Reply to interactive comment by M.C. Westhoff on Understanding Hydrologic Variability across Europe through Catchment Classification”

Anna Kuentz, Berit Arheimer, Yeshewatesfa Hundecha, Thorsten Wagener

The authors are grateful to Dr. M.C. Westhoff for his review and his interesting comments and suggestions to improve our paper. The replies to the reviewer’s different comments are detailed below (in blue font) after each comment/question (written in italics).

This paper aims to classify a large set of European catchments using a few different regression, and clustering techniques. The results are analyzed by looking at spatial patterns while the main drivers are characterized for each class.
Although I personally have no record in catchment classification methods, I judge this paper as potentially publishable. But before that, I think the paper can and have to be improved.

The first point I was triggered about was the sentence “So far we have not yet found a widely accepted classification system” (P2, L8), which made me expect that this paper would (or at least aimed) to finalize this issue. However, this is not the case, while I think you can make this attempt by reserving a part of the available dataset for validation. The used dataset is large enough and I think the results would benefit from a “calibration-validation cycle” in which the dataset is split in two randomly chosen sets, of which one is used for calibration and the other for validation. This can be done several times for different randomly chosen subsets. This exercise may tell you more about number of catchments needed in a class and how robust the chosen signatures are.

We agree that the calibration-validation exercise suggested by the reviewer would be interesting. However, our aim is not to come up with a unified classification system that would have a general application. The main aim of the work is to understand the link between different flow signatures and catchment physiographic attributes and whether these links are different for different groups of catchments that can be defined based on certain characteristics. To this end, we employed different established classification approaches to group catchments and assessed which classification leads to identification of a stronger link.
Based on the reviewer’s remark, we feel that our statement about the lack of a widely accepted classification may send a wrong message about the aim of our work. Therefore, we will remove it and try to emphasize that our aim lies in understanding what does control the signatures across a large domain and what we can learn about similarity.

A second aspect was that I had problems understanding what was done and in which order. If I am not mistaken, I think you can roughly summarize it by: 1) With a regression analysis catchment descriptors (CD) are correlated with flow signatures (FS). 2) Classes have been derived using 3 different clustering methods: one using CD, one using FS and one using a CART analysis. 3) For each class, correlations between CD and FS are derived and compared with the correlations derived in step 1. If this is indeed the case, I suggest to add e.g. a flow chart and to turn paragraph 2.2 and 2.3 around.

Thank you for this comment and suggestion that will for sure improve the clarity of the paper. Actually we could write the different steps as follows:

1. correlation analysis giving a first overview of the links between descriptors and signatures and screening of the descriptors (elimination of 13 catchment descriptors without any significant correlation);
2. classification using three different methods;

3. calibration of linear models, on one hand using the whole domain, on the other hand inside each group of the three classifications, and comparison of performance of these different models.

As also raised by Referee #1, the first part about correlation analysis is maybe confusing and a bit redundant so we plan to remove section 3.1, move the graphics to the supplementary material and only state the main conclusions of this part of the study in the main text.
We agree on the suggestion of adding flow chart and will add one.

I very much agree with paragraph 3.4 in which it is suggested that the finding can be used for ungauged basins or to parameterize large scale models. But to really benefit from the results of this paper I would encourage the authors to also publish the regression constants. This would make it possible for others to indeed parameterize large scale models, while other future classification studies can better compare (quantitatively) their results with those of this study.

We thank the reviewer for his interest on this part of our work; we will publish the regression constants in the supplementary material.

Minor comments: Be consistent in using either the term "Catchment Descriptors" or "physiographic control"

We will check this again in the revised version.

P3,L32: Give also the range of the catchment sizes

That is a good suggestion, we will include this information.

P6,L5: explain what E-HYPE is

Thank you for pointing out this oversight! This will be added. E-HYPE is a pan-European hydrological model, more information and some model results are available on http://hypeweb.smhi.se/europehype/long-term-means/

On P3,L11-12 it is stated that "No study so far, to our knowledge, has applied the results from comparative hydrology at the continental scale, also including large rivers with human alteration and ungauged basins", suggesting that this study will include basin subject to human alteration. Now on P6,L12 it is stated that stations with strong flow regulations were eliminated.

When visually checking the hydrographs of each flow station, the catchments with obvious and very strong flow regulation where removed. Though, a part of the catchments used in the study still have various forms of human alteration. This has partly been taken into account with some indices like agricultural area, urban area or irrigated area. Unfortunately we haven't been able to find a good indicator of flow regulation available over the whole Europe but this would certainly be of interest if such an index became available. Nevertheless, impact from regulation was clearly identified in the hydrological interpretation of similarities between catchments in specific groups. This is part of the results (Table 3), which is discussed in Section 3.3.
However, your remark, also supported by a comment from Referee #1 let us think that this sentence unnecessarily stresses human alteration when it's not the main object of our study, so we plan to remove the mention to human alteration here.

P12,L15: It is unclear to me to which method is referred here. Please clarify

Thank you for raising this unclarity, we rewrote the sentence as follows: "When looking at the classification based on catchment descriptors, the average of standard deviations of each catchment descriptor within all clusters was estimated to be 0.71, and the average of standard deviations the flow signatures was 0.78. For the CART classification, these numbers are 0.76 for catchment descriptors and 0.67 for flow signatures."

P12,L8: You mean actual evaporation, right? Also add this at P13,L15 and potential other locations.

This is indeed a lack of precision; we mean actual evapotranspiration and will add this precision where relevant.

P14,L9: Is it possible to quantify the strong relationship?

In both works, regression models were built to estimate BFI using geological classification. Both show that the predictive model for baseflow when geological classification was employed were strong, making a conclusion that geology is the determining factor for baseflow estimation. It is therefore difficult to give figures that quantify how strong the relationship is. In the former work (Longobardi and Villani, 2008) they showed the reduction in the prediction error when accurate spatial variability of geology was used in the classification.