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

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.
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.

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.

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.

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}).
Given that:

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

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

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}.
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}).

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.

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

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

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

Line 250: can you provide a reference that supports such hypothesis (i.e., that mean annual 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

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

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.

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 in

data 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.