Review of manuscript egusphere-2025-3455 submitted to Natural Hazards and Earth System Sciences (NHESS)

Anonymous Referee #1, 28 Aug 2025

[answers in blue]

The manuscript presents reanalyses model simulation results on convective-permitting scales over Italy with hourly resolution. The study is important since understanding precipitation and provides a method of studying events at hourly scales. Extreme precipitation trends on an hourly basis are important for assessing flood risk, especially as these trends are expected to increase with a warming climate, even in drying areas. The paper is generally well written, and the figures are well presented, however there are a few major concerns, along with some minor comments, that I hope the authors will address in order for the article to be suitable for publication.

General comments

A major concern is that the title and abstract of the paper describe extremes, while a lot of the results and figures describes the full dataset compared to the extreme analysis. The Result section should be altered to highlight the extreme analyses better and maybe reduce the description and figures around the full dataset to better suit the journal. Another major concern is that with the weight given on extremes in titles and abstract, there is only one threshold for extremes used which also is set very high, especially when considering hourly data. Above the mean of 37 datapoints which give only 18-19 events per grid cell if the events are evenly distributed around the mean (a quick calculation sets this threshold around the 9.995 percentile). The threshold applied to define extreme precipitation is exceptionally high, and with so few events included, the resulting trend estimates are highly uncertain. The robustness of the study would be improved if additional analyses were carried out using several lower thresholds, allowing for a more comprehensive assessment of trends.

We would like to sincerely thank the reviewer for the careful and thorough reading of our manuscript and for the constructive comments, which will greatly help us improve the quality of the paper.

Regarding the first major concern, we acknowledge that a considerable part of the Results section is devoted to the analysis of the full event-based dataset. Our intention in doing so was to show that the identified precipitation events are consistent with the climatological characteristics of hourly precipitation over Italy, and therefore to provide a solid basis for the subsequent analysis of extremes. Moreover, since we shared this dataset with the scientific community, a complete explanation of it is necessary. Nevertheless, we agree with the reviewer that, given the emphasis on extremes in the title and abstract, the Result section of the manuscript can be better aligned with this aim. In the revised version, we will reduce the number of figures and the descriptive content related to the full dataset, potentially moving some material to the Supplementary Material, in order to dedicate more space and emphasis to the analyses of the subset of events classified as extremes.

Regarding the second major concern, we understand the reviewer's point about the use of a high threshold for defining extremes. It is correct that our threshold corresponds to approximately 18–19 events per grid cell if events were evenly distributed around the mean. However, we may not have emphasized enough in the manuscript that extremes are not selected point by point. Instead, we follow an event-based approach, where an event is defined as a cluster of many points and is classified as an Extreme Precipitation Event (EPE) if at least one grid point within the cluster (the one with the peak value) exceeds the threshold. Combined with the fact that peak value distributions are typically strongly right-skewed with fat tails (see Coles, 2001), this leads to a

selection of EPEs that is less restrictive than it may initially seem. For instance, about 11% of summer events in the full dataset exceeds the average RX1hour threshold. This reflects the choice of a fixed threshold applied to events that vary seasonally: fewer events occur in summer, but they tend to be more intense (see Results, Section 3.1).

We acknowledge that lowering the threshold would increase the number of identified extremes, but we believe this would blur the distinction between EPEs, as described by the Extreme Events Theory (see Coles, 2001), and more moderate high-quantile events, thereby reducing interpretability. Our decision to adopt a fixed threshold based on the mean of annual maxima (RX1hour) was also guided by the approach proposed by Lavers et al. (2025) at the European scale, where fixed thresholds (in that case RX1day) were explicitly recommended for their clarity in communicating extremes.

We also recognize that the relatively small sample size of extremes may raise concerns about statistical robustness. To address this, and to mitigate spatial uncertainty in reanalysis data, we applied a moving window spatial aggregation (0.5° × 0.5°) and employed state-of-the-art trend detection methods: Sen's non-parametric slope estimate, Mann–Kendall significance testing, and field significance analysis with False Discovery Rate control (Wilks, 2006). This multi-method framework supports the statistical robustness of our trend results despite sample limitations.

For these reasons, we believe that our choice of a fixed, high threshold ensures both conceptual rigor and clarity in the identification of extreme precipitation events. We will provide better explanation of our choices regarding this aspect in the revised manuscript.

Coles, S. (2001), An Introduction to Statistical Modeling of Extreme Values, Springer-Verlag, London, https://link.springer.com/book/10.1007/978-1-4471-3675-0

Lavers, D. A., Villarini, G., Cloke, H. L., Simmons, A., Roberts, N., Lombardi, A., Burgess, S. N., & Pappenberger, F. (2025). How bad is the rain? Applying the extreme rain multiplier globally and for climate monitoring activities. Meteorological Applications, 32(2), e70031. https://doi.org/10.1002/met.70031

Sen, P. K.(1968), Estimates of the Regression Coefficient Based on Kendall's Tau, Journal of the American Statistical Association, 63, 1379–1389, https://doi.org/10.1080/01621459.1968.10480934

McLeod, A. I.: Kendall rank correlation and Mann-Kendall trend test, R package Kendall, 602, 1–10, 2005.

Wilks, D. S., 2006: On "Field Significance" and the False Discovery Rate. J. Appl. Meteor. Climatol., 45, 1181–1189, https://doi.org/10.1175/JAM2404.1

Specific Comments

Page 2. L27-29: Even drying areas experience more extreme precipitation events.

We thank the reviewer for this useful remark. We will revise the sentence accordingly to explicitly acknowledge that even regions undergoing drying trends may experience an intensification of extreme precipitation events, as noted e.g. by Donat et al. (2016)

Donat, M., Lowry, A., Alexander, L., O'Gorman, P., & Maher, N. (2016). More extreme precipitation in the world's dry and wet regions. *Nature Climate Change*, 6, 508-513. https://doi.org/10.1038/NCLIMATE2941

Page 5. Section 2.2 Median is easier than the 50th percentile, and used earlier in Introduction.

We appreciate the reviewer's suggestion. We agree that the use of the term median is more appropriate and clearer than 50th percentile. In the revised version, we will therefore adopt the term "median" throughout the manuscript, while keeping the expression 50th percentile only at its first occurrence to clarify the statistical meaning.

The Method section needs to be improved to better understand the results presented. Section 2.2 should be rewritten to increase readability. The thresholds and smoothing are first presented, and then described again in the later paragraph, maybe rewrite for better readability.

Concerning the organization of Section 2.2, we fully acknowledge the need for improved readability. Following the reviewer's advice, we will restructure the section so that the thresholds and their smoothing are introduced and described in the same place, thereby avoiding repetition and facilitating a more coherent presentation.

Line 162 – 165 More specifically, … this sentence in especially hard to follow. More than half in which instances?

We thank the reviewer for pointing out the lack of clarity in this sentence. What we intended to convey is that an event is considered *relevant* when the precipitation exceeds the median of the distribution of events greater than 1 mm within a given area and season. We agree that the current formulation is difficult to follow, and we will therefore rephrase the sentence in the revised manuscript to ensure the definition is stated more clearly and unambiguously.

I'm also curious how time is handled as an event usually lasts more than one hour. Is there any clustering in time?

We thank the reviewer for raising this important point. In our analysis, the general approach was to study events at the hourly scale, regardless of what happens before or after. However, the aspect of event duration is explicitly addressed in the final part of Section 3.3, where we analyze the average persistence of extreme events (Figure 15). This analysis provides an estimate of the mean duration of the extremes as defined in our study and shows that such events typically last only slightly longer than one hour on average. This finding not only clarifies the temporal scale of our definition of extremes but also helps ensure the robustness of the calculated trends by reducing the risk of double counting. In the revised version, we will emphasize both the strengths and limitations of this choice and note that future work could benefit from explicitly applying temporal clustering techniques.

Section 2.4 Is this Gaussian filter the same as used in section 2.2? If so, does the justification of this radius apply to the earlier smoothing, and move this part in section 2.2?

We thank the reviewer for this observation. Yes, the Gaussian filter applied in Section 2.4 is the same as the one used in Section 2.2, and the justification of the chosen radius is identical in both cases. To avoid redundancy and improve clarity, in the revised version we will unify the explanation and provide the motivation in a single place, making explicit reference to it in both sections where the filter is applied.

Table 2, The table could be improved if the names were included in addition to the short names.

We thank the reviewer for this helpful suggestion. In the revised manuscript, we will add an additional column including the full names alongside the short names to improve the readability of the table.

Page 9. L196 ERA5 is already introduced as the driver of the dataset.

We thank the reviewer for this useful remark. In the revised manuscript, we will remove the phrase *"which is a global reanalysis product with a 0.25° grid spacing"* to avoid repeating information already introduced earlier when describing the ERA5 dataset.

Page 10, l207-208 Explain better here: How many events would this give over a typical grid cell on the coast or in the mountains?

We appreciate the reviewer's request for clarification. As already discussed in response to the second major concern, extremes are not selected point by point. Instead, we adopt an event-based approach, in which an event is defined as a cluster of grid points and is classified as an Extreme Precipitation Event (EPE) if at least its peak exceeds the threshold. Moreover, precipitation distributions are typically right-skewed with heavy tails and vary both seasonally and regionally. This leads to a higher number of extremes (compared to the 17–18 events expected from point-by-point selection under symmetric distributions) and to substantial variability across regions (e.g., coastal vs. mountainous areas). This variability is evident in Figure 11. In this analysis the number of occurrences is not computed at the single grid-cell level, but within a 0.5° moving window (≈156 grid points) in order to reduce the impact of misplacement errors. In the revised manuscript, we will make this clearer at the end of Section 2.4 to improve readability and ensure that the methodology is transparent to the reader.

Page 20. L 358-359: This analysis would also benefit from a lower threshold, as two executive hours above RX1hour is extremely rare.

We agree with the reviewer that two consecutive hours above the average RX1hour are indeed extremely rare. However, this rarity is consistent with our methodological choice: by construction, extremes are derived from the mean of hourly annual maxima (RX1hour) and are meant to identify only the most exceptional events, in line with Extreme Value Theory. This makes the occurrence of consecutive exceedances intrinsically unlikely, but at the same time confirms that the analysis truly focuses on *extremes*, and supports the approximation of considering extremes of different hours as independent.

As discussed in our response to the second major concern, lowering the threshold would increase the number of identified extreme events but would also blur the distinction between genuine extremes and more moderate high-quantile precipitation events. From a practical perspective, it would also bring to more multi-hour episodes, requiring the adoption of event-tracking techniques, which falls outside the scope of this work. For these reasons, we consider our current approach the most appropriate for capturing the statistical behavior of precipitation extremes at the hourly scale.

Page 22. L 423-424, This sentence could be misinterpreted, there could be trends here that were not found because of issues with the reanalysis. Could there be a false positive trend, or a false insignificant trend?

We thank the reviewer for this important observation. We agree that the limitations of reanalysis data may indeed have masked existing trends, leading to the possibility of *false insignificant trends*, i.e. trends that are in fact present and significant but cannot be robustly detected due to the constraints of the dataset. What we intended to stress in that line is that the significant trends we do identify remain robust even when accounting for the specific limitations of the reanalysis. To avoid misinterpretation, we will reformulate the sentence in the revised manuscript to better convey this distinction.