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)

General comment

This study analyses the streamflow elasticities (considered as sensitivities) to climate described as observed percentage changes in river flow per percentage change of a climate driver. The autors have used a large hydrometeorological dataset to draw data from over 8,000 catchments, and provide a pan-European quantification of elasticities of annual mean and extreme streamflow to precipitation and temperature. In a second part, they intend to analyze the dependence of the elasticities to precipitation in using a random forest model with 20 climate and catchment factors to explain regional difference in elasticity value.

The objective is to demonstrate that such empirical strategy advances understanding of hydrological resilience of stream flows to climate change highlighting amplified and dampened streamflow response to climate which can ultimately support water management and disaster risk mitigation across Europe.

I believe that the manuscript has some potential, but additional work is required to put it at the standard required by the Journal of Hydrology and Earth System Sciences and it should be rejected in the current form. The datasets are not described in sufficient detail, the evaluation is limited, and the overall quality of the paper is affected by several shortcomings. I summarize my comments below in the hope that they will help the authors prepare an improved version of the manuscript.
Below we provide detailed point-to-point responses that addresses all raised concerns

Major comments:

1) Dataset and pre-processing.

- The authors use the comprehensive large-scale EStream dataset (do Nascimento et al., 2024), but they do not properly introduce it in the article, even though information about the sources of the hydrometeorological variables is crucial. For example, if EStream relies on ERA5 reanalysis for precipitation (as does the CARAVAN dataset cited by the authors), this would introduce limitations, since it is partly based on a model.
  In the revised manuscript, we will introduce the EStreams dataset with more detail, thereby also highlighting the source of precipitation and temperature data (which is E-OBS).
- The authors appear to have removed catchments with NaN values in either of the two seasons (L155) only for the seasonal dominance analysis. I recommend excluding these catchments from the entire analysis, as their annual elasticities may reflect only one season due to missing data.
  We agree such catchments should be excluded and they already are. In the annual analysis, we only compute annual values for years that have at least 11 valid months of data (i.e. >90% completeness). Subsequently, we only use those catchments that have at least 15 years with

valid annual values to calculate elasticities and sensitivities. Through this we avoid that the year is overrepresented by one season. We will state this more clearly in the methods of the revised manuscript.

- The methodology for minimum flow should be applied more cautiously, as the authors report several basins with instrumental errors (Q = 0 for multiple dates). Screening these basins would help improve the analysis. In addition, presenting the distributions of Qmean, Qmax, and Qmin in the supplementary material would better support the claim that Qmin is not biased by erroneous zero values.
  While the existence of zero flow values, even Q = 0 for multiple dates, could be physical, a prolonged period of over a year of zero values is not likely an actual measurement but indicating the presence of errors. In the pre-processing of the timeseries we will add a criterium that excludes the years that hold more than 90% zero values for that year, or whose 0 values follow substantial flow rates before (i.e. the flow rate before Q = 0 exceeds Q_5) as such an abrupt shift would likely reflect a measurement error. In addition, the spatial and frequency distributions of Q_{mean}, Q_{max} and Q_{min} will be presented in a new supplementary figure.

2) Flaws in the method.

- The author introduce the estimation of elasticities with too few reference from the literature and without supporting their choice in the chosen formula compare to others (Andréassian et al. 2016).
  In the revised manuscript, several more studies will be added to support the choice of the formula, and to provide broader context of the use of elasticities in the hydrological literature. In our manuscript, we deliberately picked two different methods to derive elasticities to test if the main results are method-dependent, which they were not (see figure S2 and L33-38 in the supplement).
- Equation 1 defines the elasticities. From my understanding, the elasticity to P (or T) for the extreme flows Qmin and Qmax is also related to the annual mean of P. If this is the case, the interpretation of Fig. 3 (L266–277) remains unclear to me. Since P_mean≈Q_mean x ε_mean≈ Qmax x ε_max, it follows directly from Q_mean<Q_max that ε_mean<ε_max.
  This derivation with its associated interpretation does not apply. Precipitation (P) and streamflow (Q) variables in Eq. 1 are normalized by their multi-year means. Simplifying the formula of how Q and P relate would not be $P_{mean} \approx Q_{mean} x\ \varepsilon_{mean} \approx Q_{max}\ x\ \varepsilon_{max}$ but $P_{mean} \approx Q_{mean}/\varepsilon_{mean} \approx Q_{max}/\varepsilon_{max}$. Since streamflow and precipitation are normalized by their annual means, which implies that their normalized values are all, on average, one. This applies to studying Q_{mean} or Q_{max} (or Q_{min}). Therefore, the stated concern is incorrect, as the general assumption that $Q_{mean}<Q_{max}$ does not apply.
- It is unclear why and how the authors have proceed with temperature "using the absolute annual mean of temperature". Does it means absolute temperature is used in Equation (1) ? Then the absolute operator should be used (also T is not introduced in this Eq.).
  We apologize for any confusion. In the revised manuscript, we will introduce T in equation 1 and clarify that we are referring to the annual mean temperature (without normalizing). The motivation for using temperature in this form is motivated in L113-115.
- Without normalization of T, the regression uses one normalized variable (P) and one non-normalized variable (T), so the elasticities with respect to P and T are not directly comparable. This represents a limitation of the analysis: the authors rely on a bivariate approach to estimate elasticities, but do not (and cannot) jointly evaluate these parameters.
  Indeed we use one normalized variable (P) and one non-normalized variable (T) and therefore, the elasticity and sensitivity values are not directly comparable (and have a different name),

as also mentioned in the manuscript. This choice is motivated in L113-115 and L99-109. This approach has also been applied previously (see Vano et al., 2015 cited in the manuscript). The bi-variate approach still holds with the limitation of not being able to compare them relative to each other. In the revised manuscript, we will highlight this more explicitly.

- The authors chose to first estimate elasticities using a linear model (Eq. 1), and then to relate these elasticities to catchment attributes with a nonlinear model (random forest) in order to derive feature importance (Fig. 7). However, while their nonlinear model does not accurately reproduce the elasticities ($R2 < 0.52$), they still rely on it for interpreting feature importance. It rather uncommon to trust feature importance derived from a model that fails to explain more than half of the spatial variance. Furthermore, they retained climate-related attributes such as aridity in the model; aridity emerges as the most important predictor, yet it is directly linked to P, from which the elasticity is defined. This introduces a circularity in the methodology that weakens the support for their conclusions. I would recommend not relying on inaccurate nonlinear modeling and instead keeping the analysis at the linear level, avoiding the inclusion of nested variables in the modeling framework.
  In the revised manuscript, we will further highlight that the random forest model weakly predicts elasticities and adapt the description of how links to variables should be interpreted accordingly.
  Changing to a linear modelling approach would mean that we assume elasticity to be varying linearly with all characteristics, which would be unrealistic to expect for particular characteristics (e.g. catchment size). For the revised manuscript, we will conduct a parallel analysis using a hierarchical multiple linear regression approach for comparison in the supplementary material.
  Although aridity is related to precipitation, it is a variable that describes the relationship of multi-year mean precipitation divided by multi-year mean potential evapotranspiration. This information is not information used to derive elasticity (which only depends on annual anomalies of these variables). With the elasticities we study the response of streamflow to precipitation by considering annual (or seasonal) deviation from the mean behaviour. Further, several papers address the relationship of aridity and streamflow elasticity to precipitation (see Zheng et al., 2009, Sankarasubramanian et al., 2001 and Potter et al., 2011 cited in the manuscript).

3) Limitation in the Analysis.

- Several figures are not discussed, including all PDFs in Figures 2, 4, and 6, and it is not specified whether the vertical bars represent means or medians.
  In the revised manuscript, we will add the explanation of the vertical lines in the kernel density plots. The figure 2a is discussed in L189-199. For figures 4 and 6 we will make the discussion of the kernel density plots more explicit.
- Figure 2 is interpreted mainly through visual inspection of spatial patterns in relation to other catchment attributes that are not shown in the main text but only in the supplementary material. Subsurface properties appear to play an important role, yet no figure directly supports this. I suggest providing quantitative metrics and additional figures to substantiate these claims.
  In the revised manuscript we will highlight better that the quantification of the relationship of the elasticities to different catchment characteristics is shown in 3.4.
- The strategy for selecting catchment attributes included in the random forest model is not described, and the metrics are not specified (for example, what is the VIF mentioned in L176?).
  L175-176 describe how the selection of attributes was motivated. In the revised manuscript,

we will further clarify that and the VIF will be defined to make the selection criteria more understandable.

- The concept of feature importance in the random forest, and what exactly it represents in this context, is not explained in the article.
  In the revised manuscript, we will describe in more detail how to interpret the feature importance.

4) Unclear presentation

- The elasticities to precipitation use various and different notations through the paper (in the abstract, Eq.1, Fig.2 and in the text) that must be harmonized.
  In the revised manuscript, notations of elasticity will be harmonized throughout the paper.
- Seasonal dominance is introduced but the notation is not used in figures or in the maintext.
  In the revised manuscript, we will refer to the introduced seasonal dominance notation in the text and figures.
- Significancy of the elasticity cannot be read from the Figs. 2, 4,6. And are not discussed either.
  In the revised manuscript, we will include information on the significance in the written text.

Specific comments:

- l70 : This “non-parametric approach”. The term here is irrelevant as linear regression is a parametric approach.
In the revised manuscript, the term “nonparametric” will be removed.

- Importance of Fig. 1 is not obvious and its deleting could save place.
This figure aims at highlighting how we combine the use of normalised variables (P and Q) and not-normalised variables (T), and how the bi-variate approach works despite using not-normalised temperature. We can move this figure to the supplement.

- The notion of “seasonal scale” arises lately in the paper (L135), such finer grain of analysis should be better emphasised in the Abstract and Introduction of the paper.
At present, it is mentioned in the abstract in L13 and in the introduction in L49. In the revised manuscript, we will consider if we can further clarify this early on in this manuscript.

- The notion of p-values in L145 is not clear, how it is computed?
In the revised manuscript, we will specify more in detail how p-values are calculated.

- Table 1 : human influence: Mean area equipped for irrigation: what 10/5-year resolution means?
This notation indicates a variable sampling resolution of 5 or 10 years. As we use averaged values over the entire time period, this information will be removed from the revised manuscript.

- Fig. 3 : the “binning in group of 2%” is not clear, is it the range of variation of elasticity inside each bin ?
It means binning points along the x-axis in groups of 2%, e.g. the lowest 2% of mean flow elasticity values will be binned into the first bar, the next 2% in the following, etc. This will be clarified further in the revised manuscript.

-Fig4: Pdf are note introduced in caption.
In the revised manuscript, kernel density plots will be introduced in the caption.