Reply to interactive comment by Anonymous Referee #3 on "Understanding Hydrologic Variability across Europe through Catchment Classification"

Anna Kuentz, Berit Arheimer, Yeshewatesfa Hundecha, Thorsten Wagener

The authors would like to thank Referee #3 for his/her contribution to our paper through his/her interesting remarks. Our replies to the different comments are written below (in blue font) after each comment/question (written in italics).

(1) The purpose of conducting a correlation analysis seems a bit unclear to me. Firstly, if reducing the number of variables (to be used for classification) was the goal, why is it that only physical descriptors were chosen for culling, and not flow signatures? It could easily be argued that some flow signatures (e.g., HFD, LowFr) which do not have high correlation with most physical descriptors can be removed as well. Secondly, as mentioned in Section 2.2, a PCA is performed anyways to reduce the dimensionality prior to classification. So why prescreen the variables with correlation analysis before applying PCA? Wouldn't PCA alone on the whole dataset (16 flow and 48 physical variables) do the job?

Thanks for this remark that seems to agree with reflections from reviewers 1 and 2. This first part of analysis was performed to give a first overview of the links between descriptors and signatures and to closely study the catchment descriptors to decide whether it was reasonable to keep them for further analysis. Flow signatures were selected to describe different components of the hydrological regime (as explained in section 2.1) so all of them were kept along the different steps of the study. However, as the reviewer underlines, some these signatures turned out to have low correlation with most physical descriptors and to be difficult to model. This is an interesting point that we suggest to include in our conclusions for the final manuscript.
The reviewer is completely right when saying that the PCA should be enough to reduce the dimensionality prior to classification; however, the catchment descriptors are used not only for classification but also in the next steps of the study to build linear models and find out which are the main physical controls of flow signatures in different types of catchments. For this analysis it was helpful to have previously removed the less correlated catchment descriptors.
Finally, (also mentioned in our reply to reviewer 2) even though the correlation analysis was a useful introduction for us to start the study, we agree that section 3.1 may be confusing and a bit redundant so we suggest to remove it for the final manuscript, by moving the graphics to the supplementary material and only state the main conclusions of this part of the study in the introduction to Results in the main text.

(2) In Section 3.2, the geographical patterns of classification are briefly mentioned for the physical descriptors based classification, and not at all for the flow signatures based one. I think the authors have a huge opportunity here to explain the geographical context of the spatial patterns observed in Fig 3a and b. It is mentioned (Page 11, Lines 13 and 14) that the flow and physical descriptor based classifications lead to different patterns. Why is that? Any speculation on this aspect would be quite helpful here because it directly relates to the main questions asked in this study.

We thank the reviewer for the comment. We tried to keep this description of the two first classifications short to reduce the overall length of the paper and focus more on the third "combined" classification. We agree that describing and comparing the spatial patterns of the first two classifications would be of interest. However, we were not expecting the two classifications to be similar since the basis of classification for the two are different. Even a classification based only on catchment descriptors could be different if we added or removed some descriptors. The idea we pursued was to try different classifications and get insight into which one gives us more discrimination of the relationships between flow signatures and catchment descriptors and not trying to seek a

correspondence between the groups established through the different classification methods. Thus, we suggest to better explain the difficulties for such an analysis in the manuscript but not trying to explain the differences in patterns. We will, nevertheless, add more discussion on the spatial patterns of the first two classifications when doing the revision.

(3) It might be helpful to state the proportion of total area covered by each of the 10 classes obtained through CART (Figure 3c). It is mentioned later in Section 3.4 that the regression models used for predicting flow signatures across Europe perform poorly for classes 3, 6 and 8, and perform best for classes 7, 10 and 11. Knowing the % area of Europe covered by poor and good performing classes would clarify the ability of your classification to predict flow signatures in ungauged catchments. Based on a quick look of Figure 3c, it seems to me that your best performing classes are predominantly clustered around the Alps, and majority of the Europe is covered by the poor performing class 3 (and class 6 covers large areas too). Does this mean that after going through all the efforts of two classifications + CART + regression models, our ability to predict flow signatures at ungauged catchments is only limited to wet, mountainous systems (which we already know from previous studies to be simple and easily predictable hydrological systems)?

This is indeed a very good comment and we fully agree on the reviewer's suggestion to try to better quantify how much we actually learn from the classification exercise. We suggest including the proportion of total area for the different classes and extending the discussion on this topic.
The class 3, for which we weren't able to distinguish any determinant flow signatures or catchment characteristics, covers 39% of the map area. As pointed out by the reviewer, the classes where the regression models were performing best cover small areas (resp. 2.4, 2.3 and 2% for classes 7, 10 and 11); however, the regression models showed good performances as well for at least some of the flow signatures in classes 1, 4 and 5 which cover a total of 43% of the study area and are not particularly wet or mountainous systems.
Finally, as written in our reply to reviewer #1, we also want to emphasize that the effort of the classifications and regression models wasn't mainly aiming at predicting flow signatures (even though it is an obvious and interesting use) but at gaining better understanding in the hydrological patterns across the European continent, which we were indeed able to do for most of the continent: 61% of the studied area (all classes except class 3). In the next version of the manuscript, we will try to be more precise on what we actually learnt with respect to predictions of flow signatures at ungauged catchments across Europe.