spatially coherent flood hazard from extreme rainfall. Similarly, in some regions adaptation may be driven by the greatest event within a larger (e.g. administrative) region rather than just locally.*

We also extend the text in the results section:

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.*

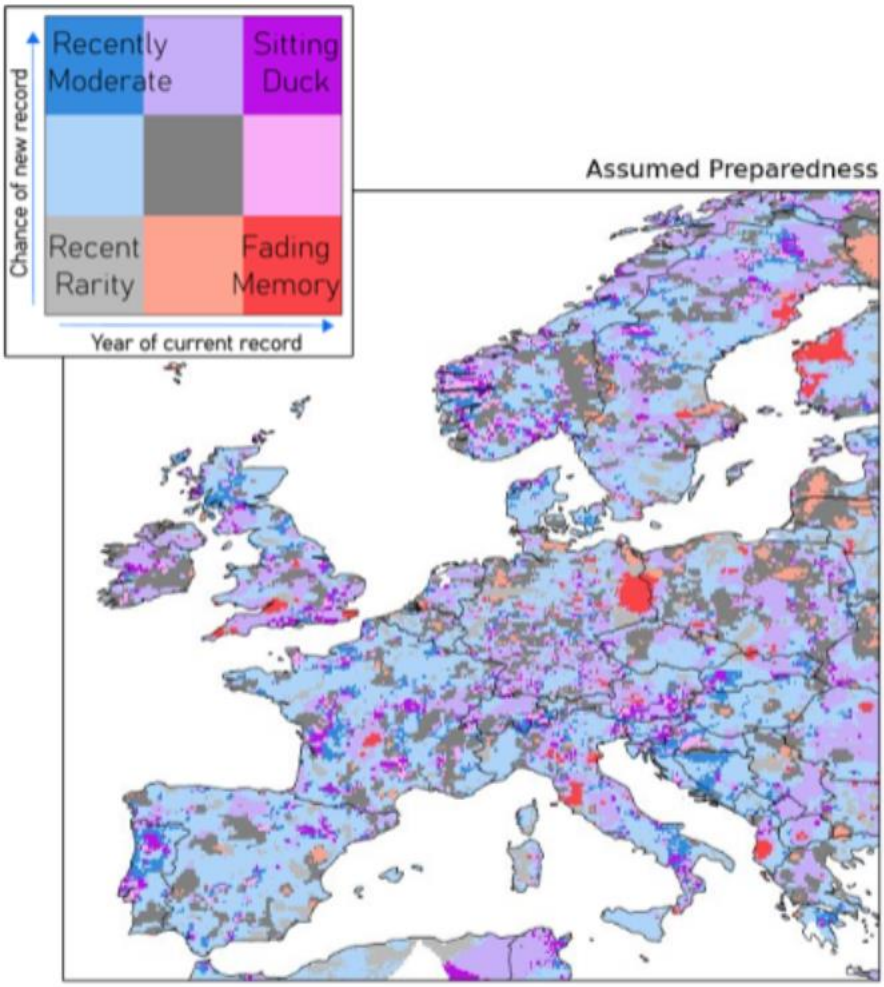
Specific comments

- Fig5: Have you considered visualizing the data with more categories for the "Chance of a New Record" and "Year of Current Record"? This could provide nuanced insights into assumed preparedness and better highlight the most unprepared regions.

We could divide Fig.5 into more categories, and we did compare different options – such as splitting into a 3x3 grid with 9 categories. After assessing these options (including colour options) we chose to show only 4, as this keeps allows the key messages to be portrayed without added complication. For different users more nuanced presentation of results could be considered – with more categories and dividing categories at specific points, perhaps of policy or historical importance. Such additional categories would likely be different for each region or use.

Here we show a version of Fig.4 with 9 categories, how included within supplementary figures and commented on in the text:

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.*

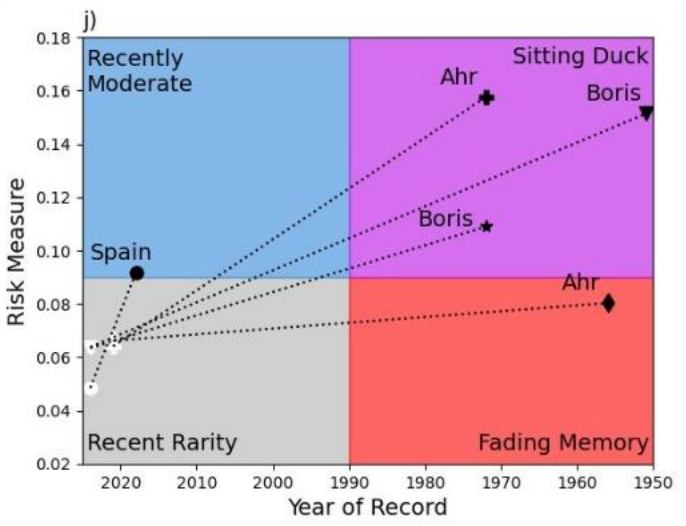


Supplementary Figure 4: Assumed preparedness across Europe. Showing each location classified by the assumed preparedness framework, using E-OBS 1950-2025. Year of record divided into 3 at 1980 and 2000. Chance of new record divided at 0.06 and 0.12.

- Fig7: consider changing the marker color from red to black for readability, and maybe linking the red/black markers to the corresponding white one, with a thin/dashed line.

Indeed, this has improved the clarity of the figure. Fig.7 has been updated, with larger markers for Fig.71-i, black markers in place of red in Fig.7j, and dashed lines linking the black markers to the corresponding white markers. Values have also changed due to a correction in the calculation of the risk measure.

Part of updated figure shown here:



- Is there a quantitative assessment of how societal preparedness improves when recent extremes are considered versus when memory of past events fades? Literature from behavioral and cognitive sciences may offer relevant insights into how recent experiences shape risk perception and preparedness.

This is an ongoing research question that we are keen to explore further. We do include more qualitative examples within our manuscript, e.g. Ahr floods (Ommer et al., 2023) and Japanese tsunamis (Pisa, 2024).

Further research quantifying how societal preparedness is altered by the timing of weather extremes would be beneficial. For example, insured losses or reported damages data from pairs or sets of similar floods could be evaluated. Collaborating with behavioral scientists would be beneficial to such work.

- The use of the term “risk” for the chance of a new record may not be optimal, as the IPCC defines risk as a combination of hazards, vulnerability, and exposure. To avoid confusion, consider using “probability” or “likelihood” instead.

Changes have been made throughout the text, though we do still refer to the 'risk metric' and 'risk ratio'. In some places the term 'risk' is now elaborated on, to make it clear it is the IPCC definition that is intended. Changes include:

Line 93: *Statistically, some locations have higher chance of experiencing record-breaking extreme rainfall,*

Line 239: *The risk metric is used as a proxy for the probability of a record-smashing extreme in a specific location*.

Line 243: *is shown as low likelihood of a record as that event gives a particularly high record compared to the 99th percentile – the areas surrounding appear to be at much higher likelihood*.

Line156: …*and the statistical likelihood of a new record event*

Line 313: …*indicating lower probability of a record-breaking event* …

Line 271: … *suggesting a lower likelihood and higher levels of assumed preparedness*.

Line 266: *Such areas may be particularly at risk of impacts compared to their neighbours*….

Line 277: *Sofia exhibits the highest likelihood of a record-breaking event*…

Line 362: …*identify locations potentially at greater risk of impacts,* …

Line 404: …*locations that might not realize that the chance of extreme weather is changing, having not experienced a recent extreme*.

Line 409: *Reliable and accurate precipitation datasets are crucial to allow us to quantify the likelihood of extreme rainfall*…