Response to comment 1

We would like to thank the reviewer for their thorough and positive review of our paper. The comments are thoughtful and very useful for improving the quality of the manuscript. We will respond to all comments below, with our responses in blue, and the comment in black. Suggested changes to the manuscript are in *italics*.

The technical note Temperature dependence of precipitation tail heaviness in the TENAX model by Thomas et al presents an investigation of the possibility of including a varying shape in the TENAX model by Marra et al. (2024). The note investigates potential consequences of the inclusion of the varying shape and probes some possible modelling choices. The note is quite interesting and it explores a very relevant topic in the modelling of extreme precipitation: changes in the shape of the distribution of natural hazards might result in dramatic changes in risk levels and as such the topic is very relevant (indeed I think a recent interesting exploration of this topic should be at least mentioned by the authors https://doi.org/10.1029/2023WR036426). The note is a bit crude in some of the analysis, it could go a bit deeper in some explorations and could present some more in-depth statistical analysis, but I imagine this is why it was submitted as a technical note and not as a research article. It is clearly an initial step towards a larger change in the TENAX approach, which I think will be an interesting development. Overall I think the authors did a very good job at presenting the material, the aim is clear, the analysis are to the point and the methods relatively well presented. This being said, while reading the manuscript I noted down some comments which I hope the authors will consider to possibly improve the presentation of the material.

Thanks for the additional citation recommendation, we will add a reference in the introduction.

Major comments:

1. "that is, how much larger extremes are with respect to the average." I don't love this definition, since it sort of defines the tail itself rather than the heaviness (which, especially for the Weibull, is often defined in relation to the exponential tail). While this is not a major comment, I do think the authors could help the reader understanding better what the changes in the tail/shape parameter might entail by providing some more intuition/explanation on the meaning of the shape parameter in the Weibull distribution, especially when this is then allowed to vary. Maybe some pictures of a fitted magnitude model might be helpful in appreciating the implications of the increase or decrease of the shape parameter?

Thank you for the suggestion, we will add a figure like figure R1 below to aid understanding and edit the text. "*The tail heaviness impacts the likelihood of extremes within the distribution*".

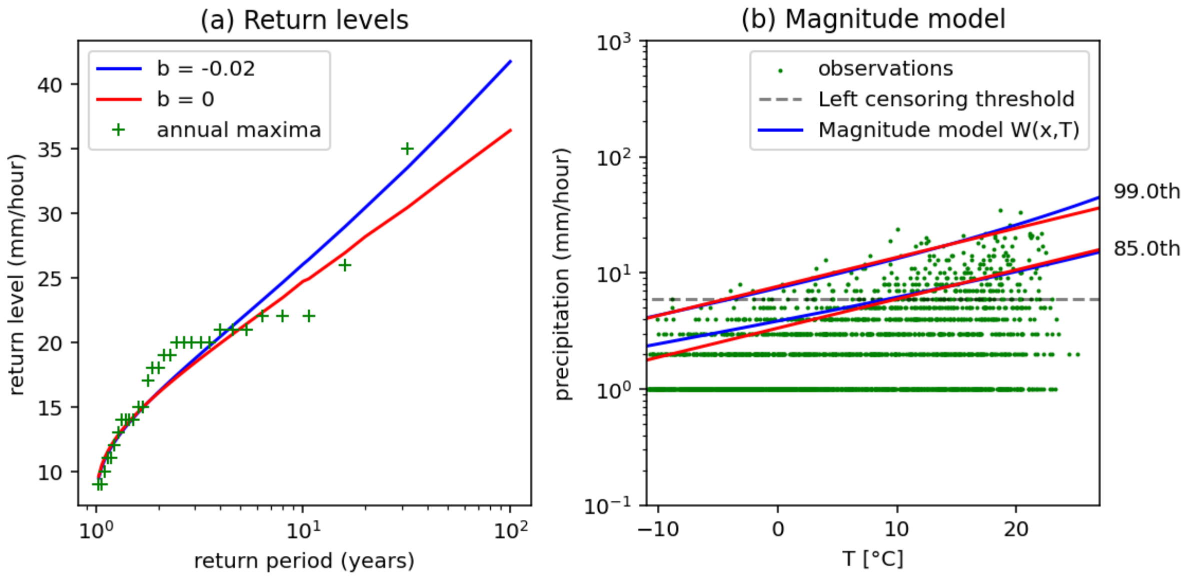


Figure R1: Example to show the impact the shape parameter has on tail heaviness and the resulting return levels from a station in north Japan (45.25, 141.85). Panel (a) shows the return levels, with shape parameter dependence on temperature in blue. No dependence on temperature in the shape parameter is in red. Panel (b) shows the magnitude model, the relationship between temperature and extreme precipitation.

We will refer to this figure in the introduction: "*For example, Fig. 1 shows the application of TENAX to rainfall data at a station in Japan. Allowing the shape parameter to depend on temperature results in the blue line, with larger return levels (Fig. 1a)*, while *the percentiles in Fig. 1b show that this dependence causes rainfall rates at higher extreme percentiles to increase faster than those at lower percentiles.*"

2. Line 90 (and throughout the manuscript): the shape parameter is assumed to be positive, I imagine this is why others have enforced this assumption by using the exponential transformation. When using a linear model, do you do anything to enforce a positive shape? Could the linear model not be problematic if one wished to use the fitted model using for example climate projection in which one might need to evaluate the model for a range of T values such that $\kappa(T) < 0$?

It is possible that the linear model would cause issues if the shape parameter became negative. In general, such variations are found to be small enough that this is not a practical possibility. However, if temperature changes were larger, a different model would need to be used and the exponential dependence examined here becomes a viable option. We will add a clarification in the methods. "*In principle, the linear dependence raises the possibility of a negative shape parameter if temperatures change* *considerably. However, our analyses showed that temperature changes* are not *large enough for this to become an issue in real cases. More in general, the exponential dependence prevents this from happening."*

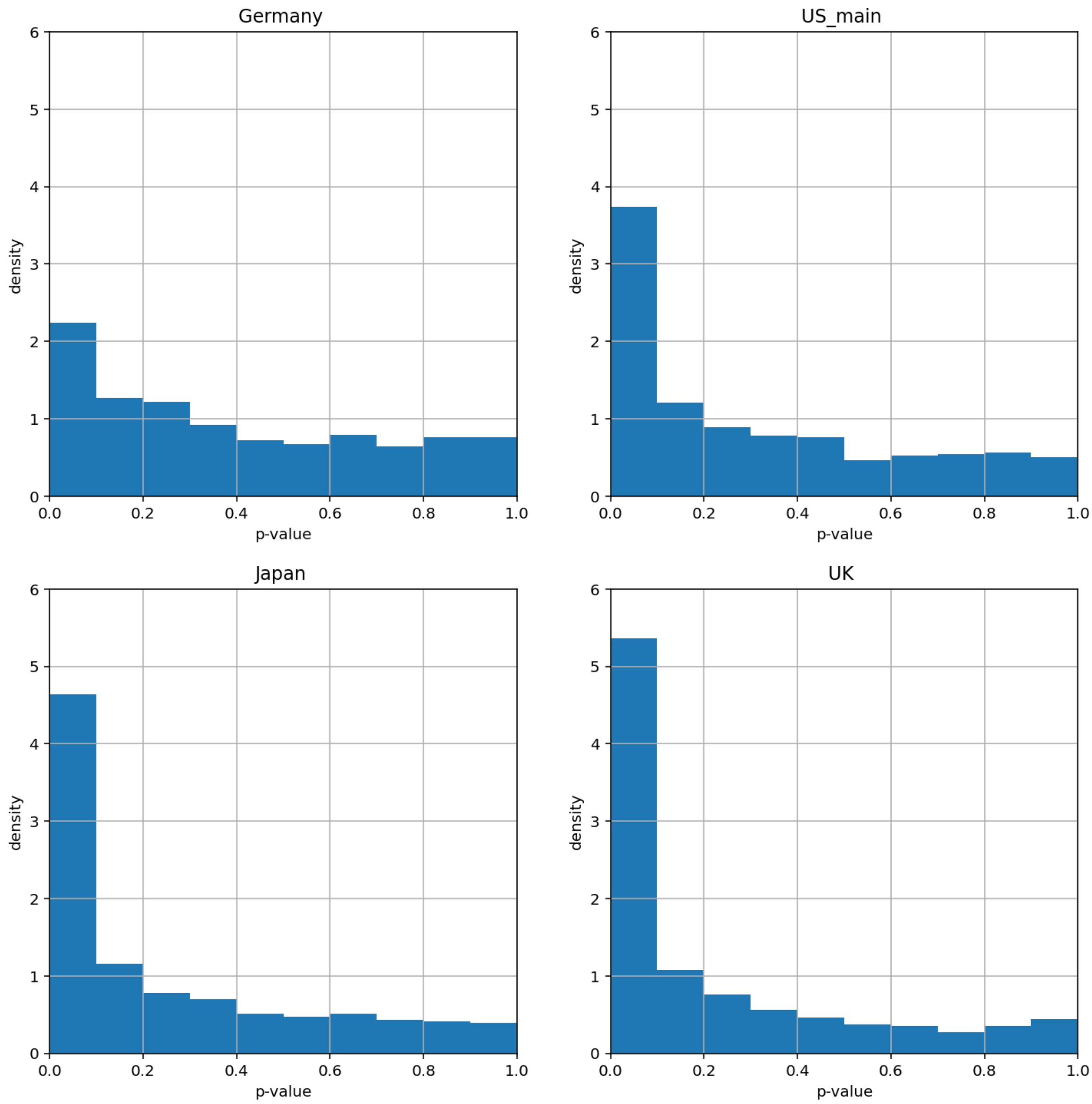
3. I found the explanation on line 118-119 quite hard to follow. You generate 5-times the number of events at each station (if so why?) or five-times the number of stations (you have one set of parameters per station and you calculate the parameters five times?)? Also I think a better wording would be that you "approximate stochastically the sampling distribution of the parameters"; I found "stochastically calculate" quite unclear (you estimate, no?). Indeed, line 120 "The distribution of these parameter" *estimates*

Thank you, we agree this was unclear. We generated 5 times the number of events at each station to improve the distribution. We will edit the text to: "*For each station in a country, we simulated precipitation and temperature events equal to the number of observed events at that station. We used the average observed parameters over the country and repeated this with $b=0$ and with b allowed to vary. From these simulated events, we recalculated the parameters*, *repeating this process five times to reduce sampling uncertainty and obtain robust distributions. Five is an arbitrary number which was chosen as a balance between computation time and stability of the results."*

4. Line 145: while I think it is OK to simply show the at-site information on the detection of the trend at this stage, I think the authors should acknowledge the issues linked to multiple hypothesis testing and field significance (see for example https://doi.org/10.1029/2021WR030172 or https://doi.org/10.1029/2007WR006268, but the topic is very present in the literature on trend detection). Also, I have found it quite useful for this type of trend detection to see the histogram of the test statistics derived for each station: how far away is this from the standard gaussian distribution (it is a cruder way of assessing the overall signal)?

To make these conclusions, we assume that the stations are independent. Of course, stations close to each other are unlikely to be completely independent. Unfortunately, there is not a clear way to resolve this, but we agree that the issue should be better underlined. We will add a sentence discussing these issues. Thank you for the references. In results we will add: "*We should note that these statistics are not completely reliable since the stations are unlikely to be completely independent which brings issues of multiple hypothesis testing.*"

The histograms of the p-values in each country are reported below and tend to support our conclusions.



5. Line 173: it is not entirely clear to me why the likelihood of outliers appearing should increase. The total number of "outliers" might increase, but why the likelihood? Also, I find the term outlier not ideal: these are situations generated by the controlled data-generating process, they are simply part of the tail of the distribution, not really outliers. Of course, they complicate the story, but they are part of the possible outcomes. Also, I don't fully understand/agree with the sentence "The increase in outliers shows that the number of events at a station affects the uncertainty in estimating the parameters." First of all, I don't fully understand how this is related to the number of events at a station. More importantly, the fact that larger sample sizes result in more precise estimation is surely not newsworthy, just like the fact that maximum likelihood is not robust (to outliers, but I think you mean something else here).

Since we are re-simulating the events to get the parameters multiple times, we are more likely to see large or small simulated parameters in the simulated ones than in the observed. We agree that these are within the tail of the distribution, and we will rephrase accordingly. We will remove the second part of the sentence, as we agree it is unnecessary.

6. Line 194: is the test used to carry out this check the LRT test mentioned in line 127? Make this more explicit if so (and actually, more details in Section 4 could be beneficial to appreciate how the testing is done). I admit I did not fully understand Figure 4 and the discussion around it, it is not clear to me what the "significantly different in one case only" category means: you are testing all the parameters? What is the one case? I think the discussion of the Figure could be improved.

Yes, the test is the LRT test mentioned in the Analysis methods. There is more information on this test in Marra et al., 2024, which we will direct the reader to in the methods. "*(see \citet{marra_predicting_2024}, section 3.2 for more details on this test)*" In section 5.3, we will add a reference to the use of an LRT to clarify this: *"We do this with a likelihood ratio test."*

The "significantly different in one case only" category means the magnitude model for all parameters is significantly different between the two time periods either for b=0 OR b=free (but not both). We will improve the figure by changing the blue label to "*either b=0 or b=free significant*" and the caption to "Maps showing where differences in the magnitude models are significant (red) or not significant (yellow), regardless of whether \$b\$ is fixed to zero or allowed to vary freely. *Blue points are locations where the magnitude model for all parameters is significantly different between the two time periods either for b=0 OR b=free (but not both)*"

7. Line 209: I am not surprised to hear that the estimation of the temperature-varying parameters is challenging and that the estimated functions tend to compensate each other. Could one think of models in which the change in shape and scale is somehow interlinked (so estimating a unique slope, modulo some constant)?

Thanks for this suggestion. However, we are not sure such a constraint would work in the case of TENAX, at least conceptually, because the dependence of the scale parameter on temperature is ultimately designed to account for thermodynamic processes (and related feedback), while the potential dependence of the shape would likely reflect dependences in the dynamics. Linking the two may bias some of the existing relations between physics and statistics in ways that are difficult to interpret, at least in the current state of our understanding. We believe in such case it would be preferrable to set the b parameter to a fixed value (or even to zero).

8. Line 274: I think the idea of a regional trend is a good one, there is plenty of literature on this carried out for similar applications

(https://doi.org/10.1016/j.advwatres.2021.103852 or https://doi.org/10.1029/2005WR004591 for example).

Added in: "If there are enough stations with sufficiently long measurements in a region of interest, b can be set to the average calculated value of b in the region, *following methods such as in \citet{kjeldsen_assessment_2021} or \citet{renard_application_2006}*."

Some other minor points:

Line 20, when mentioning the event in Italy: any references? You provide refs for the other events.

We decided to remove the reference to this event so we are only talking about the countries studied in the paper.

Line 98: a Monte Carlo approximation with $N = 2·10^5$ iterations is used *to approximate it* (or something to make the sentence clearer)

Will change the text as suggested.

Line 109: ERA-5 land cells are indeed cells, so their coordinates are representative of the cell. I don't remember (I always need to double check) if the cell coordinates are the centre of the cell or one of the corners, but I am a bit puzzled by the sentence.

This issue was an artifact of using ERA5-land data; some stations weren't within the land mask. This may have been because they were coastal or on small islands. We will add a line clarifying this.

Line 123: make clear that what is assumed to be time-invariant is the dependence on T, and that this implies that the physical relationship between temperature and precipitation is constant. It would be possible to create a TENAX model in which the parameters change with time, so the choice of investigating whether the parameter estimates change in time as a check for the validity of the model is something you should motivate (I agree that it is a sensible check). One more small point on this: can some of the differences in the estimation of the parameters in the early and later period not be also due to the difference in the distribution g(T) in the different time periods?

We will emphasise the point that the physical relationship is expected to remain constant. "As per TENAX's assumptions, the parameters of the magnitude model are required to be time-invariant. *This is motivated by the idea that the magnitude model represents the physical relationship between temperature and extreme precipitation which is not expected to change with climate change.*"

From a theoretical standpoint, the magnitude model is completely independent of the temperature model g(T).

Line 137 (and Figure 5): if you want to call the quantity you compute Bias, it should at least be Relative bias (Bias would be estimate - true). Honestly, I would simply call it ratio of estimated against true value of return levels, since you also discuss the variability of the estimates when discussing the result, so really, you are interested in the Bias/Variance tradeoffs. Also in the Figure: I would maybe delete the long y-axis label from panel b and c (and only leave it on Panel a) to keep the left sides a bit less cluttered. Also in the discussion of Figure 5 I think you should very clearly say that panel a shows what the risk of ignoring trends in the shape is in terms of underestimation. Given that you mostly found negative trends in the countries you have studied, this is a very consequential fact.

We were trying to find a way to keep the title on the figure short/simple. We will change the name to *"multiplicative bias"*. Thank you for the additional recommended improvements to the figure, we will implement these.

In the discussion of figure 5, we will add: "*This, along with the fact that our observed values of b are usually negative, show that ignoring trends in the shape parameter risks underestimating return levels*".

Line 142: just write p-value, you can use the extra 4 characters!

Done.

In Figure 2: if I understand correctly there are no boxplots to show for the lower panels of the right columns. Having those little bars made me wonder if I was looking at very precise estimations: I would leave them blank or put a cross (or a different symbol). Also: what are the black lines in the plot: make explicit in the caption or legend.

Yes, you are correct that there are no boxplots for those panels. We will edit the figure to make this clear. The lines are the minima, mean and maxima. We will make this explicit in the legend.

Line 167: I guess it is not very surprising to see that all the variation can not be explained by sampling errors, surely stations are affected by the same climate and are not independent of each other. There is a plethora of studies on this, and the behaviour shown in the Figure is to be expected: this fact is the reason behind the existence of spatial interpolation and statistics.

Thank you for this comment. We believe this is something worth mentioning because, in principle, the spatial variability of precipitation may also be determined by spatial variability of g(T) with no change in W. On a higher level, the Clausius-Clapeyron equation is invariant in space, yet we observe spatial variability in precipitation statistics. We believe it makes sense to ask ourselves how much our parameters are invariant in space.

Line 188: b is always included in the model, the point is that it is included *as an unknown parameter which needs to be estimated*

Added this in.

Line 249: "We conclude" -> this is maybe a bit strong, you don't give very much evidence for this. "This seems to indicate that ..."

Changed to suggestion.

Line 295: I imagine one could expand the recent work on Bayesian Spatial model on this type of models (https://doi.org/10.1007/s13253-025-00719-0)

We agree, and there is ongoing work in this area.