Reviewer #2 (Comments to Author (shown to authors):

The authors analyse the relative importance of soil moisture and precipitation for the generation of the average annual flood in the Yangtze river basin. This is achieved by comparing the ratios of precipitation and soil moisture before the flood event with recorded maximum of the respective variable. The relative ratio of both variables shows a positive correlation with topographic wetness index and a negative correlation with magnitude.

Overall, I think the authors present a very thoughtful analysis which addresses some important drawbacks in our current approach to flood generating processes. Instead of annually averaged results I would have liked some more event-based results. Other recommended improvements are detailed below. I encourage the authors to take them into account for a great and improved article.

Reply: We appreciate the reviewer's comments, these insightful inputs have helped to improve the quality of this manuscript. We have made our efforts to address the concerns and made corresponding revisions.

Soil moisture estimation

The authors quote two sources upon which the soil moisture routine is based on: Berghuijs et al, (2016) and Deb et al (2019). Deb et al (2019) use a water balance equation, however they did not use it to calculate soil moisture. I do not see any relevance of this reference at this point. I would recommend following the simple bucket model by Berghuijs et al (2016) to calculate soil moisture, the update version in Stein et al (2020) or to consider a modelled soil moisture product, such as ERA5. They are less prone to water balance errors.

Reply: Thank you so much for the suggestion. We are sorry about the confusion caused by the citation of Deb et al 2019, we have removed it now. Actually, we did follow the simple bucket model of Berghuijs et al. (2016) to calculate soil moisture. We removed the snowmelt component as it has little impact on flood generation in our study region, where the climate is warm and AMFs occur in summer (Yang et al 2020). We have to admit that the estimation of soil moisture is highly simplified, and may not always represent the catchment condition accurately. Thus we normalized it with maximum to reduce the impact of the biases in the estimation but keep the seasonal trend and calculated the multi-year mean to reduce the uncertainties. In this revision, we have replaced the normalized soil moisture with percentile soil moisture following the reviewer's comment (Figure A1-A5). The percentile soil moisture represents the relative saturation, and would be less influenced by the inaccurate representations at the event scale.

We agree with the reviewer that reannalysis data like ERA5 are less prone to water balance errors and could provide more event-based results with its high accuracy. The goal of this study is to present a framework to evaluate the relative contribution of rainfall and soil moisture in flood generation, we mainly focused on mean annual scale, more detailed data with high resolution would be used for validation at specific catchment in our future work. Indeed we have collected climate data with high resolution and reannalysis data to do further analysis at event scale, but this is beyond the scope of this study. We have included thes discussion in Section 4.4 Limitations. Hopefully the reviewer finds our explanation satisfactory.

Yang, W., Yang, H., and Yang, D.: Classifying floods by quantifying driver contributions in the Eastern Monsoon Region of China, Journal of Hydrology, 585, 124767, 2020.

Normalizing precipitation/soil moisture

Precipitation has more of an extreme tail than soil moisture. This is due to the fact that soil moisture has an upper limit, e.g. when the soil is completely saturated. Although this is not currently reflected in the equation used for soil storage calculation, this difference should still be taken into account. Another problem with the current normalisation approach is that some catchments in the study period will have experienced more extreme precipitation events than others, simply due to the small time period. If catchment A has experienced a 100-year precipitation event in the observed time period, but catchment B has not, then the values of catchment B will generally be higher than in catchment A. An approach to reduce this uncertainty is to use percentile values as a form of normalisation instead which is more robust (though still not perfect) to this error.

Reply: We agree with the reviewer that precipitation could have more of an extreme tail than soil moisture. We drew the figures with percentile values following the reviewer's suggestion (Figure A1-A5). As we can see, the trends seen in figures with normalized values also sustains with percentiles values: i.e., the positive correlation between antecedent soil moisture and drainage area, the negative correlation between antecedent soil moisture and drainage area, as well as the correlation between SPR and TWI which are less scatter in percentile plot. We have replaced these figures (Figure 3 – 7) in the manuscript with Figure A1 – A5.

Figure A1: Scatterplot between the drainage area and (a) the percentile of antecedent soil moisture of AMF events (the linear regression for blue dots: $R^2 = 0.46$, p-value<0.001); (b) the percentile of precipitation at the day of AMF events (the linear regression for blue dots: $R^2 = 0.61$, p-value<0.001). The green dots represent the regulated watershed, the cyan dots represent the sites on the main stream, and the rest sites are shown in blue.

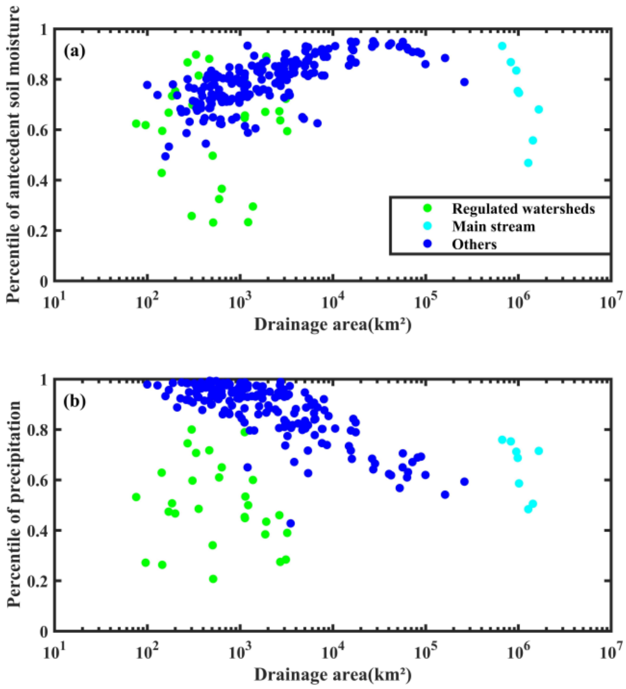


Figure A2: Scatterplot between the drainage area and the percentile of accumulated rainfall of (a) two days; (b) three days; (c) four days; (d) five days; (e) six days; and (f) seven days on AMF events.

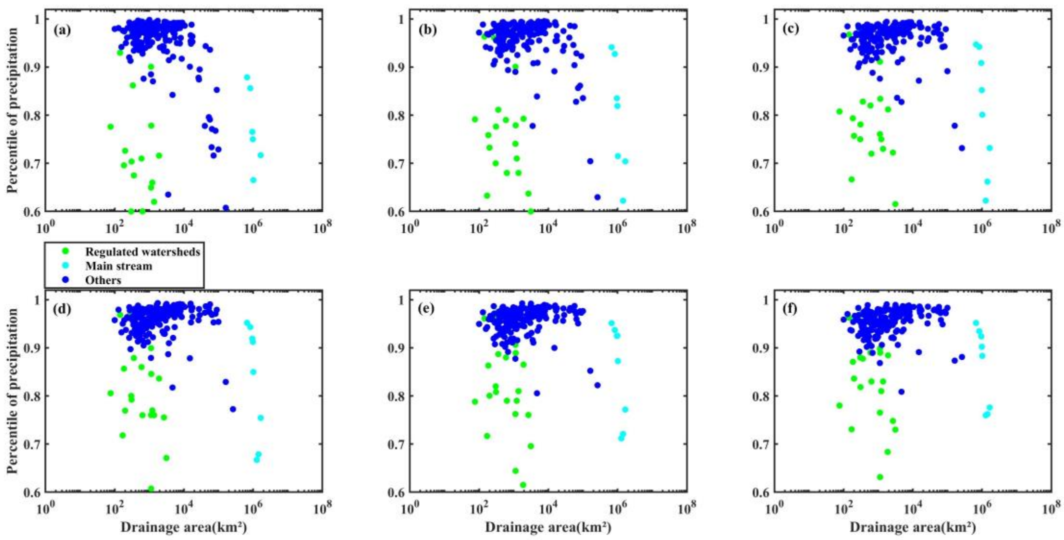


Figure A3: Scatterplot of the percentile of precipitation and antecedent soil moisture, the color represents topographic gradient and the size of circles is scaled by drainage area.

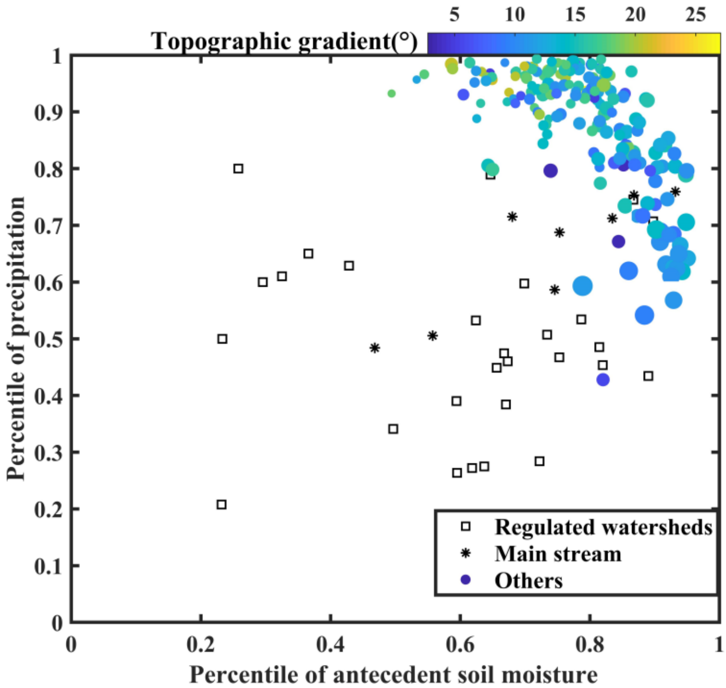


Figure A4: Scatterplots between the ratio of the percentile of antecedent soil moisture and precipitation (SPR) and (a) drainage area; (b) slope; and (c) topographic wetness index (TWI).

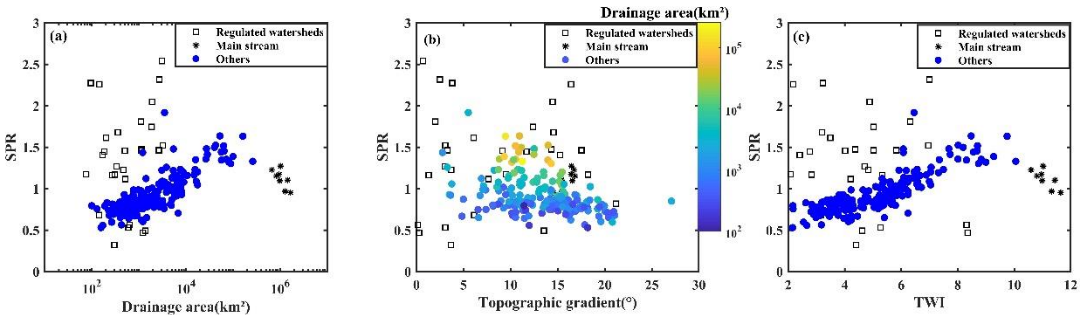
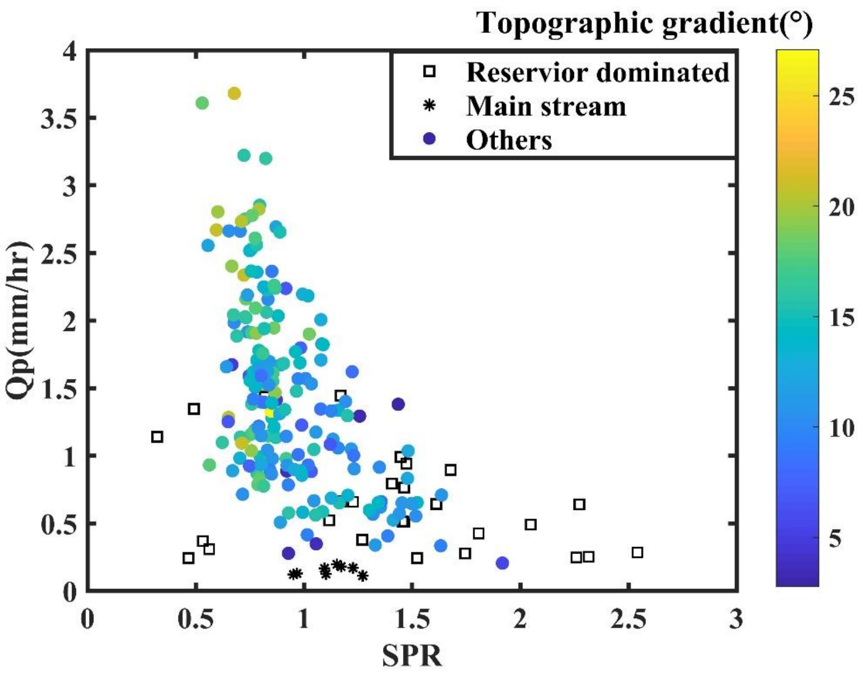


Figure A5: Scatterplot between the ratio of antecedent soil saturation rate and normalized precipitation (SPR) and area weighted annual maximum discharge (Q_P), the color represents topographic gradient.



Section 4.3.

Being able to predict average annual flood magnitude for ungauged catchments would be a valuable discovery. This should certainly be explored further in another study. However, since all results are presented at an average and not event scale, I am not convinced that these approaches would work for flood early warning. For that the diversity of flood generating processes (Stein et al, 2020) is too high and the interplay between soil moisture and precipitation too diverse (e.g. Figure 5b, Saffapour et al, 2016). Just because a catchment is dominated by soil moisture, does not mean that an extreme precipitation event will not cause a flood. I would therefore recommend removing the discussion around early warning system and focus on predicting mean annual flood for ungauged catchments.

Reply: We agree with the reviewer that given the diversity of flood generation, the SPR we derived at an average not event scale is not sufficient for flood early warning. This study is to present a framework to quantitatively evaluate the relative contribution of rainfall and antecedent soil moisture at mean annual scale, the potential application on flood early warning would need more detialed analysis at event scale. We have removed the discussion about early warning and focused on predicting mean annual flood for ungauged catchments as the reviewer suggested. Thank you again for your insightful suggestions. We hope the reviewer find our revision satisfactory now.

Minor comments

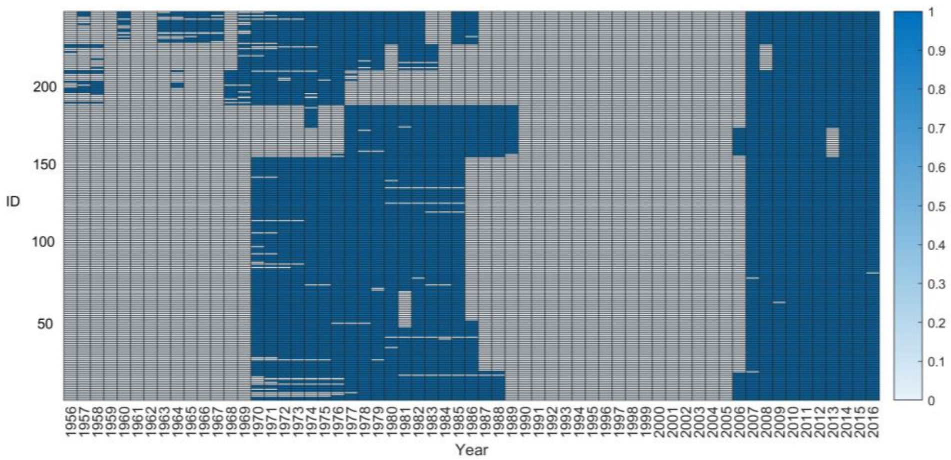
L61: Yang et al, 2020 presented an analysis on flood generating mechanisms in China.

Reply: Thank you for pointing this out, we have included Yang et al 2020 and changed this into: "Such researches were just conducted in China recently, though still limited (Yang et al 2019; Yang et al 2020)." Thank you.

L132: "with at least 20 years records from 1970 to 1990 and from 2007 to 2016 were selected". Unclear. Does that mean that some of the stations only have data between 1970 and 1990, while others only have data between 2007 and 2016? These time periods have likely different climatic conditions and the older ones might have since had dams added to their catchment. Please clarify if my understanding is correct. If yes, please discuss the implications for your analysis and add a Figure to the supplement indicating data ranges for the stations.

Reply: We are sorry about the confusion. Each of the stations used have data from both periods. As we can see from Figure A6, All the stations have at least 9-year data from 2007 to 2016, and at least 11-year data from 1970 to 1990, that is, at least 20-year records in total. We have included Figure A6 in the supplement and rewritten this sentence as: "with at least 20 years records from both the period from 1970 to 1990 and the period from 2007 to 2016 were selected (see Figure S1 for the data availability)."

Figure A6: The data availablity of each station, each column indicates each year while each row is corresponding to each station, blue grid indicates there is record of this year.



L190-193: Can be removed since it repeats information from the Introduction.

Reply: We have removed these two sentences as the reviewer suggested. Thank you.

L200-203, 237-242, 256-266: Please ensure that you are not mixing results and discussion.

Reply: These are brief explanation for the results, which we discussed further in the discussion section. But as the reviewer indicated, it is better to differentiate results and discussion. We have moved most of these discussions in Section 4 now.

L220: There are no red dots on the colour scale in Figure 2. Please clarify.

Reply: It should be 'blue dots', we have corrected it now. Thank you.

L308: Can you explain why the fact that smaller watersheds more easily reach saturation supports that they are less soil moisture dominated? They way the results by Sharma et al (2018) are mentioned might confuse some readers otherwise.

Reply: We are sorry about the confusion. What we want to say is that due to the relative small drainage area, the heterogeneity would be small, and the hydrologic connectivity would be developed quickly, it would be easier to reach saturation in more parts of the catchment. As the reviewer mentioned, this sentence could cause some confusion, we have removed the citation of Sharma and rephrased this sentence as: "This shift of dominance may be attributed to the longer confluence time in the large watersheds and less heterogeneity in small watersheds." Hopefully the reviewer finds our explanation clear.

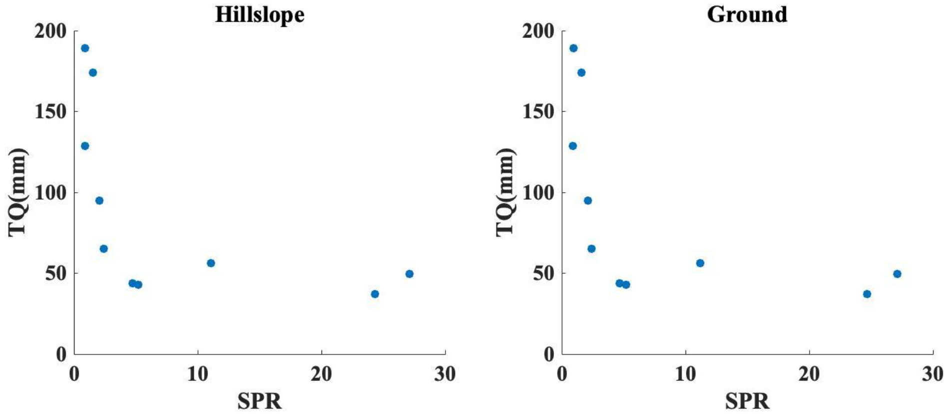
L321-322: The correlation between TWI and SPR is much weaker in the regulated watershed. It will most likely not be sufficient for any form of prediction in those catchments.

Reply: We agree with the reviewer that this correlation between TWI and SPR should only be applied to the 'natural watersheds' not the regulated ones or the ones on the main stream. We have changed it into: "That is, we could derive the relative dominance of soil moisture and rainfall in flood generation in specific watershed from its TWI for the natural watersheds without significant human intervention."

L333-336: Where are the event scale results presented? It would be most interesting to see event scale results as well. Currently, I do not see any evidence that the results can easily be transferred to event scale.

Reply: It was from the field observation in a small mountainous catchment. We had soil moisture measurements at two spots, one on upslope and the other at the bottom, as Figure A7 presents, the SPR at both spots have negative correlation with total event discharge. Since this is a very preliminary results, we just want to say that there is possibility of application of this framework. We agree with the reviewer that so far we don't have convincing evidence that the results can easily transferred to event scale. We have removed the discussion about early warning at event scale and focused on predicting mean annual flood for ungauged catchments as the reviewer suggested.

Figure A7: The scatter plot between total event discharge and SPR at two observation spots.



Figures:

Please try to avoid the use of red and green together when they are the only distinguishing feature. People with colour vision deficiency will not be able to differentiate them. For alternatives, please check Stoelzle & Stein (2021).

Reply: Thank you very much for your kind reminder, we have replaced all the red dots with cyan ones in all the figures.

Figure 5: The scaling of point size according to drainage area is barely visible. Since drainage area is covered in Figure 6b as well, I would suggest to remove this scaling.

Reply: We agree that the size of the points representing the drainage area is not distinct. Since in this figure we also want to discuss the impact of drainage area, so it might be better to include the scaling here. We have amplified the difference of drainage area (Figure A3), hopefully it is clear now.

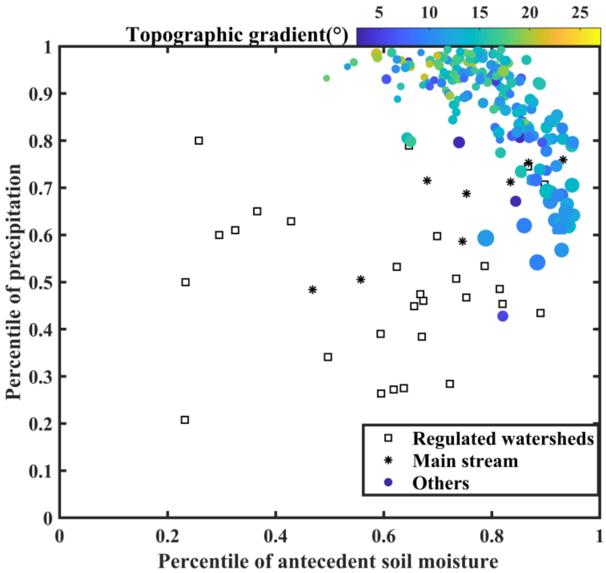


Figure 5: It is unclear what the dashed lines indicate.

Reply: The dashed line is used as illustration of the three groups of catchments: the relative large and flat catchments on the bottom right that are more dominated by soil moisture, the relative small and steep catchments on the upper left that are more rainfall dependent, and the rest of the catchments having floods with heavy rainfall on near saturated soil. But this is just a quanlitative illustration, so we have removed the dashed line to avoid further confusion and focused more on the declining trend between the percential of rainfall and percentile of antecedent soil moisture.

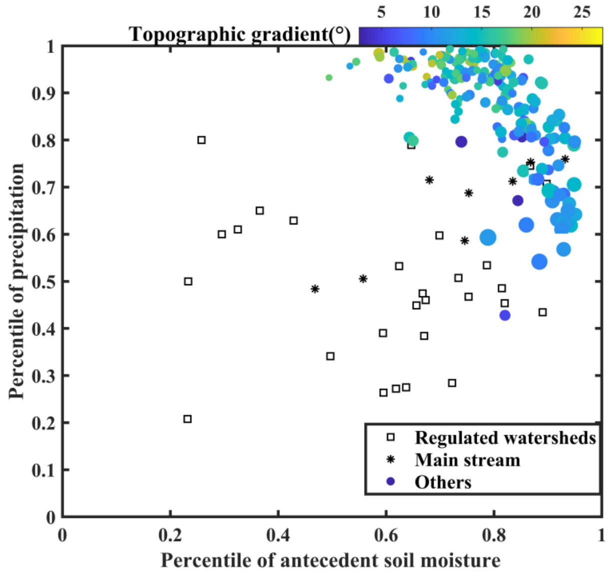
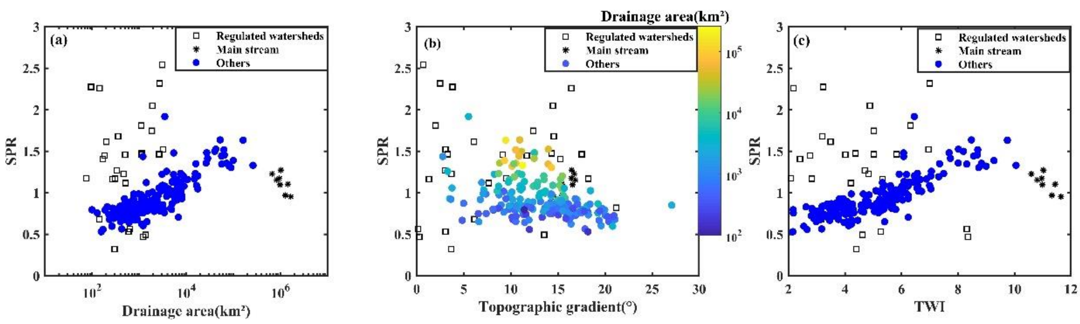
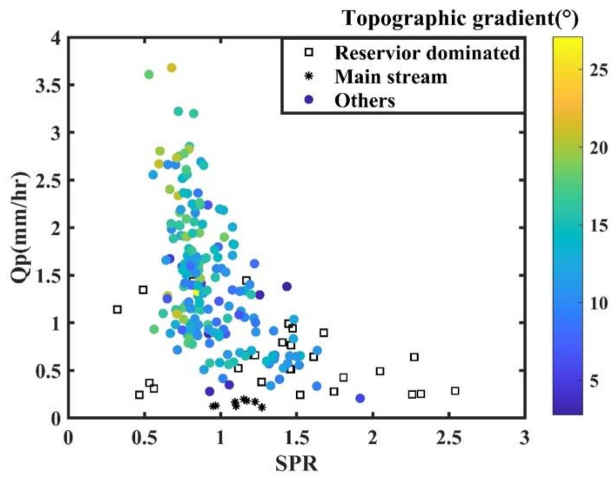


Figure 6b and 7: Since the text talks only about topographic gradient and not slope I would recommend using the same terminology in the Figures.

Reply: Thank you for the comment. We have replaced slope with topographic gradient in Figure 6b and Figure 7.





Saffarpour, S., Western, A. W., Adams, R., and McDonnell, J. J.: Multiple runoff processes and multiple thresholds control agricultural runoff generation, Hydrol. Earth Syst. Sci., 20, 4525–4545, https://doi.org/10.5194/hess-20-4525-2016, 2016.
Stein, L., Pianosi, F. and Woods, R., 2020. Event-based classification for global study of river flood generating processes. Hydrological Processes, 34(7), pp.1514-1529.
Stoelzle, M., & Stein, L. (2021). Rainbow color map distorts and misleads research in hydrology–guidance for better visualizations and science communication. Hydrology and Earth System Sciences, 25(8), 4549-4565.

Reply: Thank you for the literature input. We have included these in our manuscript now, thank you!