Reviewer #1 Comment 1: (hereafter referred to as R1C1, R1C2…) *In the paper, the authors estimate the effects of urbanization, cropland area, and dam regulation on the magnitude of annual maximum streamflow by analyzing historical data from 2739 gaged catchments in China. The authors use panel regression methods to identify these effects, and find that urbanization increases annual maximum streamflow, whereas dam regulation decreases annual maximum streamflow. Overall, the paper addresses an important question and provides new understanding about the factors leading to changes in annual floods. The paper is well-organized and clear, with ample references to previous studies, and the methods are appropriate for the questions studied. There are a few issues that could be addressed with minor revisions, which I have detailed below. First, more detail about the underlying data for the regression is necessary to fully interpret the results. Second, I would also advise against comparing the magnitude of p-values as a method to select between two different models. I have also noted some areas of the text needing clarification and included one suggestion for additional analysis that I think could be of interest to readers.*

A: Thank you for your constructive comments. We have carefully considered your suggestions and revised the paper. We have made the following major revisions in the method and data.

First, we added 3-day and 30-day total precipitation before flood peaks for each catchment in the regression to account for individual time-varying confounders. The reason for such a revision is that the delineated climate regions cannot fully control climatic confounders since the climatic drivers of floods have sub-regional spatial variability. Therefore, the regression equation has been revised as:

$$log(Q_{i,t}) = \alpha_i + g_1\left(Urban_{i,t}\right) + g_2\left(Crop_{i,t}\right) + g_3\left(RI_{i,t}\right) + \pi_{r,t}D_rD_t + D_r\left(\varphi_r P_{i,t}^{(3)} + \lambda_r P_{i,t}^{(30)}\right) + \varepsilon_{i,t}$$

where $P_{i,t}^{(3)}$ is the 3-day total precipitation before the flood peak in year t of catchment i, which accounts for the rainfall that causes the flood; $P_{i,t}^{(30)}$ is the 30-day total precipitation before the flood peak in year t of catchment i, which accounts for the soil moisture and snowmelt that cause the flood. The coefficients of $P_{i,t}^{(3)}$ and $P_{i,t}^{(30)}$, namely φ_r and λ_r, are assumed to be constant within a climatic region r. The original region term $\pi_{r,t}D_rD_t$ accounts for omitted time-varying regional confounders other than $P_{i,t}^{(3)}$ and $P_{i,t}^{(30)}$.

Second, we selected 757 non-nested catchments to fit the regression model so that the residuals of the model were not highly correlated. This revision avoids uncorrected inference about the regression coefficients due to the underestimation of their standard deviations using correlated flood samples.

The results did not substantially change after the methodology revision above. Other issues you mentioned are addressed in the following replies.

Main Comments

R1C2: *1. It would be helpful if the authors included more detail about the underlying distributions (or ranges) of the causal factors studies. For example, the authors state that the effect of urbanization is stable (i.e. linear), but it is unclear over what range of urbanization values this was tested, and if the effect could be increasing/decreasing outside of this range. Also, the panel regression specifically models within-unit (catchment) effects. This requires that there are within-catchment variations in the explanatory variables, but it is unclear how some of the variables, like dam regulation, are distributed in space and time. What percentage of catchments have changes in reservoir index, urbanization, and/or cropland area over the time period? Suggestions for ways to add this information include reporting these statistics within the text and/or adding a figure(s) of single or joint probability distributions or a time series of the causal factors.*

A: Thank you very much for your critical comment. We agree with you that a statistical summary of catchment characteristics is necessary. We will add the summary table (Table R1) in the revised manuscript.

Table R1. Summary of catchment characteristics for 757 catchments in 1992-2017. The summaries of *RI* and Δ*RI* are calculated based on 207 catchments with at least one large and medium dam.

Variables	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Area **(km^2)**	29	499	1096	3341	2763	142372
Urban **(%)**	0	0.06	0.30	1.52	1.10	65.07
Δ*Urban* **(%)**	0	0.05	0.23	1.14	0.85	24.66
Crop **(%)**	0	10.63	24.71	32.75	48.99	99.58
Δ*Crop* **(%)**	-21.58	-0.81	-0.02	0.38	0.87	32.04

RI	0.01	0.09	0.21	0.51	0.61	7.45
Δ*RI*	0	0	0	0.17	0.07	7.44

R1C3: *2. Fig. 2 indicates that the lowest p-value is used to choose between a quadratic term ("increasing effect") or square root term ("decreasing effect") in cases where both terms have a p-value less than 0.01. P-values are not intended to be used to determine which model form is correct. It would be more appropriate to use a criterion designed for model selection, such as the AIC or BIC. Alternatively, in cases where both an increasing and decreasing effect are plausible and there is not a model form that clearly fits better, the authors could report both models.*

A: Thank you for your important comment. We agree with you that p-values are not appropriate for choosing models. In the revision, we used AIC to select effect terms (i.e, function $g_1(\cdot)$, $g_2(\cdot)$, and $g_3(\cdot)$) and found no change in the terms compared with the original manuscript. We will use AIC in the revised manuscript.

R1C4: *3. In Fig. 8, the authors provide a map showing stations that had relatively small changes in urbanization or reservoir index but did have significant changes (>10%) in annual maximum streamflow. It would be interesting to extend this type of comparison for the stations shown in Fig. 6 and 7. In other words, it would be interesting to compare the observed streamflow trends to the trends predicted by changes in urbanization and RI shown in Fig 6 and 7. This comparison could identify regions where additional causal factors are involved (and thus could identify interesting areas for future research). I do not believe this analysis is necessary for publication, but would likely be of interest to readers. Comparing observed trends with trends predicted by causal factors would also be a relatively new addition to the panel regression literature within hydrology*

A: Thank you very much for your valuable comment. You provided an interesting idea to compare observed flood trends with the trends predicted by the changes in *Urban* and *RI*. However, in addition to gradual changes (trends), observed floods may also experience abrupt changes. It is difficult to determine the percentage change of floods between two specific years using observed flood data with abrupt changes, especially using data in a short period (26 years from 1992 to 2017 in this study). It is beyond the scope of this study to derive the percentage change of floods under different kinds of flood change patterns. Nonetheless, we still hope to see whether catchments with significant changes in observed floods match the ones with large changes in *Urban* and

RI. Therefore, in the revised manuscript, we will label the catchments with significant changes in observed floods in Fig. 6 and Fig. 7.

Line-by-line comments:

R1C5: *Line 31: "whether a factor affects floods?" would read more clearly as "does a factor affect floods?"*

A: Thank you very much for your suggestion. We will change the sentence accordingly in the revised manuscript.

R1C6: *Lines 47-64: In this paragraph, I find the descriptions of the existing methodological approaches to be unclear, particularly for the first method. Is the first "model-based" method referring to an empirical model, a physically-based model, or some combination?*

A: Thank you very much for your comment. The first method refers to the use of physical hydrological models, which can simulate floods in different scenarios (e.g., with and without human impacts). We will change "model-based approach" to "physical model simulation" in the revised manuscript.

R1C7: *Line 116: "the time-varying constant effects" should just be "the constant effects"*

A: Thank you very much. We will change the text accordingly in the revised manuscript.

R1C8: *Lines 119-121: "Although RI may correlate with Urban and Crop, the effects of dams and land cover on floods can be derived independently since we have controlled their common drivers (Pearl and Mackenzie, 2020) in each equation, i.e., the regional time-varying term and the individual time-invariant term". I'm not sure this is correct - it is possible that RI could be temporally correlated (within-watershed) with Urban or Crop, in which case the effects could be confounded if they are modeled separately. The results don't seem to indicate that the variables are confounded, but this could also be checked by calculating the within-unit correlations.*

A: Thank you very much for your comment. *RI* and *Urban* or *Crop* have step changes, so with-unit correlations cannot capture their dependence relationship. In the revision, we combined *Urban*, *Crop*, and *RI* into one regression model. Note that the effects of *Urban*, *Crop*, and *RI* on floods have little change after revising the regression model.

R1C9: *Line 138: "Since the pooling samples were sufficient for statistical inference,…". It is unclear what is meant by this statement, so it should be clarified.*

A: We no longer used p values to choose model forms. Therefore, this sentence will be deleted in the revised manuscript.

R1C10: *Lines 145-147: "1. No other important: …" would read more clearly as "1. There are no other important…". Likewise, "2. No interactions between…" would read more clearly as "2. There are no interactions between…"*

A: Thank you very much. We will change the sentences accordingly in the revised manuscript.

R1C11: *Line 155: different exponential formatting is used between Equations 6 and 7 (exp vs e). Equation 7 is also basically the same as Equation 6.*

A: Thank you very much. We will unify the exponential formatting in the revised manuscript. Equations 6 shows the sensitivity of Q to X, i.e., "the percentage change in Q given a fixed change in X", as described in Line 132. The fixed change, ΔX, equals 1% for X=*Urban* or *Crop* and 1 unit for X=*RI*. While Equation 7 shows the accumulated changes in Q due to the changes of X in a long period. We will clarify the differences between Equation 6 and Equation 7 in the revised manuscript.

R1C12: *Line 158: "|exp(g(Xi,t2) – 0) – 1| < 10%". I believe the "-0" should be replaced with g(Xi,t1)?*

A: Thank you very much for your comment. It should be "0". Here we hope to select catchments that are free from the impacts of urbanization and dam constructions. Therefore, the initial value of the factor (X_{i,t_1}) equals 0, which represents the condition with no urban areas and no dams. We will clarify our purpose for such a selection in the revised manuscript.

R1C13: *Lines 158-160: Are the trends in annual maximum streamflow calculated using log(Q)? This is what I expect based on the presentation of the results, but should be clarified in the text.*

A: Thank you very much. The trends in annual maximum streamflow are calculated using Q. We will clarify it in the revised manuscript.

R1C14: *Line 191 (and elsewhere): Suggest changing "large and middle dams" to "large and medium dams" or "medium and large dams".*

A: Thank you very much for your suggestion. We will change all "large and middle dams" to "large and medium dams" in the revised manuscript.

R1C15: *Line 199: "The number of catchment groups k in Section 2.2 had no optimal value." 'Was there a method used to test for an optimal value?*

A: Thank you very much for your comment. Indeed, we can use the Silhouette value of the clustering or the AIC value of the regression model to select a *k* value. However, these methods derive an optical *k* from a statistical perspective rather than from a hydrological perspective. We prefer to test the sensitivity of the model to varying *k* so that we can test the robustness of our method.

R1C16: *Line 209: "with only one exceptional type of effect for Urban and Crop". This phrasing is confusing, so I suggest rephrasing to something like: "except for two cases (Urban effect when k=10, and Crop effect when k=40)".*

A: Thank you very much for your suggestion. We will change the sentence accordingly in the revised manuscript.

R1C17: *Line 210: "percentage increase" – I believe this should be a percentage point increase? (For example, an increase from 10% to 11% is a 10 percent increase, or a 1 percentage point increase.)*

A: Thank you very much for pointing out the typo. Yes, the "percentage increase" should be "percentage point increase". We will correct the word in the revised manuscript.

R1C18: *Line 238 and 348: "more than 10% of increases in Q". If I understand correctly, this should be written as "increases in Q of more than 10%".*

A: Thank you very much. In the revised manuscript, Line 238 will be revised as "Increasing *Urban* causes increases in *Q* of more than 10% in 284 (10.4% of 2739) catchments", and Line 348 will be revised as "increasing urban areas cause increases in Q of more than 10% in 10.4% of the 2739 catchments".