Comments/Text of Keirnan Fowler (Reviewer 1) posted in black, our text in blue with additions to existing text in red.

This paper examines the common concept of streamflow elasticity and takes it one step further, examining the sensitivity of different flow percentiles to changes in precipitation. This means the impact of climate changes can be examined separately for high flows and low flows. I find this to be a worthwhile extension to an existing widely-used method. It allows the elasticity concept to be more closely related to problems of societal interest such as ecological sensitivity to changes in low flows and impacts on infrastructure due to changes in high flows.

I find the manuscript to be close to publication standard already. The methods used are rigourous, the writing is usually quite clear, the findings are well supported by quality figures, and the paper is relatively complete. I offer the following comments, in the hope of improving the paper from its already high standard. Points 1 and 7 are editorial (and thus subjective) and are suggestions only.

We first thank the reviewer for their comments and suggestions which are thorough and contribute positively to the development of the manuscript. We propose making the following 3 substantial revisions to the paper.

1. Primarily, we will move the majority of the discussion and results of the panel regression models to the appendix.
2. Second, we will re-work the abstract so that it more accurately reflects the paper's objectives and conclusions and will further clarify the aims and justification in the introduction.
3. Finally, we will adjust the phrasing around the methodology, particularly equation 2, and add additional examples in order to limit confusion about the approach.

Comments are addressed in more detail below.

1. OVERALL FRAMING OF PAPER. I think the abstract and introduction could be improved to frame the paper better and increase its impact. To me, the paper should primarily aim to be an introduction to a new concept (or, more precisely, a new variant on an existing concept), as per the existing line 400: "The intention of this paper is to provide an introduction to the concept [of elasticity curves] in a large-sample context". If this is indeed the goal, then the authors ought to aim to clearly establish: (1) the importance of the existing method; and (2) the need for the new method, couched in terms of the limitations of the existing method. Neither of these aims are achieved very well in the existing abstract and introduction.

Specifically, the tone of the introduction seems to take for granted that the existing method is important; it does not clearly explain its significance or what questions can be/have been answered by the method in the past. Likewise, although the introduction does go some way to answering (2) (line 62-63), it waits too long to do so and does not go into sufficient detail (saying only "abnormally high and low flows are associated with the greatest strain on hydrological systems"). Can we get a lot more detail here? Eg. for high flows, it could acknowledge/discuss that infrastructure is often designed according to estimates of flooding potential, so any changes to this potential are very important; likewise low flows are important eg. for riverine ecology among other things. Articulating these factors will help the reader understand why the new method is important, which will motivate them to keep reading. In my view, it is crucial to do this early, before they lose interest.

As for the abstract, the majority of the text is spent trying to articulate the different catchment "types" that have been defined for the example application. This would be fine if the paper was about a new system for classifying catchments. But if the paper is about introducing the concept of elasticity curves, then this detail is unexpected and unhelpful in the abstract. The abstract needs to be about the method, not this particular application. Readers can read the full paper if they want this sort of detail. My suggestion would be to focus the abstract on the importance/significance of the method and what it adds; limit the results to a handful (say two or three) things that were learned in the specific application. Note, existing text in the conclusion section, Line 427 - 434, contains some of the above elements and could be adapted for the abstract.

The objective of this paper is to both establish the concept of an elasticity curve, and to demonstrate how, in combination, the multiple elasticity estimates are informative separately from point estimates. We agree with the reviewer's assertion that this could be clearer in the introduction and abstract and that the abstract is overly focussed on the classification scheme. We propose adding additional clarifying sentences throughout these sections which clarify the aim and the relevance of the method. In addition, we propose re-writing the abstract so that it reads as follows:

"Streamflow elasticity is a simple approximation of how responsive a river is to precipitation. It is a ratio of the expected percentage change in streamflow for a 1% change in precipitation. Typically estimated for the annual average streamflow, we here propose a new concept in which streamflow elasticity is estimated for multiple percentiles across the full range of the streamflow distribution in a large-sample context. This "elasticity curve" can then be used to develop a more complete depiction of how streamflow responds to climate. Representing elasticity as a curve which reflects the range of responses across the distribution of streamflow within a given time period, instead of as a single point estimate, provides a novel lens through which we can interpret hydrological behavior. As an example application, we calculate elasticity curves for 805 catchments in the United States and then cluster them according to their shape. This results in three distinct elasticity curve types which characterize the streamflow-precipitation relationship at the annual and seasonal timescales. Through this, we demonstrate that elasticity estimated from the central summary of streamflow, e.g. the annual median, does not provide a complete picture of streamflow sensitivity. Further, we demonstrate that elasticity curve shape, i.e. the response of different flow percentiles relative to one another, can be interpreted separately from between-catchment variation in the magnitude of streamflow change associated with a one percent change in precipitation. Finally, we find that available water storage is likely the key control which determines curve shape."

2. CLARITY OF METHOD. I feel there is a strong possibility of readers misunderstanding the method. Specifically, the focus on different flow percentiles (or ranges of percentiles) may lead readers to believe that the method only focusses on precipitation that falls during the relevant percentile/range. For example, the reader might believe that the method is asking "how sensitive is low flow to precipitation that falls concurrently with times of low flow?" whereas my understanding is that the intent is to use the same seasonal or annual average of precip & PET regardless of which flow percentile is in view. Is this correct? Can the authors make this clearer please? Perhaps via some more concrete examples? An explanatory figure may also help.

The reviewer has correctly understood what we have done, and we agree that this could be made clearer throughout the document with a series of examples. In addition to some changes in phrasing, we propose adding the following clarifying statement: "As presented in this study, the elasticity curve characterizes the sensitivity of different percentiles of annual and seasonal streamflow to changes in the average annual or seasonal precipitation. For example, an elasticity of 0.5 for the 15th percentile of annual streamflow would indicate that a 1% change in the overall mean annual precipitation would correspond to a 0.5% change in the 15th percentile of annual flow."

Additionally, we will change the text which describes the model (currently lines 127-130) to read:

"where $Q^{q}_{i,t}$ is the natural logarithm of a streamflow percentile (q) calculated for time period (t) for catchment (i), $\alpha_{i,t}$ is the intercept, $\ln(P_{i,t})$ is the logarithm of catchment averaged mean daily precipitation for the time period of interest (year or season), and $\ln(E_{i,t})$ is the logarithm of catchment averaged mean daily potential evaporation in that period. Note that mean seasonal and annual climate time series are used, not percentiles equivalent to the streamflow percentile of interest (denoted with the superscript "q"). The point estimate of precipitation elasticity is represented by the regression coefficient: ε^{q}_{P} and potential evaporation elasticity is represented by ε^{q}_{E} The error term is $\eta^{q}_{i,t}$."

3. GREATER JUSTIFICATION OF "CAUSAL" NATURE OF ANALYSIS. The panel regression model is described as a "causal" model (eg. line 152). Can the authors please provide more justification for this? I am not an expert in this area, so I am looking for more information here - it seems to me as if this method is a variant on linear regression, with additional care to hold confounding factors constant. However, even if the authors manage to hold every available confounding factor constant, it does not resolve the problem that correlation does not imply causation. Have other authors made similar claims of panel regression, and what is their reasoning? Given there exist specialised causation methods (ie. methods that were directly formulated to try to distinguish correlation and causation such as https://doi.org/10.1126/science.1227079), it is not a claim that I would be making lightly. Even if the authors agree with me, this is no reason to change to experimental design; merely the way it is described.

Panel models have been widely used for causal inference, especially in combination with graphical models, like a directed acyclic graph (D.A.G), to make the modelling assumptions clear. We used in the creation of the models but did not include them with the manuscript. Examples of such studies include Anderson et al., 2022; Blum et al., 2020; and Yang et al., 2021 in hydrology and Ferraro et al., 2019 in environmental studies, as well as a wide range of other research in fields including econometrics and health sciences. In fact, even the simple linear regression model can be used for "causal inference" if the assumptions are explicit enough (Pearl, 2009) although this is not a particularly robust tool. Causal inference implies a different way of approaching a problem, where the intent is explicitly to infer relationships which persist across changing conditions and which cannot be clearly defined by distribution functions alone, relative to more typical statistical inference approaches which focus on assessing the parameters of a distribution and establishing statistical significance. There is an enormous amount of overlap in methodologies, but the underlying principles differ.

Two-way linear fixed effects panel regression models are robust to many aspects which would normally bias regression approaches and this makes them a useful tool for this purpose. For instance, many catchment-level confounding variables (which are time-invariant) are controlled for by the streamgage-specific intercepts as in (Blum et al., 2020). Additionally, they address the majority of omitted variable bias by requiring that confounding variables either be directly measured or be invariant along at least one dimension of the data, for instance, time (Anderson et al., 2022; Nichols, 2007). Simple linear regression models, or any single-site regression models for that matter, cannot address this. We direct the reviewer to Section 4.3. in Anderson et al., 2022, for a more detailed description.

That said, we also recognize that "causal inference" is relatively new in hydrology and that the use of causal terminology in this way may be misunderstood by unfamiliar readers. This is especially true in the context of the current manuscript, where we have used a very simple model design, have not presented the D.A.G as part of the manuscript, and where the inclusion of causal language is not essential to the argument. Further, we agree that while panel regression models are a robust statistical tool which offer a range of benefits for hydrological science, their validity for this purpose relies on having well-constructed, explicit, modelling assumptions (Imai & Kim, 2021).

Thus, while we feel confident that our modelling assumptions are valid and are comfortable making causal inferences, we propose limiting causal terminology in the paper and will instead rely predominantly on language appropriate for statistical inference approaches (e.g. associations, etc.). We will additionally move the majority of the discussion of the panel models to the appendix, as is suggested by the reviewer in a comment below. We will retain mention of the panel regression model in the text to demonstrate the robustness of the approach presented in the paper and will briefly elaborate on its causal applications in the text when directing the reader to the appendix. Because of the panel regression model's robustness to omitted variable bias, we feel that it is important that it not be entirely excluded from the paper.

4. JUSTIFICATION OF TWO METHODS. The results of the two methods are very close (a key difference is the uncertainty bounds, but these may be closer than they look - see following point). The results are so close that one wonders whether the two methods are actually doing almost the same thing. Two recommendations arise from this:

- Are both methods really needed? They are taking up valuable real-estate and they really detract from the story because it becomes more about comparing the methods rather than reflecting on what the results actually mean in this first-of-its-kind study. Perhaps one of the methods could be moved to an appendix?

- If the authors elect to keep both methods, I suggest they bolster their justification for why the two methods are different.

The panel regression models lend credibility to the concept because they are substantially more robust than blending together individual single site models. Therefore, we will not exclude the approach from the paper. Instead, we will move the majority of discussion of the panel regression model design, results, and associated figures to the appendix. In the text, we will state that panel regression models with a similar parameterization as the single site models were applied and that the results were remarkably similar; in this way we can reduce the space taken up by model inter-comparison and increase discussion of the results.

5. REPORTING OF UNCERTAINTY. If the authors elect to retain both methods, then the following becomes relevant. With respect to Figure 2: At first glance, the uncertainty bounds appear to be the

key difference between the two methods. However, the comparison is apples with oranges, and if the authors correct this then the uncertainty bounds may be much more similar. Specifically, for the panel model, the confidence intervals are plotted, whereas for the single catchment models, IQR is plotted, which is a lot closer to prediction intervals (which plot uncertainty about individual predictions) than it is to confidence intervals (which plot uncertainty about the underlying assumed relationship). Is it possible to retain the IQR for the single catchment models and then swap to prediction intervals for the panel regression?

We thank the reviewer for this useful comment. We will calculate prediction intervals for the panel regression models, which can be included in the appendix with those figures.

6. LIMITATIONS. Regarding line 320-322, this speaks to a significant limitation here: the method assumes a flow response in the same time period. So it might be that certain flows are very sensitive to changes in climate, but if this sensitivity is subject to a delay longer than the time period used, this will not be detected here. Is this true? If so, I suggest this be included / discussed.

This is true, but we do not feel that it is detrimental to the work. This is a typical assumption of hydrologic elasticity work, although it is rarely discussed as explicitly as we have done, and while this is particularly relevant at the seasonal scale, at the annual time scale, this seasonal element is less influential.

Basically, an elasticity estimate is always just going to provide an approximation of how responsive streamflow is to precipitation change in the time period you are looking at, which is why storage and evaporation are such important controls on elasticity. For example, if storage is seasonal in nature (e.g. snow), seasonal streamflow elasticity (e.g. in winter) might be close to 0, but the elasticity of summer streamflow to winter precipitation might be quite high, because local hydrology relies on winter snow melt to drive dry season flows. The method doesn't assume a flow response in the same time period, instead, the method sheds light on the extent to which a response occurs in that time period without ruling out the possibility of a delayed response. Specific investigation into delayed responses would be an interesting direction for future work, but is not the point of this paper. We propose adding the following clarifying comment at the end of the paragraph referenced, and may add further clarification of the same type throughout the discussion:

"The seasonal elasticity estimates specifically consider the influence of in-season precipitation on streamflow within that same season. Streamflow in many rivers is driven by out-of-season precipitation, for example, snow which falls in winter and fall may drive spring and summer streamflow as it melts, particularly in high altitude regions. Thus, while flat seasonal elasticity curves and low percentile-specific point estimates indicate a muted hydrologic response, they do not rule out the possibility that the timescale for response is merely longer than that which is considered."

This sentence, on lines 342-344 already tries to convey this point, "The range of type B elasticity curves which is present across the seasons is washed out at the annual scale, demonstrating that the catchment storage which leads to a uniform response across the distribution of streamflow generally operates at a timescale of less than year," but seems to fail at this objective, so we will try to further clarify our intended meaning here.

7. EDITING FOR EASIER COMPREHENSION: Overall the text of the paper is admirably concise, but it's still very dense. I suggest the authors review the existing text specifically to try to make it easier to

comprehend. One technique that might help is the use of tables since these facilitate a visual structure to the information. This is a suggestion only - I understand that this can be time-consuming to do!

We are happy to edit the text to try to simplify the flow of the paper to improve clarity. However, it is not immediately apparent to us how we could incorporate tables in a manner which contributes towards this goal.

ADDITIONAL MINOR COMMENTS, BY LINE:

Figure 1b: I think there's potential for confusion here: the line is monotonic increasing, but the unfamiliar reader might wonder whether it *must* be so, or whether it could be different. For example, the median (purple) point here could have higher elasticty than both the high and low flow, yes? If so, consider changing the figure to a non-monotic relationship to make it clear this is a possibility.

This is a good point and is accurate. We propose changing the figure to include a non-monotonically trending line in order to better reflect this.

Line 111-12: "We estimated ... catchment boundary". Perhaps rephrase for better clarity.

We will replace this sentence with: “We estimated average daily precipitation (mm/-

day) annually and for each season, averaged within the upstream drainage area (watershed boundary) of each gaging station.“

Line 116: Unclear. Does "we recalculated these values in order to accurately represent the time period of the analysis" mean "we recalculated these because the existing dataset did not cover our desired period"?

This is correct, we will change the wording to reflect this.

Line 118: Would the sentence "Annual values..." fit better after the sentence that follows it? Also "fall into corresponding “years”" - is "water years" the intention here?

We will change this to say: “Annual values were calculated for water years (defined here as September to August), and seasonal values were estimated for winter (December, January, February), spring (March, April, May), summer (June, July, August) and fall (September, October, November) within each water year.”

Figure 3: Unclear that column c is the same as Fig 2, partly because it is named differently. I suggest to name it exactly the same previously ("clusters") and potential add the words "(from Figure 2)" to make this explicit. I realise the current title is aiming to clarify normalised versus non-normalised but I feel this clarification can occur in the Figure 3 caption.

This is a good suggestion and we will implement this change.

Line 365: The introduction of example catchments interrupts the flow of the paper. Could this be done in the methods section instead?

Yes, we will introduce the example catchments in the methodology section instead and work to improve the flow of the paper.

References

Anderson, B. J., Slater, L. J., Dadson, S. J., Blum, A. G., & Prosdocimi, I. (2022). Statistical Attribution of the Influence of Urban and Tree Cover Change on Streamflow: A Comparison of Large Sample Statistical Approaches. *Water Resources Research*, *58*(5), e2021WR030742. https://doi.org/10.1029/2021WR030742

Blum, A. G., Ferraro, P. J., Archfield, S. A., & Ryberg, K. R. (2020). Causal Effect of Impervious Cover on Annual Flood Magnitude for the United States. *Geophysical Research Letters*, *47*(5), e2019GL086480. https://doi.org/10.1029/2019GL086480

Ferraro, P. J., Sanchirico, J. N., & Smith, M. D. (2019). Causal inference in coupled human and natural systems. *Proceedings of the National Academy of Sciences*, *116*(12), 5311–5318. https://doi.org/10.1073/pnas.1805563115

Imai, K., & Kim, I. S. (2021). On the Use of Two-Way Fixed Effects Regression Models for Causal Inference with Panel Data. *Political Analysis*, *29*(3), 405–415. https://doi.org/10.1017/pan.2020.33

Nichols, A. (2007). Causal Inference with Observational Data. *The Stata Journal*, *7*(4), 507–541. https://doi.org/10.1177/1536867X0800700403

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys*, *3*(0), 96–146. https://doi.org/10.1214/09-SS057

Yang, W., Yang, H., Yang, D., & Hou, A. (2021). Causal effects of dams and land cover changes on flood changes in mainland China. *Hydrology and Earth System Sciences*, *25*(5), 2705–2720. https://doi.org/10.5194/hess-25-2705-2021