Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC1
https://doi.org/10.5194/hess-2022-151-RC1, 2022

Review of the manuscript „River flooding mechanisms and their changes in Europe revealed by explainable machine learning"

Larisa Tarasova (Referee)

Referee comment on "River flooding mechanisms and their changes in Europe revealed by explainable machine learning" by Shijie Jiang et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2022-151-RC1, 2022

Review of the manuscript „River flooding mechanisms and their changes in Europe revealed by explainable machine learning"

This manuscript proposes a new method for classification of flood generation mechanisms using machine learning that provides the information on the importance of different indicators on the generation of the particular flood event. The method has a potential to overcome the subjective choice of classification thresholds of the previously developed methods. It was tested across European catchments with particular focus on how flood mechanisms were changing in the past decades.

The proposed method has certainly some clear advantages compared to the previous methods and provides a new perspective on classification of flood events. It a very well-written manuscript and I have enjoyed reading it very much. I am certain that this is a substantial contribution to the current knowledge on flood generation processes and their changes. However, although the proposed method has the potential to avoid subjective thresholds, I do not think that the authors have succeeded in overcoming this issue completely. A more prominent attention has to be paid to this issue in the manuscript. I also have some minor suggestions that might help to improve the manuscript and clarify its novelty. Please see my detailed comments below.

Kind regards,

Larisa Tarasova

General comments

Abstract: I think the abstract puts too much focus on the changes in flood mechanisms in Europe that is not the most novel findings and instead fails to elaborate on the machine learning approach used here and its advantages compared to previously existing methods. The method implemented in this study is the real novelty, while there are already several studies on changes in flood generation processes in Europe. Therefore, I suggest the authors to consider to put more stress on the methodological aspects in abstract to show how this study stands out.

Selection of thresholds: The proposed methodology has a very strong advantage that it can avoid arbitrary decisions on how the indicators and their threshold are selected for the event classification. However, the authors did not avoid that issue as they have selected the periods for which the effect of recent and antecedent precipitation was accumulated to avoid additional computational effort. This pragmatic choice is understandable and is in line with the subjective choices previous classification studies were making, but it has to be properly stated in the manuscript and a sensitivity analysis on the effect of this choice on the results of the study will be very welcome. Please also see my detailed comments to the corresponding part of the manuscript.

Detailed comments

Line 40-42 I suggest to also mention here the study of Kemter et al (2020) on changes of flood mechanisms in Europe and global analysis of Stein et al (2020)

Line 48-49: Here I miss mentioning the study of Kemter et al (2020) that did exactly that.

Line 88: Please indicate if the size of catchments was limited to avoid the effect of human influence or was there any other reason for this selection?

Line 100: Please indicate also the lower boundary of catchment sizes to clarify if the size study catchments comparable with the spatial resolution of the hydrometeorological datasets.

Line 103: Please elaborate how the catchment boundaries from two datasets were merged. Are they identical?

Line 104-106: I miss here a more motivated choice for the day duration as an indicator for classification. It seems to me that essentially it is a combination of the location and

day of the year information. Please provide more information on the nature of the preliminary test performed, particularly if other potential indicators were examined.

Figure 2: The interpretation arrow is not so clear, why does it return back to the input layer? At this point in the manuscript the meaning of the integrated gradients for the features is not yet explained and looks confusing in this Figure. Please add clarification in the caption. Consider indicating the target maximum annual flood in panel a as a point and not as a window. The panel c is rather confusing as there is only one event is being displayed in the panels a and b and the cluster plot is not set in any particular space (i.e., the axes are not indicated). Consider omitting this panel, I think that idea of clustering is understandable without this example only brings more confusion.

Line 139-144: I agree that presenting LSTM model in detail is not necessary here, but I think I more detailed explanation on how the structure of LSTM suited for capturing short-term and long-term interaction will be very helpful here, as it can provide the readers with the insights on why particularly this method is more applicable than classifications that are based on subjectively selected thresholds.

Line 145: It is not clear. Does it mean for each catchment? Please clarify.

Line 153: What is the sample in this case? Maximum annual floods? Please clarify.

Line 185-187: I do understand authors' arguments on why they had to make this decision and restrict quantification of the effect to 7 and 180 days only. Although, I find it somewhat disappointing. The authors have stated earlier in the manuscript the main advantage of the proposed ML-based method is that one can avoid selecting subjective indicators and their thresholds. In my opinion selecting here 7 and 180 days is nothing else but exactly that kind of subjective threshold that partially impairs the main advantage of the method. If clustering indeed is very time consuming (which is actually surprising to me as in my experience k-mean clustering is not the most time consuming procedure and computational power is hardly a limitation with cluster resources available) at least a sensitivity analysis has to be performed to analyze how the selection of these thresholds affects the results.

Line 188-189: I cannot agree with this statement. The duration will be strongly affected by catchment size and mechanism. The build-up period of snowmelt floods in larger catchments can take up to several months. I also do not think that the provided reference is up to date. Please revise.

Line 189-190: It is consistent with previous studies, but they also did not examine if these thresholds are appropriate. Please revise.

Line 192: It is not clear what is the role of multiple-peak discharges here and how they were considered. Please clarify.

Line 197-203: Please clarify if clustering procedure was performed for all catchments simultaneously or if they were considered individually. If it was performed simultaneously for all catchments, does it mean that if a catchment has very local and specific mechanisms they likely not to be detected by the procedure?

Line 206, 277, 402 and elsewhere: Are these maximum annual peak discharges? If yes, please indicate it clearly here and elsewhere.

Figure 3: Does this figure display NSE only for annual maxima or for the complete streamflow time series? Please clarify.

Line 294: I think it is a rather a stretch to call streamflow generation that occurs due to excess of soil storage capacity and heavy precipitation as we cannot guarantee that heavy precipitation generates overland flow. In case it is first contribute to increase of soil moisture storage the physical process of streamflow generation will be the same for both drivers. Please revise.

Line 303 and elsewhere: I think "mixed mechanisms" is not an accurate term here as it refers to the occurrence of different mechanisms in the same catchment, but not necessarily simultaneously. Consider using "mixture of mechamisms" instead.

Figure 6: Please add an explanation for the mixtures in the caption. Please also clarify how the classes for two processes are formed, do the corresponding two processes have to generate more than 70% of annual maxima?

Line 338-340: I think it will be helpful to relate here to the findings of Stein et al (2021) (doi: 10.1029/2020WR028300) on the controls of catchment characteristics on the dominance of different flood mechanisms

Line 350: Consider using term "pre-defined" criteria instead of "manual" as it is not so clear.

Table 1: Consider also adding catchment sizes to the comparison as I expect that there is a difference between these studies also in that regard.

Line 379: I think the study of Kemter et al 2020 also should be mentioned here.

Line 381-383: This note would be more helpful earlier before the comparison of the results. Consider moving this part up.

Line 385-389: I think it might be worth mentioning here the work of Tarasova et al 2020 (doi: 10.1029/2019WR026951) that investigates how using different data sources for the same indicator affects event classification

Line 404-407, 427-431: These parts would be more suitable in the dedicated Method section

Figure 7: Please indicate how many catchments are the basis for Sankey plot in the caption. Please also clarify the origins of the p value in the caption. The information provided on methodological aspect of trends in this caption is not sufficient. Please add a corresponding section in the Methods. Panel b: I am wondering if the results of trend analysis are not so clear due to regional differences in the direction of trends. Looking at the results of Kemter et al (2020) it seems that there are disparate trends for different regions that can be obscured when mixed together. Perhaps something worth mentioning in the corresponding text.

Line 455-459: It would be helpful if this information is provided in the dedicated Method section.

Figure 8: It is not clear why the lines of the plot do not correspond to the whole extent of time axis. Please clarify or correct. In region 1 and region 2 it seems that there is certain periodicity in the data, it would be helpful if the authors would add a short discussion on suitability of monotonic trends analysis in such cases. Please also consider adding geographical indications for regions instead of numbers. This will make this figure easier to interpret. Please also add the number of catchments in each of the considered regions.

Line 492, 499. Caption of Figure 9: It is not clear which length is meant here. Please clarify.

Line 486-504: This part is not very well connected to the previous narration and provides yet another new results for which methods were not clearly elaborated in the Method section. Consider omitting it or revise.

Line 539-543: I would recall here how “recent” and “antecedent” precipitation were defined in this study, because despite what this part claims the definition of these two indicators were set arbitrary by selecting corresponding number of days during which the effect was evaluated.

Line 549: The term “perspective of catchment average” is not very clear here without the context. I think it would be clearer to just indicate that these methods did not perform an event-based classification and instead identified one single dominant driver per catchment.

Conclusion section: A statement about the dependence of the results on the performance of the ML model for the proposed classification method would be very welcome in this section. Moreover, same as for abstract more focus on the newly developed ML-based classification method instead of changed in the mechanisms will be welcome here to highlight the novelty of this study.

Line 563-565: I think the authors have to be more cautious here with this statement, because there might be strong regional differences (i.e., there are disparate patterns in precipitation changes in Europe). Moreover, the term extreme precipitation is much more often related to very short precipitation (i.e., less than 1 day), while 7-day long precipitation can substantially affect the storage of the catchment and lead to soil moisture excess floods and hence the resultant magnitude of the flood will depend much more on the initial storage conditions compared to the floods that are generated by short and extreme precipitation. Finally, the authors have examined here maximum annual 7-day precipitation which does not guarantee that this is the same 7-day precipitation sum that have caused a maximum annual flood in the corresponding year.

Editorial comments

Line 264: regions with winter snowpack accumulation

Line 276: catchments associated