Response to Reviewer 1

Brief communication: Contextualizing the September 2024 extreme precipitation in Austria within the climatological record

Responses to reviewer comments coloured in blue.

General comments

The paper presents a statistical characterization of the extreme event happened In Austria in 2024. The applied methodology seems correct and reasonable (I have a few comments on that, listed below), and results moderately interesting.

We thank the reviewer for the assessment and useful comments. We respond point-by-point below with suggestions to improve the manuscript.

However, my main concern is on the relevance of this work with respect to the NHESS Brief Communication scopes, as reported in the journal page: "These may be used to (a) report new developments, significant advances, and novel aspects of experimental and theoretical methods and techniques which are relevant for scientific investigations within the journal scope; (b) report/discuss significant matters of policy and perspective related to the science of the journal, including "personal" commentary; (c) disseminate information and data on topical events of significant scientific and/or social interest within the scope of the journal." Maybe the paper could be in the (c) case, but I suggest to the author to more clearly state the relevance of their work, now it seems to me a case study with limited new added knowledge.

We appreciate the opportunity to clarify how our manuscript aligns with the NHESS Brief Communication scope. We agree that the work fits under Criterion (c) (*disseminating information on topical events of significant scientific/social interest*), but it also contributes to Criterion (a) (*advances in experimental/theoretical methods relevant to the journal's scope*) by offering insights that extend beyond a localized case study.

Below, we outline how the study addresses these specific scopes:

1. Alignment with Criterion (c): Disseminating data on a topical event of high scientific and social interest

- The extreme precipitation associated with Vb-type cyclone 'Boris' in September 2024 is of climatological and societal relevance. It led to widespread severe flooding, landslides, and flash floods across northeastern Austria and neighboring countries. The exceptional severity and unprecedented nature renders this a topical event of high scientific and social interest.
- This unprecedented event was a major outlier, with a third of the study area setting a new 5-day maximum precipitation (`Rx5day`) record compared to the historical baseline since 1961. Roughly 7% of Austria experienced totals at least 50% above previous records. Disseminating these return periods and observational statistics is critical for updating regional hazard models and design thresholds.

2. Alignment with Criterion (a): Methodological Advances in Extreme Value Analysis

We argue that rather than presenting a simple local case study, this manuscript addresses an interesting and fundamental methodological challenge in natural hazard research: how to estimate return periods for unprecedented extreme events when standard historical data-series under-sample the far tail.

- We employ a robust POT approach using a spatial Bayesian hierarchical framework with regional pooling of the GPD shape parameter.
- We show how regional pooling dynamically 'borrows strength' from neighboring stations to stabilize extreme-value distributions and reduce uncertainty. This methodology allows us to transparently propagate uncertainty, preventing short-record stations from introducing statistical noise while still leveraging their observed extreme signals.

- Our model successfully confirms a stable, heavy-tailed Fréchet regime, offering a robust framework that other researchers can apply to similar catastrophic events worldwide where frequentist methods can yield vast uncertainties making them less useful.

Specific comments

Here the list of specific comments.

Line 11: briefly mention what “Vb storm track” is

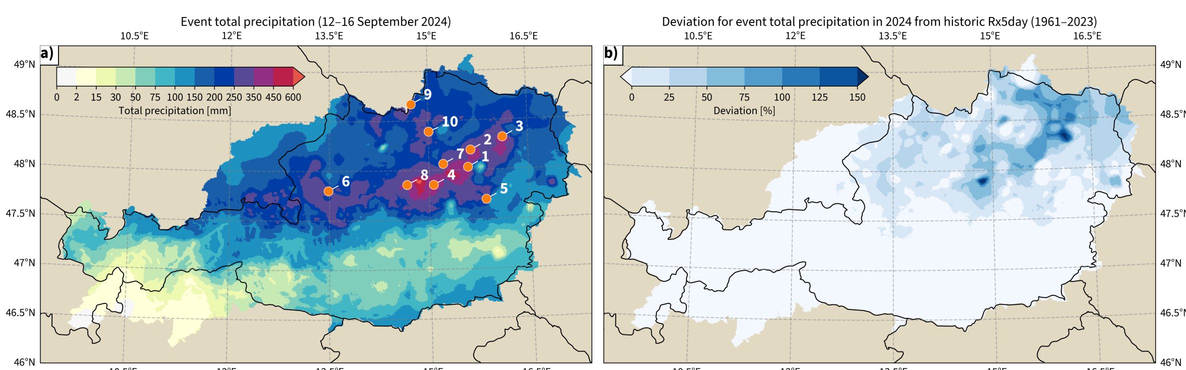
The Vb storm track refers to the characteristic path of cyclones originating over the Gulf of Genoa and moving northeastward toward Hungary and then further into Central and Eastern Europe. We will rearrange and slightly rephrase the text to make this connection more explicit and improve readability. The revised sentence reads: *”The system originated as a surface low over the Gulf of Genoa and progressed northeastward towards Hungary and further into Central and Eastern Europe, thereby closely following the classical Vb storm track (van Bebber, 1891; Messmer et al., 2015; Hofstätter et al., 2018), which is well known for producing extreme precipitation in northeastern Austria (Hofstätter et al., 2016, 2018)”.*

Line 30-31: which return period? Then at line 149: Different from your estimates?

The return periods mentioned refer to the analysis of Kimutai et al. (2024). We will improve the wording to clarify that our results confirm the events’ extreme nature shown in Kimutai et al. (2024), while demonstrating the added value of spatial pooling for narrowing the confidence intervals. The revised sentence in the introduction reads: *”Return periods of total precipitation of the September 2024 event have been estimated using a non-stationary extreme value model based on block maxima by Kimutai et al. (2024), but uncertainties inherent to the standard 100-year time-series limit the conclusiveness of these estimates”.*

Line 49: why using the Rx5day2024 and not the event one? I mean, this way you are characterizing the 2024 maxima, not specifically the event, as also mentioned in the first sentence of figure1 caption (for some cell, the 2024 maxima happened during the September event, for others, no)

We agree that the event-specific precipitation total is more appropriate for this figure than the annual Rx5day2024 statistic. We will therefore revise the figure to show the accumulated precipitation during the event period (12–16 September 2024) instead of Rx5day2024. While the overall conclusions remain unchanged, we agree that this representation more accurately reflects the purpose of the figure. In addition, we updated the analysis to use the recently released SPARTACUS v3 dataset instead of v2.1, which was not available when the original manuscript was prepared. Compared to v2.1, SPARTACUS v3 incorporates a larger number of homogenized station observations, resulting in improved data quality and making it more suitable for the climatological analyses. The SPARTACUS v3 dataset is available on the GeoSphere Austria DataHub at `https://data.hub.geosphere.at/dataset/spartacus-v3-1d-1km`. The main conclusions remain the same. The updated figure is shown below:



Line 58: what st indicate in Yst ?

Thank you for pointing this out. This was a typo and the correct term is Ys, which denotes the station-specific exceedance.

Line 64 and line 98: this result is based on just 4 stations. Why not using all? Is this valid also for the more distant stations such as 6-9-10?

We thank the reviewer for raising this point. The restriction of the tail dependence test to the four stations with series lengths exceeding 50 years is a direct consequence of evaluating extreme dependence at the 99th percentile. Evaluation of χ at $u = 0.99$ requires a sufficient number of concurrent exceedances across station pairs. For short-record stations the number of joint exceedances is very small, introducing small-sample estimator bias and large variance. To prevent these sampling artifacts from distorting the regional tail dependence inferences, short series were excluded. We will rephrase the sentence on lines 98-99 to make this rationale clearer. The revised sentence reads: *"Tail dependence diagnostics at the 99th percentile for the four longest records (≥ 50 years; stations ranked 2, 3, 5 and 8 in Tab 1) to ensure sufficient joint exceedances indicate significant extremal correlation* $(0.3 \leq \xi \leq 0.55)$*, strongly supporting regional pooling of the GPD shape parameter."*.

Line 65: bias on what? And how much is that bias?

We agree that the use of the terms 'bias' and 'variance' in this context was not sufficiently precise. We will therefore rephrase the sentence, and remove these terms to avoid ambiguity. The revised sentence reads: *"Thresholds were selected to retain sufficient sample size for parameter stability without violating the asymptotic assumptions of the GPD tail."*.

Line 67: which others non-pooled baseline models? Mention or show their performance

We performed leave-one-out cross-validation (LOOCV) for both the pooled model (with a shared shape parameter) and a non-pooled model (stations with specific shape parameters). The LOOCV was conducted at the observation level, resulting in `N = length(all_exceedances)` validation members, and was designed to assess how well each model predicts an unseen exceedance at a station with an existing observational record. The LOOCV results showed that pooling improved overall fitting stability across the existing station network. We will add these details into the methodology section for the extreme event analysis. The added sentence reads: *"(iii) A leave-one-out cross-validation on observation level for a pooled and a non-pooled model with station specific shape parameter showed that pooling improves overall fitting stability across the existing station network."*.

Lines 87-92: I find not coherent numbers and explanations between eq.2, these lines, and figure 1b. For example, when you say "more than double it", do you mean delta (eq.2)=100 ? But deviations in figure 1b > 200. I suggest to use same metric in the text and figure.

Thank you for pointing out this inconsistency. We have revised Figure 1b to ensure that it reflects the definition of δ given in Equation 2. In the previous version, the value of 100 % shown in the map represented the historical reference value, while Equation 2 defines the historical reference as 0 %. This inconsistency was purely numerical and does not affect the pattern or the drawn conclusions. In the revised manuscript, the text and figures will be harmonized and consistently follow the definition of δ in Equation 2, where a doubling of total precipitation corresponds to $\delta = 100$ %.

Lines 110-125: I think that too many lines are dedicated to the explanation regarding the inclusion of the two stations with very short length. Also ... adding them "leaves the shape parameter unchanged" (line115), so how they "help stabilize the join EV distribution" (line 119) ? what if you don't include them?

Thank you for this comment. We understand the concern regarding the length of this explanation. However, because the inclusion of very short records is unconventional in standard extreme-value analysis, we believe a detailed justification is necessary to reassure readers of the methodological rigor. To clarify the specific mechanisms and address your questions:

- How they stabilize the joint distribution while leaving the shape parameter unchanged: The shape parameter is shared across the network and is robustly determined by the longer-record stations; adding the short records does not distort this regional parameter (demonstrating model robustness). However, because these short stations captured the extreme event coherently with neighboring stations, including them provides additional spatial replicates of the event. This stabilizes the joint distribution (specifically the scale parameters and spatial correlation across the network) rather than shifting the regional shape.
- What happens if they are excluded: Excluding these stations would discard real, observed data of a major extreme event. This would artificially reduce the spatial footprint of the event in our dataset, leading to weaker spatial inference and a loss of valuable information.

To make these points clearer and more concise for the reader, we will revise and streamline the text in Lines 110–125 as follows:

However, the shortest records, Schwarzau (3 years) and Unterach (6 years), benefit less from this stabilizing effect. Even under a POT framework, effective sample sizes remain small. Return-period estimates inferred from these short series should therefore be treated with caution and regarded as indicative ranges, with relative rankings being more defensible than their absolute magnitude. Nonetheless, including very short records is mathematically justified within a spatially polled Bayesian hierarchical framework. Adding these series leaves the shared shape parameter essentially unchanged, demonstrating robustness, while the scale parameters and station-level return periods appropriately reflect the associated uncertainty. For example, at the three-year Schwarzau station, the posterior return period of 12 years is accompanied by a very wide credible interval (3–162 years). Crucially, because the extreme event occurred coherently across stations, these short records provide genuine signal to the pooled tail. This stabilizes the joint extreme-value distribution and reduces posterior return period estimates and their uncertainty across the network, whereas excluding them would discard observed extremes and weaken the model's spatial regularization.

Line 133: not clear to me what do you mean with "the most extreme intensities can decouple from the regional mean at the distributional limits due to localized effects and hydrometeorological regimes". Which localized effects and hydromet. regimes?

We revisited the explanation in that paragraph and included specific local effects and hydrometeorological responses that were analysed in Komma et al (2025). The revised sentence reads: *"Specifically for the September 2024 event, Komma et al. (2025) showed that the extreme response resulted from a complex interaction of meteorological and catchment processes. Mesoscale convective cell regeneration in the Alpine foreland and sharp elevation-dependent thermal contrasts, which limited snow storage to high Alpine catchments, interacted with rapid soil saturation thresholds and short-duration, high-intensity convective bursts."*.

Line 134: which kind of consistency you are referring here?

Thank you for this comment. We will rephrase this sentence to remove ambiguity and directly address what we mean with consistency, which here links to the hydrometeorological responses discussed in Komme et al. (2025). The revised sentence reads: *"Such localization of extreme peaks under regional synoptic forcing physically reflects the interplay of tributary synchronization and local hydrological controls, rather than a uniform regional flood response (Komma et al., 2025)."*

Line 145: stable within 0.15-0.25 ... but before you mentioned stability around 0.22 at high percentile.

We agree this is a potential source of confusion for the reader. We will remove the two statements 'ξ stable within 0.15 - 0.25)' and '($\chi > 0$)', as they add no value in the conclusions and are described within the main text in more detail already.

Figure 2: showing the time series on the left seems a quite basic plots; maybe better to move to supplementary: Caption mention the "overall goodness of fit": have you assessed it or this is a by-eye evaluation?

We agree that timeseries are indeed a very basic visualisation, however, we think it helps communicate the extraordinary nature of this event compared to the long station records, which helps contextualize the high return periods as well. Therefore, we would prefer to keep the Figure as is. The overall goodness-of-fit was assessed by visual inspection. We will update the figure caption accordingly for clarification.

Line 156: what do you mean with "accounting for local controls"? and how you demonstrate its importance with your work?

Thank you for pointing this out. While our study demonstrates the benefits of pooling, it does not explicitly test or prove the influence of local control variables. We will rephrase the sentence to more accurately reflect the specific conclusions supported by our analysis.