Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-54-AC5, 2020

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Future streamflow regime changes in the United States: assessment using functional classification" *by* Manuela I. Brunner et al.

Manuela I. Brunner et al.

manuela.i.brunner@gmail.com

Received and published: 27 May 2020

Reviewer 3: Genevieve Ali

The PDF I tried to upload in our earlier reply does unfortunately not display correctly. This is why we provide the responses in plain text here.

COMMENTS In this manuscript entitled "Future streamflow regime changes in the United States: assessment using functional classification", two main goals are pursued: (1) develop a catchment classification scheme for streamflow regimes, and (2) use this scheme to evaluate changes in future flow regimes. Contrary to the majority of

previously published catchment classification efforts, here the authors decided not to rely on stream-flow indices. Instead, they are using a functional approach via which the shapes of mean annual hydrographs are classified, this in order to retain temporal autocorrelation information. Overall, the manuscript is of appropriate length, well written and with good-quality figures and tables. The dual focus of the manuscript on catchment classification and climate change impact assessment is very interesting, and I agree with the authors about their description of the advantages of functional classification. I did find that a few statements made in the manuscript warranted clarification, and that some details regarding the datasets, process interpretations, or linkages with existing literature were lacking (see specific comments below). With revisions, I believe that this manuscript will be interesting to the HESS readership, and a great addition to our growing body of literature on catchment classification.

Reply: Thank you very much for your thorough review. We provided missing details on the dataset, added a few clarifications, and extended the discussion as suggested. The changes are discussed in detail below.

SPECIFIC COMMENTS N.B.: page and line numbers are noted as PX (page X) and LX (line X).

Section 2.1: Given the international readership of HESS, I think that more detailed information is needed about the catchment selection criteria. For people not familiar with the CAMELS dataset, it is quite unclear what is meant by "minimum human impact": is the human impact assessed in terms of catchment-wide land use (that would mean no agricultural or urban catchment), or river regulation? And how may the answer to that question affect the generalization potential of the manuscript conclusions? In other words, the authors should discuss the limitations associated with not considering human-impacted catchments in the present study... Also, how was the 1981-2018 data period chosen for the analysis?

Reply: Thank you for pointing out the need for clarification. We focused on catchments with minimum human impact to be able to look at the effects of climate change on flow regimes in isolation. The catchments belong to the HCDN-2009 network [Lins, 2012], which is a set of stations deemed suitable for analyzing hydrologic variations and trends in a climatic context. The dataset consists of catchments with natural flow conditions undisturbed by artificial diversions, storage, and other activities in the drainage basin or the stream channel and show less than 5% imperviousness as measured by the National Land Cover Database [Jin et al., 2013]. We added this information to the text. We also specified that the period 1981-2018 was chosen 'as data for this period was available for most stations in the dataset'. As suggested, we added a discussion on the limitations of the classes in the case human-impacted catchments are of interest: 'The streamflow regime classes identified here do not comprise classes of catchments with major flow alterations as the clustering was performed using streamflow regimes from catchments with minimal human impact. The five classes proposed here are therefore of limited use if a problem requires including catchments with strong human flow alterations. A flow regime of a regulated stream may still be attributed to one of the five regime classes identified if the altered regime shows similarities with the flow seasonality and variability of one of the 'natural' classes. However, if flow alteration leads to the emergence of regimes clearly distinct from those observed under natural conditions, additional regime classes would be necessary. In addition, the relationships between catchment characteristics and class memberships would need to be revised to enable the assignment of ungauged catchments to one of the classes in the updated set.'

P5 L120: There is a reference to characteristics with missing values. Which characteristics (or types of characteristics) are the ones with missing values? Did omitting them lead to biased results?

Reply: Among the 33 characteristics available, 2 had missing values (i.e. 'second most common geologic class in the catchment' and 'subsurface porosity'). They both belong to the class of geological characteristics comprising 7 characteristics in total,

which means that we were still able to consider 5 characteristics related to geology. We found, however, that these geological characteristics were of minor importance for explaining regime class membership. We specified the two classes with missing values in the manuscript.

Section 2.5, specifically L180-182: How was the comparison made, exactly, from a quantitative or statistical standpoint? Using contingency tables or crosstabs? Or something else? This is a bit unclear to me.... Maybe because I was expecting a statistical comparison when in fact, it is not what was done...

Reply: We checked whether the predicted future class corresponded to the class of the reference simulation. The outcome of this check is binary: 0: predicted future class corresponds to reference class, 1: predicted future class differs from reference class. The results of this comparison are shown in Figure 6 (bars on the left). For the catchments with regime changes, we then identified the direction of change using a contingency table of counts (Figure 6, colored bars on the right). We specified in the text that: 'We then compare the predicted future classes to the class of the corresponding reference simulation using a contingency table of counts.'

Figure 3: The different (graphical) features of the boxplots should probably be described in the figure caption. I assume that the horizontal black lines refer to the medians.... what do the whiskers represent, though: 1 interquartile range (IQR), 1.5 IQR, min and max values, or something else? Are there no statistical outliers associated with each cluster, i.e., each individual box?

Reply: Thank you for pointing out the need for specification. We added the following text to the caption: 'The black lines in the boxplot indicate the median, the upper and lower whiskers correspond to 1.5 * RIQ, where RIQ is the inter-quartile range. Outliers are not displayed.'

P9 L203-204: That should not be a surprise, given that the flood and drought definitions are hydrograph-based.... or am I missing something?

Reply: The droughts and floods were determined using a threshold-level and a peak-over-threshold approach, respectively while the regime classes was determined using the mean annual hydrographs where extremes are smoothed out. But yes, we would expect some correspondence between the streamflow regime of a catchment and the types of extreme events it experiences. We here show that flood and drought event characteristics of the different streamflow regime classes are indeed distinct (Figure 3 in the manuscript).

P9-10, L207-210: The text description, here, does not underline that strong of a contrast between the weak winter regime and the strong winter regime. Maybe it can be rephrased for the contrast to be expressed more strongly?

Reply: We added the following sentence highlighting the differences between catchments with a weak and strong winter regime: 'Compared to catchments with a weak winter regime, catchments with a strong winter regime lie at higher elevations, show higher fractions of snow and are characterized by larger flood magnitudes.'

P10 L217-218: That would explain why there is such a large degree of spatial contiguity/spatial autocorrelation within each cluster. However, it is a bit unclear to me, from the text, whether a RF classification using climatological variables only performs equally as well as – or better than – a RF classification that used both climatological and physiographic variables.

Reply: You are right, we did not discuss whether the random forest model profits from including additional physiographical characteristics in addition to climatological ones. We just discussed variable importance in the context of the 'full model' including all potential explanatory variables. If physiographical variables are excluded from the random forest model, the prediction error increases from 10% to 12%, which corresponds to a marginal decrease in model performance. So yes, a random forest classification using climatological variables performs almost equally as well as a model also including physiographical variables. We added the following sentence to the text: 'Excluding

these physiographical explanatory variables from the random forest model results in only a small decrease in prediction performance (prediction error 12%).'

P10 227: The authors stated that "However, our clustering scheme avoids the formation of very small clusters seen in Jehn et al. (2019)." First, what might explain this? Second, the authors seem to imply that having very small clusters is an inconvenient, and I am not sure I agree – very small clusters could represent very local conditions or hotspots, which are real. The authors should either rephrase or at least nuance their statement to clarify what they mean.

Reply: We did not intend to suggest that forming small clusters is necessarily a bad thing and therefore reformulated the sentence using neutral wording:' However, our clustering scheme results in larger clusters than the ones seen in Jehn et al. (2019).' This is mostly related to the fact that we chose to work with fewer clusters. If we further increased the number of clusters to e.g. 7 instead of 5 clusters (Figure 1 in this response to the reviewer), we would also introduce very small clusters. We would further split up the melt-regime cluster and the New-Year's-regime clusters. This does, however, not further improve cluster distinctiveness as measured by the mean silhouette width.

P10 L230-234: The authors wrote that "The strong link between regime classes and meteorological and physiographical catchment characteristics allows for the attribution of ungauged catchments, where streamflow data are not available, to one of the regime classes, which is potentially very useful for the prediction of streamflow characteristics in ungauged basins". I am not sure where that statement is coming from, as ungauged catchments were not examined in the present study. I agree that the present study might have interesting implications for predictions in ungauged catchments, but this statement, as written, reads as a result when in fact it is an interpretation. In the same line of thought, I wonder whether it would be possible to have separate Results and Discussion sections in the manuscript. There are a few instances, in the text, where it can be tricky to distinguish whether a plain result/fact is being stated, or whether a hypothesis/interpretation is being put forward.

Reply: We split up the Results and Discussion section into two sections in to more clearly distinguish between the results of our study and their implications. It is correct that the focus of our study is not on prediction in ungauged basins. However, we show that a random forest model fitted to climatological and physiographical characteristics is well able to attribute a catchment to one of the regime classes without having any information on streamflow (class prediction error 10%, see l. 118-124 and l.215-119). We add the following sentence to the methods section: 'To further investigate the physiographical and climatological controls on regime class membership and to check whether regime classes can potentially be predicted for ungauged catchments, we perform a random forest classification'. Thanks to its low prediction error, this random forest model enables attributing of ungauged catchments to one of the regime classes. As we do not go into detail on this aspect, we moved the statement to the new Discussion section and clarified it as follows: 'The strong link between regime classes and meteorological and physiographical catchment characteristics enables attributing ungauged catchments, where streamflow data are not available, to one of the regime classes. This attribution can be achieved by using the first random forest model fitted in this analysis enabling predictions of regime class membership using physiographical and climatological characteristics. The ability to attribute of an ungauged catchment to one of the regime classes is potentially very useful to predict of streamflow characteristics in ungauged basins.'

Figure 4: This figure is quite interesting but the comparison of "climate sensitivity" between observations and simulations appears quite qualitative. I wonder: 1) How were the five example catchments showcased in this figure chosen (or, are those sites representative of median cluster conditions)?; and 2) Was a quantitative method of comparison between observations and simulations used for all catchments?

Reply: We chose one regime per cluster and the sites do not necessarily represent median cluster conditions. Yes, we also applied a quantitative method to evaluate 'climate

sensitivity' over all catchments (l.241-243). We added that: 'The sensitivity gradients are computed on the response surface of each catchment in the horizontal direction for temperature and in the vertical direction for precipitation.' The results of this quantitative evaluation are summarized in Figure 2 in this response to the reviewers. The statement: 'Higher mean precipitation leads to higher mean discharge independent of the catchment and regime. The reaction of streamflow to temperature, however, seems to depend on the catchment because the relationship between mean temperature and mean discharge is generally weak and can be positive or negative. (l.238-240)' can therefore be generalized to the entire dataset. We preferred to show the sensitivity grids for a few catchments as we think that these examples nicely illustrate the mechanisms we see for the whole dataset.

P11 L240: The authors refer to a "visual analysis"; were all plots for all 605 catchments visually analyzed?

Reply: We computed such sensitivity grids for each catchment and used them to compute horizontal and vertical sensitivity gradient as outlined in the response to the previous question.

P11 L243-244: The Methods section should explicitly state what the Kolmogorov-Smirnov test was used for, the assumptions being it, and the null and alternate hypotheses (so that readers know what the test results mean). Also, a test cannot be rejected: we can only reject or fail to reject a null hypothesis, so that sentence should be reworded.

Reply: We rephrased the sentence to: '(Kolmogorov-Smirnov test does not reject the null hypothesis that observed and simulated gradients were drawn from the same continuous distribution at level of significance alpha=0.05.)'

Figure 5: Lines are a bit difficult to distinguish on this figure; making it larger and changing the symbology might help.

Reply: Thank you for your suggestion. We rearranged the plot into a 3 rows, 2 columns format to increase the size of the individual subplots. In addition, we darkened the color of the control regime to increase contrast with respect to the regimes simulated using the GCM output.

P12 L258-259: The authors wrote “In contrast, regimes with a strong seasonality such as strong winter and New Year’s regimes are well simulated”. What about the melt regime, which is also highly seasonal?

Reply: This statement is also valid for melt regimes and we added this regime type to the list.

Figure 7: If the black circles mean no regime change, the legend should state so.

Reply: Yes, black circles refer to no regime changes. We added this to the legend of Figure 7.

COMMENTS SPECIFIC TO DISCUSSION ELEMENTS WORTH INCLUDING IN THE MANUSCRIPT

Discussion comment #1: In the present study, regime clusters appear equivalent to clusters derived based on physiographic similarity and clusters derived based on climatological similarity... this is contrary to studies published by Ali et al. (2012) and Oudin et al. (2010) – in a comforting way, I might add – and this should probably be discussed. The "overlap" or agreement between the different classifications bodes well for using climatic and physiographic information as a proxy for streamflow regime types. The fact that an agreement was found in the present study and not in others may be due to the fact that here, functional data were used instead of select streamflow indices.

Reply: Thank you for suggesting to expand the discussion on this aspect. We added the following discussion point: ‘We find functional data clustering to be a useful tool for identifying clusters of catchments with not only similar streamflow regimes but also similar catchment, meteorological, flood and drought characteristics. This similarity

corroborates findings by Bower et al. (2004) and McCabe and Wolock (2014) who established a clear link between similarity in streamflow seasonality and climatic and physical similarity. However, it is in contrast to findings by Ali et al. (2012) who found that catchments similar with respect to a set of flow indices are not necessarily physically similar. Explicitly including seasonality or information on the temporal autocorrelation of regimes may therefore help to identify clusters of catchments which are not only hydrologically but also physically similar.’ A reference to Oudin et al. (2010) was added to the introduction.

Discussion comment #2: It is not a study limitation per se, but the authors may want to discuss the rationale for using functional streamflow data classification (to preserve temporal information) while NOT using climate time series (e.g., mean annual hyetograph) for classification purposes. When I started reading the manuscript, I was puzzled by the fact that a classification based on temporally autocorrelated data (i.e., whole annual hydrographs) was going to be compared to a classification based on climate indices. In other words, I wondered how the analyses would turn out given that different regions may have similar values of mean annual precipitation, even though the temporal distribution of that precipitation may be skewed in some places but not elsewhere. In the end, the authors found that they could neglect the temporal information included in climate time series and still manage to use that climate information (i.e., the climate index class) as a good proxy for streamflow regime class (which, itself, is based on temporally autocorrelated data). That warrants discussion, I think, as it is a bit counter-intuitive (to me, anyway...)

Reply: Our functional streamflow regime clustering approach is indeed solely based on the mean annual hydrographs and the temporal autocorrelation contained therein. It does not rely on climate time series. The information on climate characteristics is only used to see whether the hydrological regime clusters are also climatologically meaningful. We clarify this in the introduction by saying: ‘This scheme makes better use of the seasonal and temporal information stored in the hydrological regime than index-based approaches and is solely based on streamflow information (i.e. no climatological information is used).' We indeed find that these clusters formed according to mean annual hydrographs are distinct in terms of climate and physiographical characteristics (Figure 3 in the manuscript). The good predictive power of a random forest model in correctly attributing catchments to a regime cluster based on climate and physiographical characteristics supports this (l.215-217).

Discussion comment #3: The authors may want to use the concepts of resistance, resilience and synchronicity discussed by Carey et al. (2010): those concepts partly echo what the authors are referring to as "climate sensitivity".

Reply: Thank you for this suggestion. We extend the introduction to the climate sensitivity analysis as follows: 'In the climate sensitivity analysis, we assess whether the hydrological model reacts to changes in mean temperature and precipitation in the same way as observations. In terms of precipitation, this corresponds to checking whether the model captures the resistance of a catchment, i.e. the degree to which runoff is coupled with precipitation Carey et al. (2010).'

EDITORIAL SUGGESTIONS P2 L30: "illustrate the hydrological functioning" seems more appropriate than "govern the hydrological functioning", since the authors are referring to streamflow regimes. P2 L31: I think that the phrase "influencing streamflow variability" should be changed.... Otherwise the whole sentence read as "The characteristics of streamflow regimes [influence] streamflow variability and seasonality", which reads as a circular statement.

Reply: We rephrased this sentence to: 'The characteristics of streamflow regimes, as described here by mean annual hydrographs, include streamflow variability and seasonality and influence the hydrological functioning of a catchment.'

P10 L217: "shows that the the most important variables for" SHOULD BE CHANGED FOR "shows that the most important variables for"

Reply: We eliminated the duplicate 'the'.

P11 L243: "Klomogorov–Smirnov" SHOULD BE CHANGED FOR "Kolmogorov-Smirnov"

Reply: We fixed this typo.

P13 L274: "In contract" SHOULD BE CHANGED FOR "In contrast"

Reply: We fixed this typo.

REFERENCES CITED IN THIS REVIEW Ali, G., Tetzlaff, D., Soulsby, C., McDonnell, J. J., and Capell, R. (2012), A comparison of similarity indices for catchment classification using a cross-regional dataset. Advances in Water Resources, 40, 11-22. doi:10.1016/j.advwatres.2012.01.008

Carey, S.K., Tetzlaff, D., Seibert, J., Soulsby, C., Buttle, J., Laudon, H., McDon-nell, J., McGuire, K., Caissie, D., Shanley, J., Kennedy, M., Devito, K. and Pomeroy,J.W. (2010), Intercomparison of hydroclimatic regimes across northern catchments: synchronicity, resistance and resilience. Hydrological Processes, 24: 3591-3602. doi:10.1002/hyp.7880

Oudin, L., Kay, A., Andréassian, V., and Perrin, C. (2010), Are seemingly physically similar catchments truly hydrologically similar? Water Resources Research, 46, W11558,doi:10.1029/2009WR008887

References used in this response to the reviewer

Ali, G., D. Tetzlaff, C. Soulsby, J. J. McDonnell, and R. Capell (2012), A comparison of similarity indices for catchment classification using a cross-regional dataset, Adv. Water Resour., 40, 11–22, doi:10.1016/j.advwatres.2012.01.008.

Bower, D., D. M. Hannah, and G. R. McGregor (2004), Techniques for assessing the climatic sensitivity of river flow regimes, Hydrol. Process., 18(13), 2515–2543, doi:10.1002/hyp.1479.

Carey, S. K. et al. (2010), Inter-comparison of hydro-climatic regimes across northern catchments: Synchronicity, resistance and resilience, Hydrol. Process., 24(24), 3591–3602, doi:10.1002/hyp.7880.

Jin, S., L. Yang, P. Danielson, C. Homer, J. Fry, and G. Xian (2013), A comprehensive change detection method for updating the National Land Cover Database to circa 2011, Remote Sens. Environ., 132, 159–175, doi:10.1016/j.rse.2013.01.012.

Lins, H. F. (2012), USGS Hydro-Climatic Data Network 2009 (HCDN–2009): U.S. Geological Survey Fact Sheet 2012–3047, Reston, VA.

McCabe, G. J., and D. M. Wolock (2014), Spatial and temporal patterns in conterminous United States streamflow characteristics, Geophys. Res. Lett., 41(19), 6889–6897, doi:10.1002/2014GL061980.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-54, 2020.

Clusters 7

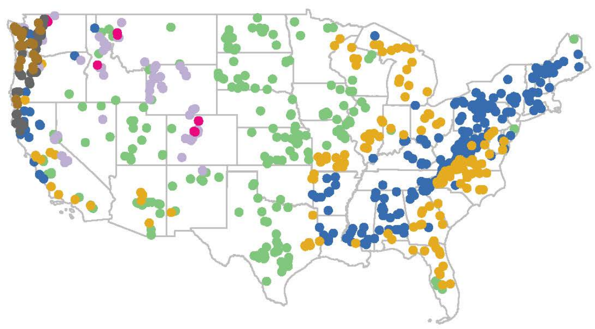

Fig. 1. Map of 671 catchments in the dataset clustered into 7 streamflow regime classes. Each color represents a different class.

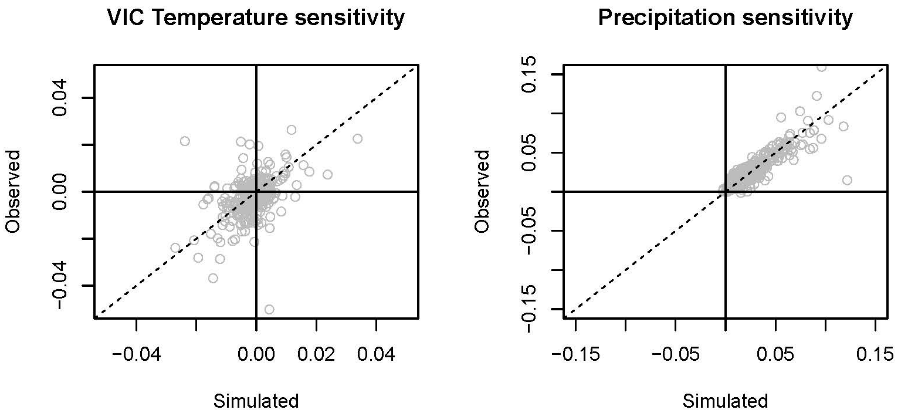

Fig. 2. Observed vs. simulated sensitivity gradients for temperature (left) and precipitation (right) over all catchments computed using climate sensitivity grids as displayed in Figure 4 of the manuscript.