Dear Reviewer,

Thank you for this constructive review. Please find our responses below (**in bold**).

Best regards,
Anna Luisa Hemshorn de Sánchez (on behalf of all authors)

Review of Sensitivity of mean and extreme streamflow to climate variability across Europe, submitted to HESS.

The study presents an empirical analysis of the sensitivity of mean, maximum and mininum annual streamflow to mean annual precipitation and temperature, performed by using a European dataset comprising several thousand river basins. It also tries to link the identified sensitivities to catchment descriptors.

The manuscript is generally well-written (except for recurring typos - at least I believe them to be so - that at times make understanding the meaning of sentences difficult; see minor comments for details). However, it seems to contain little novelty besides the use of a large dataset (HESS journal evaluation criteria #2: Does the paper present novel concepts, ideas, tools, or data?). The use of linear regressions to relate streamflow to precipitation and temperature metrics and estimate sensitivities (elasticities) is not new. The use of random forest analysis "to reveal the underlying physical processes" is also not new (and the ability of such an analysis to actually identify "processes" is debatable).

While we respect this opinion, we would like to emphasize that Reviewers 1 and 2 acknowledge the novelty of our work and that our manuscript is novel in that:

- **it goes beyond the commonly studied sensitivity of mean streamflow to climate by considering maximum and minimum annual and seasonal flows,**
- **it gives a pan-European overview of this, and**
- **while answers are not definitive, it provides linkages with a wide range of catchment properties.**

In addition, I would like to raise the following three points.

We address these points individually below and thereby strengthen the manuscript. The main changes that we will implement in the revised manuscript are the following:

- **Emphasize the contribution of our work more clearly and revise our methodology to select and validate catchment descriptors considering the suggested review Tarasova et al. (2024) (e.g. adding cross-validation to the random forest model).**
- **Set a minimum station density for the E-OBS climate data to reduce related uncertainties.**
- **Clarify the use of normalized variables further to avoid confusion.**

1. The authors state among the motivations of this analysis "the limited understanding of what shapes spatial differences in streamflow sensitivities" (line 75-76) and the fact that "while several studies link elasticities to catchment characteristics, these links remain uncertain" (line 59). Their approach to addressing this issue (i.e., applying random forest analysis to a set of catchment descriptors), however, has been used widely before and indeed it does not produce new knowledge, as the authors' themselves reported in the abstract (lines 21-22: the fact that the filtering effect of catchments is controlled by combinations of catchment properties is known). As it is, this part of the study reduces the value of an otherwise interesting (although not especially new) empirical analysis.

   We understand the concern but studying the elasticities of mean, maximum, and minimum flows and how they are related to a wide range of catchment characteristics beyond climate characteristics as presented in this study has not been done in this form before, to our knowledge. A valid result of this analysis can still be that linkages with catchment properties still need further development. In the revised manuscript we will better emphasize the contribution of our work and the motivation for choosing the method and catchment descriptors.

   A recent review (Tarasova et al. 2025) provides some ideas on how new insights may be gained by works that use catchment descriptors. Apart from the derivation of new more informative catchment descriptors, which may not be viable for this study, other suggestions may instead apply, like for example i) hypothesis-oriented selection of catchment descriptors, ii) the derivation of functional catchment descriptors (notice, e.g., how several descriptors with high feature importance relate to how catchments use their storage), iii) cross-validation to test the actual predictive power of descriptors.
   It is also recommended to clearly explain how catchment descriptors are selected. The current explanation ("we choose those that can be physically connected to elasticities", line 175) is neither clear nor exhaustive.

Thank you for referring to this review, which is indeed very interesting and relevant to the manuscript. We agree that incorporating suggestions from this review into our revised manuscript will strengthen this part of the analysis. In the revised manuscript we will include more information on how we selected catchment descriptors and revise the selection where needed to be more hypothesis-oriented. We will also add a cross-validation of the random forest model in the revised manuscript to test the predictive power of the descriptors. We will further explore the feasibility of deriving functional catchment descriptors with the available data, as following the given example of Janssen and Ameli (2021) in Tarasova et al. (2024) would require data on the long-term median water input intensity, shallow soil hydraulic conductivity, depth to bedrock, soil porosity, slope, soil thickness, soil-bedrock conductivity contrast.

2. Climatic data are taken from the EOBS dataset, which is the result of data interpolation. Does the interpolation influence the spatial variability of elasticities discussed in the paper (i.e., are we seeing real spatial variabilities, or spatial patterns produced by the interpolation method)? I ask this because two out of four key limitations of the dataset (Potential inhomogeneities in the input stations records may lead to spurious climate signals; Artifacts from the statistical interpolation method may occur in areas with very low density of stations (e.g., circum-Mediterranean, and eastern Europe), reported in the Quality information of the dataset) pose some concerns in this regard.

   To reduce the uncertainty related to low station density, we will set a minimum station density of E-OBS as a criterion for the catchment selection for the analysis of the revised manuscript. We will state this procedure and remaining uncertainties more clearly in the revised manuscript. We only use the E-OBS data to get annual mean P and T, which reduces the impact of this uncertainty.

3. The rationale for using absolute temperature in the regression instead of normalized values as done for precipitation and streamflow is not very convincing (although the choice may be legit). In particular, it is not clear to me why the motivation given at line 114 (also referred to in the authors' reply to Referee #1 on this issue) should matter: as for streamflow and precipitation, the mean annual temperature of different years would be normalized by the long-term mean annual temperature. Hence, the reference to zero degrees being an arbitrary reference point seems out of context.

   The existence of an absolute zero reference point is crucial when normalizing variables. Here is a short example: A streamflow of 20m^3/s is twice as big as 10m^3/s. With a long-term mean of 5m^3/s the normalized value of 4 would also be twice as big as 2. With temperature

(°C) instead the zero value does not represent an absolute zero point but a point based on the freezing point of water at standard atmospheric pressure. Therefore, 20°C is not twice as hot as 10°C but 10°C warmer. Normalizing with a mean temperature of 5°C would result in values of 4 and 2 that would give the wrong illusion that the first temperature is double as warm as the second. In the revised manuscript we will add a sentence to explain the importance of the reference to zero.

4. I was also puzzled by the discussion of Figure 3. Having read the reply provided to Referee #1 (which is at the moment rather confused), I would like to suggest the following. Apart from the notation used, the remark of Referee #1 would be correct if $Q_{max} > Q_{mean}$ (or if both these variables are normalized by the long-term Q_{mean}).

   To clarify Q_{max} is not normalized with the long-term mean of Q_{mean} but the long-term mean of Q_{max}. Therefore, while it is always valid that $Q_{max}>=Q_{mean}$ it is not necessarily valid that normalized Q_{max} is always larger than normalized Q_{mean} (while that might hold for most cases).

   Given that:

   $Q_{mean} \sim \varepsilon_{mean} * P_{mean} \rightarrow P_{mean} \sim (1/\varepsilon_{mean}) * Q_{mean}$

   $Q_{max} \sim \varepsilon_{max} * P_{mean} \rightarrow P_{mean} \sim (1/\varepsilon_{max}) * Q_{max}$

   This leads to:

   $(1/\varepsilon_{mean}) * Q_{mean} \sim (1/\varepsilon_{max}) * Q_{max}$

   And hence $\varepsilon_{max} > \varepsilon_{mean}$ directly follows from $Q_{max} > Q_{mean}$.

   I understand that the assumption that $Q_{max} > Q_{mean}$ is not valid, because those are values that have been normalized by their long-term respective means. In other terms, we are looking at Q_{mean} / Long-term Q_{mean} and Q_{max} / Long-term Q_{max}.

   That's correct.

   I suggest clarifying this (or making the term by which the normalization occurs explicit in Eq. 1), as lines 112-113, where the normalization is introduced, remain ambiguous (at least, I was not sure whether, e.g., Q_{max} was normalized by the long-term Q_{max} or instead by the long-term Q_{mean}).

In the revised manuscript we will clarify the use of normalized variables further to avoid confusion.

Minor comments

Some non-exhaustive minor comments are reported here. I hope they may help improve the manuscript, should you decide to revise it.

Streamflow elasticities/sensitivities to precipitation/temperature are called in many different ways throughout the manuscript (e.g., Line 48: streamflow elasticities of precipitation; Line 50: precipitation elasticities; Line 53: streamflow elasticities to precipitation; Line 262: annual elasticities of mean flow elasticity of maximum flow; Line 283, 370). I believe these are mostly typos, but they make reading the text difficult, because one wonders what the authors are actually referring to. Please choose one way to call them and use it consistently.

In the revised manuscript line 48 will be changed to "streamflow elasticities to precipitation" and line 262 will be changed to "relationship of annual elasticities of mean flow <u>and</u> <u>annual</u> elasticit<u>ies</u> of maximum flow (blue) and minimum flow (orange) to annual precipitation." The choice of referring to "precipitation elasticities" or "elasticities of mean flow" only (like in line 50 or 270) was done in paragraphs where we already referred to the long version "streamflow elasticities to precipitation" to make the text more readable.

Line 89: what are "suspicious day"? This is quite a subjective criterion to remove catchments.

This is a flag provided in the EStreams dataset. In the revised manuscript we will add more detail to this variable.

Lines 154-155: why do you exclude catchments with elasticities lower than -0.5?

Because there are only few values with elasticities lower than -0.5 and the more negative the elasticities are they become less feasible physically. In the revised manuscript we will add a sentence to motivate this choice.

Lines 155-156: do you mean elasticities with nan values? How were those values obtained?

Yes, here we mean elasticities with nan values. In the revised manuscript we will specify that we mean elasticities with nan values and add a sentence on how they are obtained.

Line 235: does it mean that, by calculating metrics at the annual scale, the authors are making the implicit hypothesis of water storage that does not last longer than a year? Please state this assumption explicitly.

We characterize the sensitivity to annual variations in climate without making explicit assumptions or hypothesis on storage. We acknowledge that longer term storage variations just like other factors can affect obtained elasticities as supported by Zhang et al. (2022), that is referenced in the manuscript.

Line 250: can you provide a reference that supports such hypothesis (i.e., that mean annual precipitation is correlated with maximum precipitation)?

In the revised manuscript we will add a reference to support the hypothesis that mean precipitation is correlated with maximum precipitation.

Lines 283-284: I was surprised not seeing a reference to Muller et al. (2021) in this discussion of how catchments may dampen or amplify precipitation variability, given that that study suggests mechanisms by which the amplification may occur. Muller et al., Catchment processes can amplify the effect of increasing rainfall variability, Environmental Research Letters, 2021. https://doi.org/10.1088/1748-9326/ac153e

This is indeed an interesting paper to connect to. We will reference to Muller et al. (2021) in the revised manuscript.

Line 301: what does the term "flow type" indicate here?

It refers to mean, maximum and minimum flow. In the revised manuscript we will type them out to avoid confusion.

Line 457: so, the conclusion is that climate appears to be the strongest control of the streamflow elasticity to climate. Recalling the comment above on the use of functional catchment descriptors, it would perhaps be more informative to strengthen the discussion of results in terms of the catchment water balance, and how this modulates the climate signal.

See reply to comment 1.

Given that the study investigates sensitivities of streamflow and discuss them in term of resilience, I was surprised it does not compare its results to the "Resilience of river flow regimes" paper, and instead only mention it for introducing the term resilience. Botter et al., Resilience of river flow regimes, PNAS, 2013. https://doi.org/10.1073/pnas.1311920110.

This was surprising especially because several results of this work seem to contradict the results of that study (if I am not mistaken). For example, (line 135) sensitivities decrease with longer timescales in that study (see Fig. 3C), and (lines 405-409) arid basins show lower sensitivity to precipitation forcing than more humid ones (see Fig. 3C,E), once discounted for the exposure (i.e., the difference variability of precipitation recorded indata for humid and arid basins). Although the investigated metrics are different (annual means vs probability distributions of the original variables), it would be interesting to comment on why such differences arise.

Although this paper also studies the resilience of river flow, they look at a different component of resilience, which makes a direct comparison of results very difficult. Botter et al. (2013) derive an index which is the ratio of the mean interarrival of flow producing precipitation events and the mean catchment response time to differentiate between erratic regimes and persistent regimes. When flow producing precipitation events are frequent and the mean interarrival time is shorter than the duration of the flow pulses leading to less variable and more predictable flows the regime is described as persistent. When the mean interarrival between flow producing precipitation events is longer than the typical duration of resulting flow pulses leading to a wider range of observed streamflow the regime is described as erratic. They describe erratic regimes as being more resilient due to their reduced responsiveness. The focus here is on the relative timing (between flow producing precipitation events compared to the mean duration of the flow pulse). In our manuscript we look at a fixed time window (annual and seasonal) to compare magnitudes of (normalized) variation of precipitation (and temperature) and streamflow.

The sensitivities displayed in Fig. 3C of Botter et al. (2013) refer to the ratio between the regime instability and the exposure (S=RI/E), where the regime instability is "defined as the relative fraction of probability shifting from one flow range to another in response to hydro-climate fluctuations" and the exposure index represents "the sum of relative variations of the shape and rate parameters of the flow distribution". This means that they do not define sensitivities as how annual streamflow varies per annual precipitation variation as we do in our manuscript.

References

Botter, G., Basso, S., Rodriguez-Iturbe, I., and Rinaldo, A.: Resilience of river flow regimes, Proc Natl Acad Sci U S A, 110, 12925–12930, https://doi.org/10.1073/pnas.1311920110, 2013.

Janssen, J. and Ameli, A. A.: A Hydrologic Functional Approach for Improving Large-Sample Hydrology Performance in Poorly Gauged Regions, Water Resour Res, 57, https://doi.org/10.1029/2021WR030263, 2021.

Tarasova, L., Gnann, S., Yang, S., Hartmann, A., and Wagener, T.: Catchment characterization: Current descriptors, knowledge gaps and future opportunities, https://doi.org/10.1016/j.earscirev.2024.104739, 1 May 2024.

Zhang, Y., Viglione, A., and Blöschl, G.: Temporal Scaling of Streamflow Elasticity to Precipitation: A Global Analysis, Water Resour Res, 58, https://doi.org/10.1029/2021WR030601, 2022.