Point-by-point response to the reviewers with corresponding relevant changes

As the three reviews pointed out that the goals of our study could be misunderstood, we have reworked different parts of the introduction and conclusion to better emphasize our aim to gain better understanding about hydrological signature patterns and their controls across Europe. We have also tried to lighten and rephrase different parts of the manuscript to make it more straightforward.

The section 3.1 has been moved to the supplementary material, responding to both reviewers 1 and 3. Section 3.2 ("Catchment classifications and regression analysis") has been divided in two sections ("Catchment classifications" and "Using regression analysis to understand controls on individual signatures"), in order to make the paper easier to read and to highlight our motivation of understanding controls. The usefulness of the regression models for reaching this understanding has been further explained in the introduction of the new section 3.2 ("Using regression analysis to understand controls on individual signatures").

More specific actions taken in regard to the reviewer's comments are written below (in purple) after each comment (in black) and response (in blue).

Referee #1

This paper uses established methods to classify regions with similar physiographic characteristics and with similar flow signatures to determine the best predictive relations at ungauged locations. For this reason, the manuscript reads more as a report rather than a novel contribution to the literature. For example, in lines 13-15, it seems the manuscript goals do not seem to be driven by scientific hypothesis but more by having a large set of data and wanting to develop/explore some relations which may (or may not) be useful at some later point. In this way, I think the motivation for the study seems weak as a scientific contribution. Despite this, I do believe that the novelty of the manuscript is in the application of these methods over such a large spatial domain. As hydrologic modeling efforts expand to cover continental scales, the ability to upscale existing approaches for model calibration across large ungauged regions becomes a limiting factor in these efforts. This point should emphasized more in the manuscript to elevate the impact of the work beyond an application of existing methods to a larger region than had been tested previously.

The reviewer is completely right in his/her assumption that the scientific contribution of this paper is not in developing any new methods – but applying existing methods to learn more about nature, in this case hydrological controls across Europe. This is emphasized both in the Introduction and in the Discussion and Conclusion of the paper.

Action taken: We have reworked different parts of the introduction to emphasize more this aspect.

1. The selection of flow signatures needs more details as to how they were selected. Olden and Poff (2003) do not from my remembering of the paper - as the authors indicate in line 16 - provide 9 signatures. Their paper attempts to reduce redundancy in the 200+ statistics that have been used for hydro-ecologic classification but they do not provide a definitive reduced list. More details need to be provided as to why these signatures were selected, particularly because their usefulness in applications is not part of the analysis in the paper. This reads as quite an arbitrary choice.

Thank you for raising this point, the choice flow signatures was indeed made very carefully in this work. The reviewer is right in claiming that Olden and Poff (2003) do not provide a definitive reduced list of signatures, but they do suggest a way to select such a list when saying "One could reduce the population of indices to a minimum of nine, each of which exhibits the highest absolute loading for the first principal-component axes for each of the nine distinct components of the flow regime (Table III). This ensures that the majority of the variation is accounted for and that different facets of the flow regimes are adequately represented in subsequent analyses. Furthermore, given the particular ecological question being addressed, additional indices within each flow component could be selected (from the remaining significant principal components), which would not result in a substantial increase in redundancy".
According to their suggestion we selected the nine indices with the highest absolute loading for the first principal-component axes for each of the nine distinct components of the flow regime, except for number of zero-flow days describing the component "duration of low events", because this index was to specific to intermittent rivers (a very large majority of the rivers in our domain had a value of 0 for this index). For this component of the flow regime we selected instead the index with the highest absolute loading for the second principal-component axe.

2. In lines 12-13 (p. 3), the comment is made that this type of analysis has not been applied at the continental scale "including large rivers with human alteration…" Do the catchments examined here have human alteration? This is not noted in the methods? Does this bias your results?

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.

Action taken: Human alterations have been mentioned at the beginning of section 2.1, and altered flow at the end of section 2.1.
Note that some impacts of human alteration are analyzed in the discussion part (p. 20, l. 11-15 of the original manuscript)

3. In line 18 (p. 8), the statement is made that "identified gauging stations that should be further explored and filtered out…" Was this actually done?

Thank you for pointing out this imprecision. These stations were filtered out, but no further analysis was done on them yet. The sentence will be modified to make it clearer.

Action taken: the sentence has been changed to "…identified gauging stations that were filtered out for the following analyses."

4. In Section 2.3, how were variables determined to be significant in the regressions? What diagnostics were used? How many variables were allowed to enter in each equation? It may be useful as an explanatory tool to see which variables are significant but to make predictions (which is the *goal of this work), one needs to adhere to good statistical practices. How were these practices followed?*

We agree with the reviewer on the importance of providing this information. The significance testing is described in section 3.1: significance of correlations was tested based on a t distribution with a threshold of 0.05. We agree that this information should be available in section 2.3 as well and will be added l.13. The way the variables (and the number of variables) were selected for each regression is described in section 2.3 p. 8 l. 24-29 (stepwise regression based on the Bayesian Information Criterion). The built regression models were evaluated using statistical measures such as the coefficient of determination. These different steps constitute an established statistical procedure to build and evaluate regression models.

However, we want to point out here that, as stated in the introduction, the main goal doesn't lie in the prediction itself but in gaining better understanding in the hydrological patterns across the European continent. The regression models, like the classifications, are used as a tool to reach this better understanding by exploring the relationships between descriptors and signatures and highlighting the main controls of flow signatures in different types of European catchments.

Action taken:

- The significance has been defined in section 2.3: "Significance of correlations was tested based on a t distribution with a threshold of 0.05."
- We have reworked different parts of the introduction and conclusions to better emphasize our actual aim of understanding the hydrological patterns and their main controls.

5. I found myself questioning the value of Section 3.1. I do not think this offers any additional information beyond what can be determined from the CART and regression analyses. This section also contributes to the manuscript reading more as a report as this section seems to explain what could be characterized as exploratory data analysis that is completed before one settles on an approach and hypothesis to test. I also think that the manuscript is a bit laborious in its reading and removal of this section would help streamline the manuscript.

Thank you for this suggestion for improving the readability of the paper. 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 (13 of them were removed). We agree to move most of section 3.1 to the supplementary material and only state the main conclusions of this part of the study in the main text.

Action taken: The section 3.1 has been moved to the supplementary material. This analysis has been mentioned in section 2.3 together with the exclusion of some of the catchment descriptors.

6. Section 3.2 seems to be missing a reference to how the classification was applied to the data. At the very least, reference Section 2.2 to describe how the classification was completed.

We agree with this suggestion, the way the classification was applied to the data is indeed described in section 2.2 and a reference to this section will be added in the result section.

Action taken: The first sentence of section 3.1 (former section 3.2) has been modified to refer to section 2.2: "An automatic clustering based on flow signatures was performed first as explained in section 2.2."

7. I may have missed this but I think it is necessary to develop regression on the flow signatures using the entire dataset to compare to the regression results obtained for the classes. This analysis would determine the objective improvements provided by first classifying the data. If this analysis has been completed, please refer to this in the text when discussing the results.

The reviewer is right in raising the importance of comparing the regressions obtained for the classes with models calibrated using the entire dataset. This has been done in our study as described in section 2.3 (p. 8 l. 21-24). Both regressions using the entire dataset and regressions obtained for the classes are analyzed in the result section 3.2. As following the reviewer's suggestion, a reference to section 2.3 will be inserted in the result section (p. 12 l. 19).

Action taken: A sentence has been added in the result section (first sentence of new section 3.2) to refer to the method description: "As explained in section 2.3, multiple regression models for signature prediction were developed both using the entire domain and within each group of the three classifications and their results were compared."

8. There are two papers that I direct the authors to for potential citation. Singh et al. (2014) used CART to classify model parameter behavior across the United States and may be helpful to motivate some of other contexts in which CART has been utilized for model parameterization at ungauged locations. Oudin et al. (2010) ask almost the same question as this paper in how physiographic similarity is related to hydrologic similarity, although they answer this question using actual model results.

We thank the reviewer for bringing these interesting papers to our attention, they are indeed completely relevant in the context of our work and we will add a reference to them in the revised version.

Action taken: We have included a reference to Oudin et al. (2010) in the introductions section and to Singh et al. (2014) in the methods section.

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

Singh, R., S.A. Archfield, and T. Wagener, 2014, Identifying dominant controls on hydrologic model parameter transfer from gauged to ungauged basins - a comparative hydrology approach, Journal of Hydrology, doi: 10.1016/j.jhydrol.2014.06.030, 2014.

Referee #2 M.C. Westhoff

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.

Action taken: we have modified the sentence in the introduction to not give the impression of trying to find a universally accepted classification system.

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.

Action taken:

- A flow chart has been added at the beginning of section 2.
- The section 3.1 has been moved to the supplementary material. This analysis has been mentioned in section 2.3 together with the exclusion of some of the catchment descriptors.

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.

Action: The regression constants have been added in part E of 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.

Action taken: the use of these terms has been homogenized: "catchment descriptors" refers to the 35 different variables used in the study, some of them turning out to be physiographic controls of the hydrological response in some classes.

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

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

Action taken: the range of area has been included: "…for 35,215 European catchments with a median size (total upstream area of the outlet) of 493 km^2, ranging from 1 to 800,000 km^2 (Fig. 1)."

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/

Action taken: the sentence has been change to "…2,690 flow gauges across our study domain selected based on agreement between catchment size in metadata and the delineation in the pan-European hydrological model E-HYPE (Donnelly et al, 2012)."

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.

Actions taken:

- The sentence in the introduction has been changed to "No study so far, to our knowledge, has applied comparative hydrology at the continental scale, <u>therefore</u> including large rivers with human alteration and ungauged basins."
- More details were added about the elimination of stations with strong regulation (end of section 2.1): "This quality assurance mainly eliminated heavily regulated stations, obviously erroneous hydrographs or wrong time steps (e.g. monthly), still keeping stations with moderately altered flow".

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."

Action taken: as described in the answer.

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.

Action taken: "Actual" has been added 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.

Referee #3

(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.

Action taken:

- More discussion has been added in section 3.4 about which flow signatures were easier or more difficult to model.
- The section 3.1 has been moved to the supplementary material. This analysis has been mentioned in section 2.3 together with the exclusion of some of the catchment descriptors.

(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.

Action taken:
The introduction has been modified and a sentence has been added to section 3.1 to precise that we are not expecting to find similar patterns with different classification systems:

- In the introduction: “*Furthermore, the two approaches do not necessarily group catchments in the same way since the data sets used for the classification are different. Therefore, one needs to derive functions that link flow characteristics and catchment attributes within each group of catchments classified in either way. Ultimately, we believe that a catchment classification framework has to achieve the advantages both approaches offer to be useful […].*”
- In section 3.1: “*Correspondence between the two classifications is not expected as the two classifications were performed using different sets of data*.”

(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.

Action taken:

- The percentages of map area covered by each class have been added in Table 3.
- More details (type of catchments) about the "well performing" and "poorly performing" classes have been added in section 3.4.
- The third conclusion point has been developed with more information about the concerned regions.

Marked-up manuscript

Understanding Hydrologic Variability across Europe through Catchment Classification

Anna Kuentz[1], Berit Arheimer[1], Yeshewatesfa Hundecha[1], Thorsten Wagener[2, 3]

[1] Swedish Meteorological and Hydrological Institute, 601 76 Norrköping, Sweden
[2] Department of Civil Engineering, University of Bristol, BS8 1TR, Bristol, UK
[3] Cabot Institute, University of Bristol, UK

Correspondence to: Anna Kuentz (anna.kuentz@smhi.se)

Abstract. ~~This study contributes to better understanding the~~We studied physical controls on spatial patterns of pan-European flow signatures ~~– taking advantage of large open datasets for catchment classification and comparative hydrology. We explored~~by exploring similarities in 16 flow signatures and 35 catchment descriptors ~~across entire Europe. A database of catchment descriptors and selected flow signatures was compiled~~ for 35,~~–~~215 catchments and 1,366 river gauges across Europe. Correlation analyses and stepwise regressions were used to identify the best explanatory variables for each signature ~~resulting in a total of 480 regression models to predict signatures for similar catchments~~. Catchments were clustered and analyzed for similarities in flow signature values, physiography and for the combination of the two. ~~From the statistical analysis, w~~We found: ~~(i) about 400 statistically significant correlations between flow signatures and physiography;~~ (i~~i~~) a 15 to 33% (depending on the classification used) improvement in regression model skills ~~using~~ when combined with catchment classification versus simply using all catchments at once.~~ the full domain; and~~ (ii~~i~~) 12 out of 16 flow signatures ~~to be~~were mainly controlled by climatic characteristics~~, while topography was the main control for flashiness of flow and low flow magnitude, and geology for the flashiness of flow.~~, especially those related to average and high flows. For the base-flow index, geology was more important and topography was the main the main control for the flashiness of flow. For most of the flow signatures, the second most important descriptor is generally land cover (mean flow, high flows, runoff coefficient, ET, variability of reversals).

(iii) ~~Classifying catchments based on flow signatures or on physiographic characteristics led to very different spatial patterns, but~~Using a classification and regression tree (CART)~~ allowed us to predict flow signatures-based classes according to catchment physiographic characteristics with an average percentage of 60% of correctly classified catchments in each class. As a result~~, we further show that Europe can be divided into ten classes with both similar flow signatures and physiography. The most dominant separation found was between~~We noted the importance of separating~~ energy-limited ~~catchments from~~and moisture-limited catchments~~ to understand catchment behavior~~.~~ For improved understanding, we interpreted characteristics in hydrographs, flow signatures, physiography and geographical location to define dominant flow-generating processes.~~ The CART analyses also separated different explanatory variables for the same class of catchments. For example, the damped peak response for one class was explained by the presence of large waterbodies for some

catchments, while large flatland areas explained it for others catchments in the same class.~~We found that rainfall response, snow-melt, evapotranspiration, damping, storage capacity, and human alterations could explain the hydrologic variability across Europe. Finally, we discuss the relevance of these empirical results for predictions in ungauged basins across Europe and for dynamic modelling at the continental scale.~~ In conclusion, we find that this type of comparative hydrology is a helpful tool for understanding hydrological variability, but is constrained by unknown human impacts on the water cycle and by relatively crude explanatory variables.

1 Introduction

Hydrological systems exhibit a tremendous variability in their physical properties and in the hydrological variables we observe such as streamflow and soil moisture patterns (Bloeschl et al., 2013). At the catchment scale, we assume (or at least hope) that the aggregated response behavior, e.g. the hydrograph, is related to average or dominating characteristics and that smaller scale differences are less relevant. Although the extent of validity of t~~T~~his assumption can be questioned (Beven, 2000; Oudin et al., 2010), it is the basis for statistical hydrology where it allows us to regionalize certain flow characteristics related to floods or low flows. We generally make the same assumption in the search for a catchment classification framework where ~~we want~~our aim is to group catchments that somehow exhibit similar hydrologic behavior (McDonnell and Woods, 2004). ~~So far we have not yet found a widely accepted classification system though~~ While the preferred classification system will depend to a degree on the specific objective of a study or the data availability, it is generally agreed upon that even the search for such an organizing principle is an important undertaking for hydrology (Wagener et al., 2007).

Many studies have attempted to ~~A range of approaches have been taken to~~ organize the catchments we find across our landscape. Approaches include the use of physical and climatic characteristics (e.g. Winter 2001; Brown et al., 2013; Buttle, 2006; Leibowitz et al., 2016), ~~or~~ the use of hydrologic signatures (e.g. Ley et al., 2011, Olden et al., 2012; Sawicz et al., 2011; Singh et al., 2016), or by also including water quality (Arheimer et al., 1996; Arheimer and Lidén, 2000). ~~–~~The advantage of the first approach is that physical characteristics such as topography and land cover are now available for any location on earth (though with varying quality of the data available), while the second approach groups catchments directly by the characteristic we mainly care about, i.e. their hydrologic behavior (see discussion in Wagener et al., 2007). The disadvantages are that the first framework does not ensure that physically/climatically similar catchments also behave similarly, while the second is not directly applicable to ungauged catchments. Furthermore, the two approaches do not necessarily group catchments in the same way since the data sets used for the classification are different. Therefore, one needs to derive functions that link flow characteristics and catchment attributes within each group of catchments classified in either way. Ultimately, we believe that a catchment classification framework has to achieve the advantages both approaches offer to be useful, i.e. it has to be applicable to any catchment and provide insight into its expected hydrological behavior.

Here we assume that flow signatures are one~~a~~ relevant way towards quantifying hydrological behavior and therefore form a sensible basis for a classification framework. They condense hydrologic information ~~that is~~ derived from streamflow observations (alone or in combination with other variables) (Sivapalan, 2005). The choice of the specific signatures used for classification can be guided by: (i) the attempt to describe basic hydrological behavior (e.g. Ley et al., 2011, Sawicz et al., 2011; Trancoso et al., 2016); (ii) the need to relate to societally relevant issues such as floods and droughts (Wagener et al., 2008); (iii) the objective to characterize ecologically relevant characteristics of the catchment response (e.g. Olden et al., 2012); or (iv) in relation to subsequent hydrologic modeling (Euser et al., 2012; Hrachowitz et al., 2014; Donnelly et al., 2016). Studying differences and similarities in flow signatures as well as in catchment characteristics ~~will~~ can also improve our understanding of hydrological processes ~~under current and~~ under potential future conditions (Sawicz et al., 2014; Berghuijs et al., 2014; Pechlivanidis and Arheimer, 2015; Rice et al., 2015). Linking catchment descriptors (physical and climatic) and hydrological response signatures enables the inclusion of ungauged basins and provides the potential for assessing environmental change impacts across large domains.

Despite the significant world-wide research performed during many decades to both understand and predict hydrologic variability using physiography, work has largely addressed small or medium-sized and pristine catchments when delineating regions of similar flow controls (e.g. Yaeger et al., 2012; Ye et al. 2012, Patil and Stieglitz, 2012). Often different studies have resulted in conflicting relationships between some catchment responses and some of their physiographic controls, as a result of catchment size and geographical location. For instance, some studies have found ~~out~~ that forest cover reduces catchment streamflow (e.g. Hundecha and Bárdossy, 2004; Brown et al., 2005; Buytaert et al., 2007), while an increase in streamflow has been found in some others (e.g. Bruijnzeel, 2004). It would, therefore, be worthwhile to identify the physiographic controls of catchment responses and their relationships using a consistent approach across a larger geographic domain, which is subdivided into catchments of different spatial scales. A large sample of observed data from different physiographical and hydrological conditions, enable comparative analysis of dominant drivers for flow generation (Falkenmark and Chapman, 1989). No study so far, to our knowledge, has applied ~~the results from~~ comparative hydrology at the continental scale, ~~also~~ therefore including large rivers with human alteration and ungauged basins.

~~This~~ Our study aims at exploring and understanding the physical controls on spatial patterns of pan-European flow signatures by taking advantage of large open datasets. ~~Better understanding would enhance our ability to predict hydrological variables in ungauged catchments for more efficient water management.~~ We explore the relationships between catchment descriptors and flow signatures by analyzing 35,~~–~~215 catchments which cover a wide range of pan-European physiographic and anthropogenic characteristics. A database of catchment descriptors ~~and selected flow signatures is estimated~~ for all catchments and of hydrologic signatures using 1,366 flow gauges across Europe has been gathered. Based on this database, we make use of a set of established classification and regression approaches to learn more about physical controls of flow generation.~~, providing material for a first level of analyses of statistical and spatial distribution~~. ~~Correlation analyses are subsequently used to identify the best explanatory variables for each signature and to build regression models for predictions in ungauged basins. Catchments are clustered and analyzed for similarities in flow signatures, physiography~~

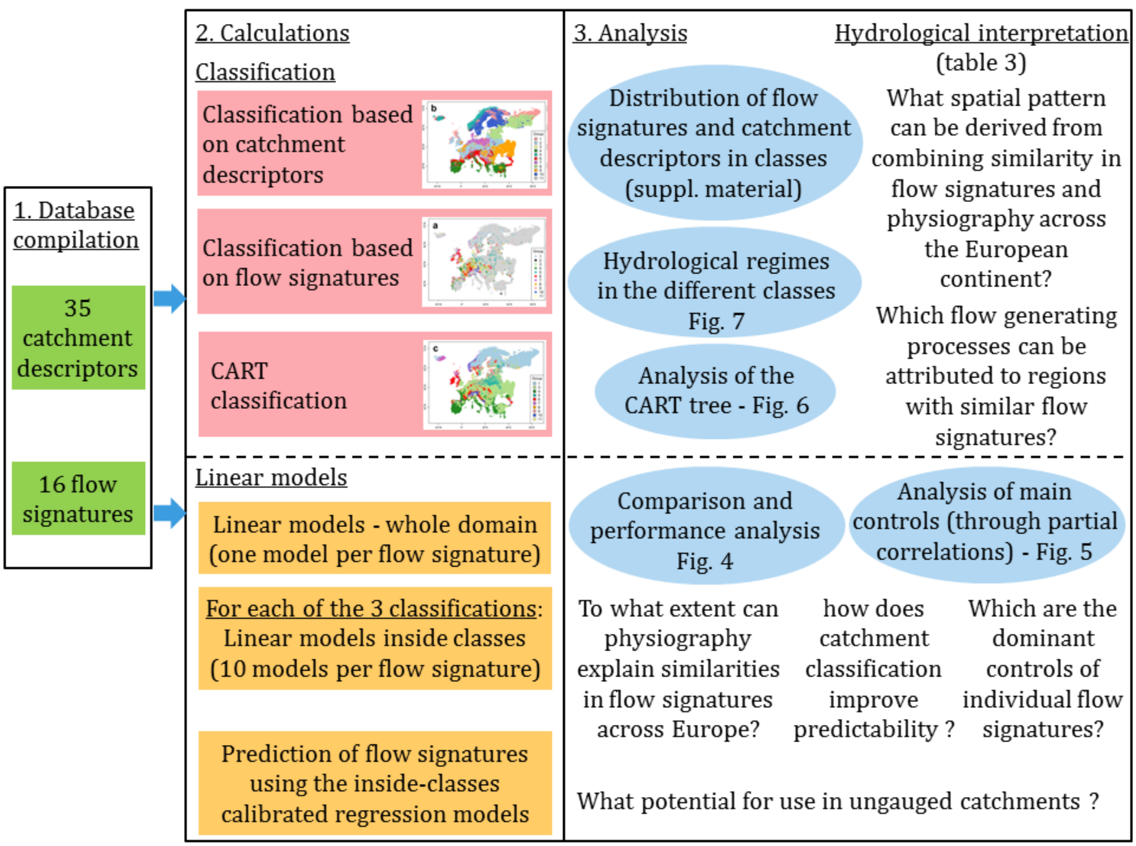
~~and combination of the two, to further improve the predictability and to detect similarities in flow generating processes across the large domain.~~

~~The ultimate aim of our study is to better understand hydrological patterns across the European continent~~Our study is guided by the following science questions:

1. To what extent can physiography explain similarities in flow signatures across Europe?
2. What spatial pattern can be derived from combining similarity in flow signatures and physiography across the European continent?
3. Which flow generating processes can be attributed to regions with similar flow signatures?

2 Data and Methods

This paper summarizes a complex workflow including numerous datasets, calculations, analyses and interpretations, which are summarized in Fig. 1. The data and methods are described in the following sub-sections.



Figur 1: Flow chart of the different steps followed in the study.

2.1. Database of catchment descriptors and flow signatures

A database of ~~physiographical characteristics~~catchment descriptors (climate, physical and human alteration) was compiled for 35,215 European catchments with a median size (total upstream area of the outlet) of ~~214~~ 493 km^2~~-~~, ranging from 1 to 800,000 km^2 (Fig. ~~1~~2). The geographical domain (8.8 million km^2) was delineated according to plate ~~-~~tectonics boarders ~~combined with~~and catchment boarders ~~of rivers~~ all the way down to the European coast and to the Ural Mountains in the East.

For each catchment, 48 catchment descriptors~~physiographical descriptors~~ were assigned using upstream topography, climate, soil types, land ~~use~~cover (including human alterations) as well as geology from open data sources (Table 1). Descriptors were estimated as spatial means of the upstream area and assigned to each catchment outlet.

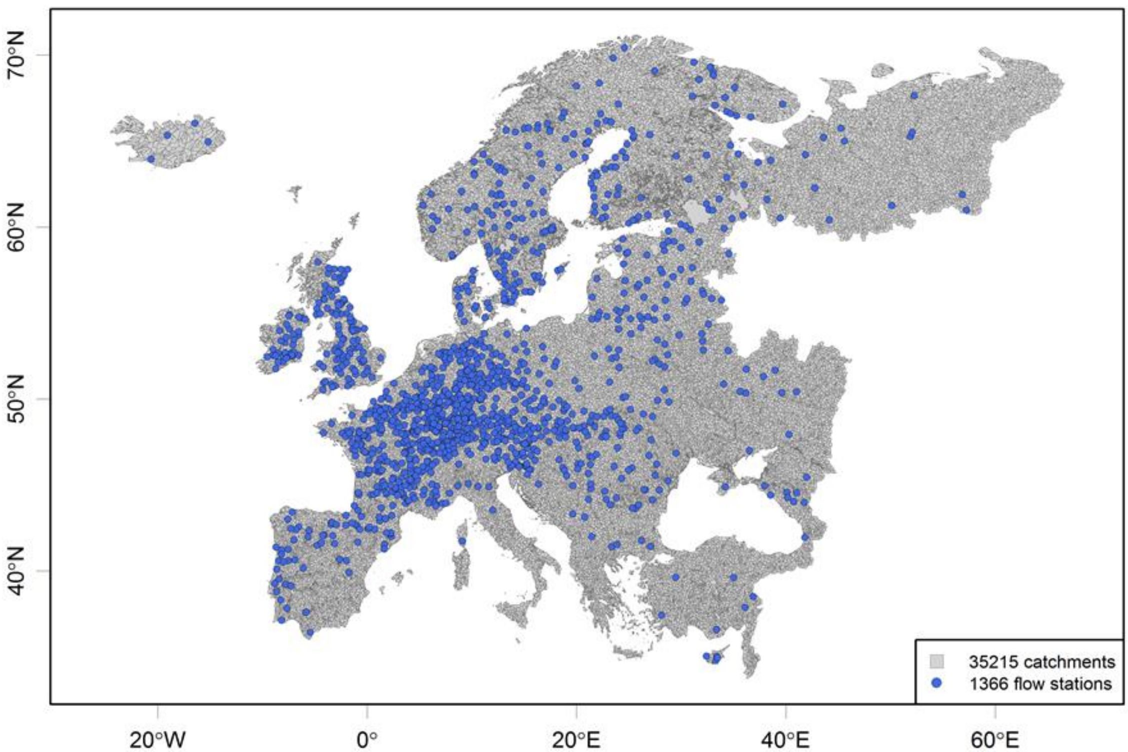


Figure 2~~1~~: Spatial extent of the study showing catchments division and selected river gauges.

Table 1. Catchment descriptors and the original source of information. Type of descriptor is indicated in brackets after variable name (T=topography; L~~U~~C=land ~~use~~cover; S=soil type; G=geology; C=climate). Variables marked with grey color were removed from the analysis because no significant correlation was found between these and the flow signatures (see Section 2.3).

Variable	Unit	Data source	Description
Area (T)	Km^2	SMHI: E-HYPE (Donnelly et al., 2016) http://hypeweb.smhi.se/	Total upstream area of catchment outlet
meanElev (T)	m	USGS: Hydrosheds and Hydro 1K (for latitude >60°) (Lehner et al., 2008)	Mean elevation
stdElev (T)	m	(same as above)	Standard deviation of elevation
meanSlope (T)	-	(same as above)	Mean slope
Drainage density (T)	Km^{-2}	(same as above)	$\frac{Total\ length\ of\ all\ streams}{Area}$
10 Land ~~use~~ cover variables (L~~U~~C)	-	CORINE; GLC2000 (Bartholomé et al., 2005) (for areas not covered by CORINE); GGLWD (lake area, distribution, Lehner and Döll, 2004); EIM (EU scale irrigation, Wriedt et al. 2009); GMIA (global scale irrigation, Siebert et al. 2005)	% of catchment area covered by the following land ~~use~~ cover types: water / glacier / urban / forest / agriculture / pasture / wetland / open with vegetation / open without vegetation / irrigated
7 soil variables (S)	-	ESD (Panagos 2006); DSMW	% of catchment area covered by the following soil types: coarse soil / medium soil / fine soil / peat / no texture / shallow / moraine
21 geological variables (G)	-	USGS Geological maps of Europe and the Arabian Peninsula (Pawlewicz et al, 1997, Pollastro et al., 1999)	% of catchment area covered by the following geological classes: Cenozoic (Cz), Cenozoic-Mesozoic (CzMz), Cenozoic-Mesozoic intrusive (CzMzi), Cenozoic volcanic (Czv), Mesozoic (Mz), Mesozoic-Paleozoic (MzPz), Mesozoic-Paleozoic metamorphic (MzPzm), Mesozoic intrusive (Mzi), Mesozoic metamorphic~~-~~ (Mzm), Mesozoic volcanic (Mzv), Paleozoic (Pz), Paleozoic intrusive (Pzi), Paleozoic metamorphic (Pzm), Paleozoic-Precambrian (PzpCm), Paleozoic-Precambrian metamorphic (PzpCmm), Paleozoic volcanic (Pzv), intrusive (i), metamorphic (m), Precambrian (pCm), Precambrian intrusive (pCmi), Precambrian volcanic (pCmv)
Karst (G)	-	World Map of Carbonate Soil Outcrops V3.0	% of catchment area marked as "carbonate outcrop" in the World Map of Carbonate Soil Outcrops V3.0
Pmean (C)	mm	WFDEI (Weedon et al., 2014)	Mean annual precipitation
SI.Precip (C)	-		Seasonality index of precipitation: $SI = \frac{1}{\bar{R}} \cdot \sum_{n=1}^{12} \left| \bar{x}_n - \frac{\bar{R}}{12} \right|$ $\bar{x}_n$: mean rainfall of month n, $\bar{R}$: mean annual rainfall
Tmean (C)	°C	WFDEI (Weedon et al., 2014)	Mean annual temperature
AI (C)	-	Precipitation, Temperature and wind from WFDEI (Weedon et al., 2014)	Aridity Index: PET/P where PET is the potential evapotranspiration calculated with Jensen-Haise algorithm (Jensen and Haise, 1963)

Flow signatures were compiled using daily hydrograph time-series of the Global Runoff data Center (GRDC) and European Water Archive (EWA) databases from initially 2,690 flow gauges across our study domain selected based on agreement between catchment size in metadata and the delineation in the pan-European hydrological model E-HYPE (Donnelly et al, 2012). A subsample of this database was selected for this study according to data availability. In order to ensure the reliability of the analyses on flow signatures, only gauging stations with at least five whole calendar years of continuous daily data have been selected (2016 stations). Others subsamples with longer time series (such as 10, 15, 20, 25, and 30 years) were extracted for result evaluation. No missing data was allowed over the period and the longest continuous time-series was used at each gauge. This means that time periods differ between gauging stations but consistent descriptors of precipitation and temperature were always used to match the observed period. Finally, all hydrographs of the resulting subset of flow gauges were visually checked ~~over~~ for a 10-year period. This quality assurance mainly eliminated heavily regulated stations ~~with strong flow regulation~~, obviously erroneous hydrographs or wrong time steps (e.g. monthly), still keeping stations with moderately altered flow. After this selection, the final set of streamflow stations used in the study included 1,366 gauging stations.

For ~~the set of~~each river gauge~~s~~, 16 flow signatures were computed (Table 2). The choice of flow signatures has been guided by a study by Olden and Poff (2003), which provides recommendations for selection of nine indices describing flow regimes with importance to hydro-ecology. In addition, five flow signatures commonly used in hydrology have been added for comparability (Qsp, CVQ, Q5, Q95, RBFlash) and two variables describing catchment response were calculated (RunoffCo and ActET).

2.2 Cluster analysis for catchment classification

We classified the catchments based on their similarities in 1) flow signatures for gauged sites only, 2) ~~physiographic~~ catchment descriptors, and 3) ~~physiographic~~ catchment descriptors selected from regression tree analysis on the classes identified using method 1.

For the first two analyses, we used the same clustering method ~~analysis approach~~. The catchments were grouped into classes of similar characteristics (of physiography or flow signatures, respectively) using a hierarchical minimum-variance clustering method. The method groups clustering objects (catchments) so that the within class variability is minimized using a combination of the k-means algorithm (Hartigan and Wong, 1979) and Ward's minimum variance method (Ward Jr., 1963). Clustering ~~was~~ started with the k-means algorithm with a large number of classes (50 ~~classes in this work~~) and classes were merged hierarchically using Ward's minimum variance method. Two classes are merged in such a way that the increase in the sum of the within class variance of the classification variables weighted by the respective class size across all classes is ~~the~~ minimal~~um~~. After each merging step, the k-means algorithm was applied to the reduced number of classes. The optimum number of classes was established by evaluating the changes in the sum of the weighted variance of the variables across all classes between successive merging steps. The point where the rate of change becomes steeper was~~is~~ set as the optimum number of classes.

Table 2. Description of the 16 flow signatures studied.

Component of flow regime		variable	Unit	Description
Magnitude of flow events	**Average flow conditions**	skew	-	skewness = mean/median of daily flows
		Qsp	$L.s^{-1}.km^{-2}$	mean specific flow
		CVQ	-	coef. of variation = st. deviation / mean of daily flows
	Low flow conditions	BFI	-	Base flow index: 7-day minimum flow divided by mean annual daily flow averaged across years
		Q5	$L.s^{-1}.km^{-2}$	5th percentile of daily specific flow
	High flow conditions	HFD	-	High Flow discharge: 10th percentile of daily flow divided by median daily flow
		Q95	$L.s^{-1}.km^{-2}$	95th percentile of daily specific flow
Frequency of flow events	**Low flow conditions**	LowFr	$year^{-1}$	total number of low flow spells (threshold equal to 5% of mean daily flow) divided by the record length
	High flow conditions	HighFrVar	-	coef. of var. in annual number of high flow occurrences (threshold 75th percentile)
duration of flow events	**Low flow conditions**	LowDurVar	-	coef. of var. in annual mean duration of low flows (threshold 25th percentile)
	High flow conditions	Mean30dMax	-	mean annual 30-day maximum divided by median flow
timing of flow events		Const	-	Constancy of daily flow (see Colwell, 1974)
rate of change in flow events		RevVar	-	Coef. of var. in annual nb of reversals (= change of sign in the day-to-day changes time-series)
		RBFlash	-	Richard-Baker flashiness: sum of absolute values of day-to-day changes in mean daily flow divided by the sum of all daily flows
Catchment response		RunoffCo	-	Runoff ratio: mean annual flow (in mm/year) divided by mean annual precipitation
		ActET	$mm.year^{-1}$	Actual evapotranspiration: mean annual precipitation less mean annual flow (in mm/year)

We performed classification using 16 flow signatures and 35 of the catchment descriptors, which have some correlation to flow signatures (correlation significance tested on Pearson correlation using a t distribution with a threshold of 0.05). In order to reduce the effect of possible correlations between the different ~~physiographic~~ catchment descriptors or flow signatures, we applied principal component analysis (PCA). PCA enables derivation of a set of independent variables, which could be much fewer than the original variables, thereby reducing the dimensionality of the problem. The number of principal components selected for further classification was fixed so that they account for at least 80% of the total variance of the original variables.

The third classification was done for all catchments – both gauged and ungauged, using a predictive regression tree, so called CART (Breiman et al., 1984), calibrated to match the classes identified with method 1. CART stands for Classification And Regression Trees and gathers algorithms based on recursive partitioning, aiming either at classifying a sample or at predicting a dependent variable (here the class of the flow stations classification) based on a set of explanatory variables (here the set of ~~physiographic variables~~catchment descriptors). ~~The resulting model can be represented as a binary tree: a~~At the different consecutive levels (nodes of the tree), two groups of catchments are divided based on a logical expression using one of the explanatory variables (dominant catchment descriptors). ~~The~~ Our idea was to obtain a classification close to the one based on the flow signatures but available for the whole set of catchment. Using ~~the~~ CART ~~methodology~~, a regression tree was first adjusted to predict the classes of the flow signature classification using criteria based on catchment descriptors, and then this tree was used in a predictive way to classify all catchments in the domain. It was calibrated using an automatic recursive partitioning based on methods described by Breiman et al. (1984) and provided in the R package “rpart” (see Atkinson and Therneau, 2000). CART has been used previously for understanding controls on groupings of catchments in relation to their hydrologic behavior (e.g. Sawicz et al., 2014) or of hydrologic model parameters or model input and their regional predictors (e.g. Singh et al., 2014; Deshmukh and Singh, 2016).

2.3. Analysis of physiographic controls of flow characteristics

To examine the link between physiography and flow regimes across ~~the~~ our geographical domain, matrices of correlation coefficients between all pairs of catchment descriptors and flow signatures were computed using three different correlations: Pearson correlation, Spearman correlation and distance correlation (e.g. Székely and Rizzo, 2009). Significance of correlations was tested based on a t distribution with a threshold of 0.05. This analysis, which results are presented in section A of the supplementary material, revealed significant correlations between some of the variables, generally consistent with our a priori knowledge (e.g. Donnelly et al, 2016). However, a number of catchment descriptors did not show any significant relationship with any of the flow signatures and were thus removed from the set of variables for the rest of the analyses. These variables are written in grey color in Table 1.~~Catchment descriptors, which did not have any significant relationship with any of the flow signatures, were removed from further analysis.~~

The correlation matrices were accompanied ~~with~~ by a visual analysis of scatterplots of all pairs of variables for quality control to avoid disinformation ~~and misunderstanding in the following analysis~~. Statistical distributions of flow signatures were plotted for different subsets of stream gauges according to the minimum length of the period of continuous daily data availability. Unrealistic values, such as runoff ratios above 1, identified gauging stations that ~~should be further explored and~~ were filtered out for the following analyses. Similarly, spatial distributions of all catchment descriptors and flow signatures were plotted as maps. Most of the maps show rather coherent patterns across Europe and could thus be compared to other sources and local knowledge for ~~further~~ additional visual quality control.

To evaluate the importance of catchment classification, we compared performance of multiple regression models when ~~calibrated~~ developed for~~over~~ the whole domain versus those where regressions were derived separately for each class of

~~classified~~ grouped catchments. ~~Multiple regression models were established using a stepwise algorithm for each flow signature as functions of catchment descriptors. This was done for the whole domain and for the three different clustering approaches (above).~~ For a given flow signature, models were explored using a forward regression, starting from a simple model using only the best correlated descriptor (according to Pearson's linear correlation) and up to a model including all descriptors. At each step, the descriptor giving the best improvement ~~of~~ with respect to BIC (Bayesian information criterion) is added, and the algorithm stops when no further improvement can be obtained. The coefficient of determination of each model was then plotted and the final number of variables was determined based on this plot. For a given classification, as many models as the number of classes in the classification were calibrated for each of the 16 flow signatures and their joint performances were evaluated~~. at the scale of the whole set of stream gauges.~~ To be consistent, regression models were only analyzed for clusters with more than 30 gauging stations, and therefore 17 gauging stations (from 2 classes of the catchment descriptors classification and 1 class of the flow signatures classification) were removed from this analysis because they ended up in classes with fewer stations. In total, 480 regression models were ~~used~~ developed in ~~the~~ our analysis. For each classification method and flow signature, we explored the influence of different ~~physiographic variables~~types of catchment descriptors by examining their partial correlations ~~of different types of descriptors~~ in the regression.

To gain better understanding of processes behind the hydrologic variability, we further examined similarities in both flow signatures and catchment descriptors for each of the clusters based on the CART classification. Each cluster was described by ~~analyzing~~ geographical locations, most characteristic physiography and flow regime. Based on this analysis, hydrological interpretation was used to ~~define~~ identify potential drivers of hydrological processes, which are dominant in each cluster. The analysis was assisted by several sources of information for classes and sub-classes, such as boxplots of variability in both flow signatures and catchment descriptors, matrices showing the median characteristics in each class, visualization of hydrographs in diagrams, and mapping spatial patterns geographically (most of this material is found in the supplementary material).

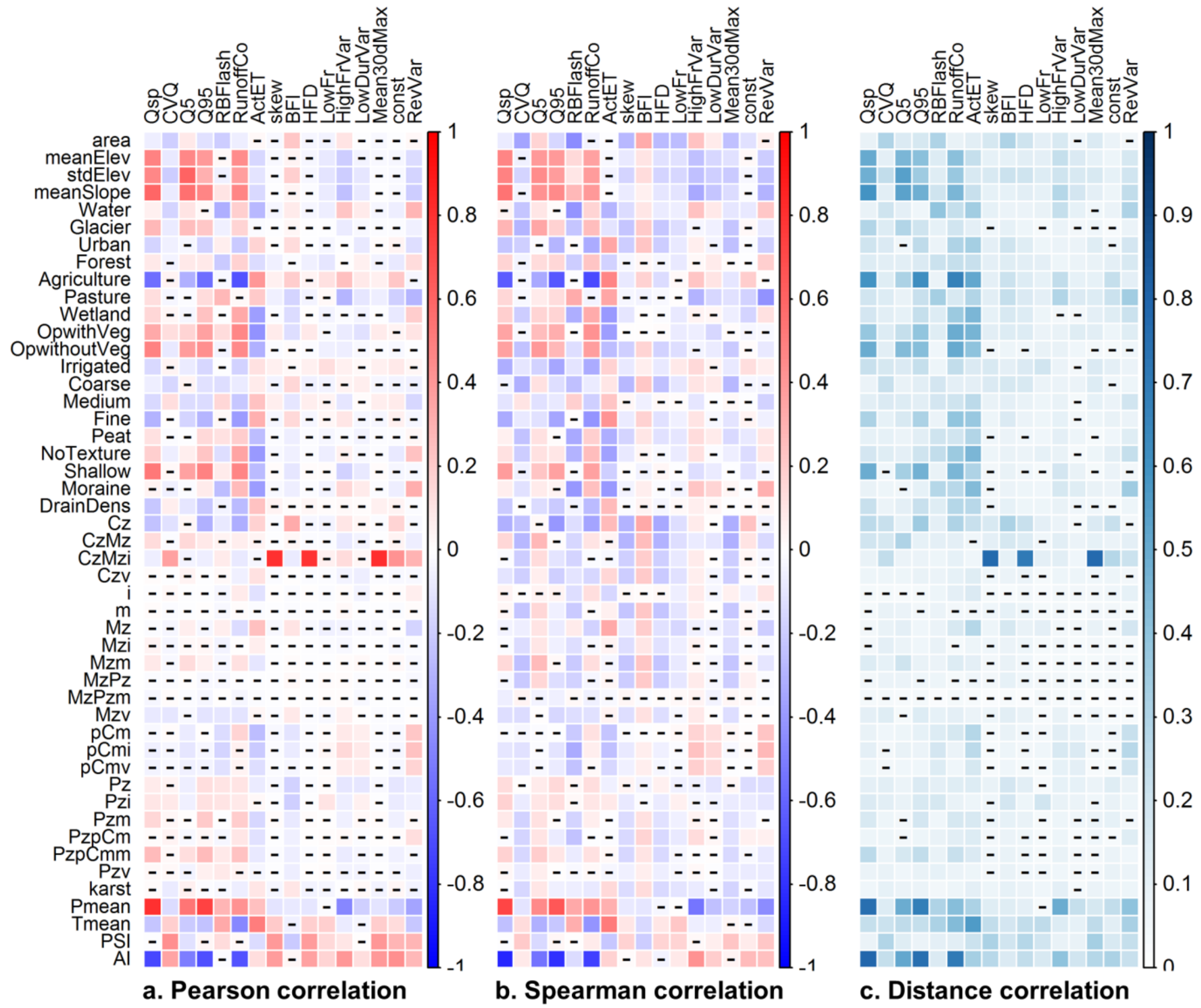
3 Results and Discussion

~~3.1 Correlation analysis~~

~~Significant correlations (significance were tested based on a t distribution with a threshold of 0.05) were found for 400 (out of 786) relations between flow signatures and catchment descriptors from open-data sources (Fig. 2); for instance positive correlation between mean slope and specific flow or low flows, and negative correlation between agricultural area and runoff ratio, and between aridity index and specific flow, 5th and 95th percentiles and runoff ratio. Overall, these relationships seem to be consistent with our a priori knowledge (e.g. Donnelly et al, 2016).~~

~~As shown in Fig. 2, there were no big differences between the three types of correlations compared. As expected, more significant correlations appear when using Spearman correlation than Pearson correlation, but the coloring shows that the differences are not very large. Pearson and distance correlation matrices have similar patterns (putting aside that distance~~

~~correlation cannot be negative). More significant correlations were found when using distance correlation compared to Pearson, but still less than when using Spearman correlation. One exception appears however, which is the percentage of "Cenozoic-Mesozoic igneous" (CzMzi) geological class. This percentage appears to be highly correlated with some of the flow signatures including the skewness of daily flow, the high flow discharge and the mean 30-days maximum. These high correlations (around 0.8) are noticeable in the Pearson (Fig 2a) and distance matrices (Fig. 2c), but absent from the Spearman matrix (Fig 2b). This high correlation led to further examination of the scatterplot between the two variables. It appeared that only a few catchments contain a significant percentage of this geological class and the scatterplot shows that the high correlation is only due to a few points with high values for both "CzMzi" and the concerned flow signature. These high correlations were thus ignored in the following analysis. In addition, combined analysis of the Pearson correlation matrix and the scatterplots showed that a number of catchment descriptors did not have any significant relationship with any of the flow signatures (or only very low correlations – below 0.15). To simplify the process by reducing the number of variables, this led us to remove the following geological catchment descriptors for the rest of the analyses: CzMzi, Czv, i, m, Mzi, Mzm, MzPz, MzPzm, Mzv, pCmv, PzpCm, Pzv and karst. The low correlation of these variables could be due to small areal representation in the geographical domain, poor data quality or small influence of subsurface geology on surface hydrology.~~



~~**Figure 2. Correlation between catchment descriptors and flow signatures using a) Pearson, b) Spearman and c) distance correlations, respectively. Non-significant correlation according to the significance test are indicated with a dash (-).**~~

3.~~2~~1 Catchment classifications~~ and regression analysis~~

An ~~classification~~ automatic clustering based on flow signatures was performed first as explained in section 2.2.~~ and~~ W~~w~~e found that 11 classes were optimal for the database used in this study. The same number of classes was then chosen for the classification based on catchment descriptors. As described in section 2, the third classification (through CART analysis) was based on the classes from the classification of flow signatures. However, ~~the~~ class no.2, which contains only 4 gauges (all situated in Cyprus), was excluded from the CART analysis for consistency~~ purposes~~. As a result, the classification derived from the CART tree only contains 10 classes (numbered 1, 3-11).

~~Concerning~~ During the CART analysis and classification, we found that 20 nodes in the tree was a good compromise to allow all 10 classes to be predicted while minimizing the complexity of the tree (to make the relationships between catchment descriptors and signatures interpretable) and ~~while~~ maximizing the probability for correct classification of catchments (relative error=0.59; minimum probability of correctly classified stations at a node = 0.35). The average percentage of correctly classified gauged catchments in each class was 60% (ranging~~e~~ between 35% and 88% ~~for each~~across leaf node, see Table A in supplementary material). It should be noted that one node (node 3a, see Fig 6) contained more than a third of the catchments (13,~~-~~645 catchments) and only 35% of the gauges in that node were correctly classified. Efforts to further classify catchments in this node through an increase of the complexity of the tree did not result in a good compromise. Indeed, to reach a level of 40% of correctly classified gauges at all nodes, the tree had to be detailed up to more than 400 nodes, making any hydrological interpretation of the splits impossible.

The first two classifications, based on clustering of either the flow signatures or the catchment descriptors alone, resulted in very different spatial patterns of similarity across Europe (Fig. 3, note that there is no correspondence between the numbering of the catchment classes used in maps 3a and 3b). Correspondence between the two classifications is not expected as the two classifications were performed using different sets of data. The third classification – where we predict the flow-based classification from the catchment descriptors – exhibits spatial patterns that are rather similar to the flow signatures-based classification, which is expected since the former is derived from the later through a CART predictive regression tree. Detailed discussion of results in terms of the classification based on flow signatures will, therefore, be focused on results obtained from the CART based classification.

In order to analyze the specific characteristics of the different clusters in terms of catchment descriptors and flow signatures, boxplots representing the distribution of each variable within the clusters were plotted (see sections D~~C~~.1 and D~~C~~.2 of the supplementary material). For the classification based on flow signatures (Fig. 3a), some clear distinctions appear between clusters in terms of mean specific flow and coefficient of variation of daily flow. For example, clusters no. 7 and 10 have the highest mean specific flows while clusters no. 2 and 4 have the highest coefficients of variation. Concerning percentage of agricultural area, some clusters cover a wide range of values (no. 3, 4, 5, 11) while others contain mostly catchments with low percentages of area covered by agriculture (no. 1, 7).

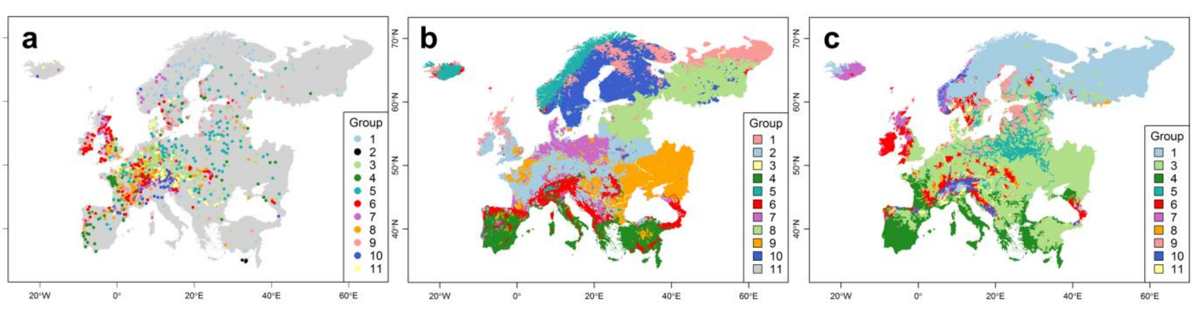


Figure 3. Spatial patterns of catchment classification across Europe based on a) flow signatures at flow gauges, b) catchment descriptors, and c) CART predictive regression tree.

The spatial pattern in Figure 3b (based on catchment descriptors) shows geographically coherent patterns with for example cluster No 6 bringing together mainly mountainous areas, No 4 gathering southern warm catchments , No 7 representing plain regions of the Netherlands, northern Germany, Denmark and Poland. Analysis of the distribution of the different variables in the classes (see boxplots in section D~~C~~.2 of the supplementary material) showed for example that cluster No 5, which is mainly located in Western Norway and Iceland, gathers catchments with low mean temperatures and high mean precipitations with high proportion of open areas without vegetation. In terms of flow signatures, these catchments have high mean and high flows, high runoff ratios and low actual evapotranspiration. Cluster No 11contains 323 catchments but none of them correspond to a stream gauge included in the study. Thus, no observations are available to characterize flow signatures for this class. Observations are limited as well for cluster No 3 as only 13 of the 152 catchments that belong to this class correspond to a flow station. These two classes were thus excluded from further analysis.

Only clustering using catchment descriptors or CART can be applied for the whole domain, i.e. in ungauged catchments. The CART-based catchment classification (Fig. 3c) was chosen for more detailed analysis (in Section 3.3) on similarities in flow generation processes as the clusters were more homogenous. When looking at the classification based on catchment descriptors, the average standard deviations of each catchment descriptor within all clusters was estimated to be 0.71, and the average 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.~~Average of standard deviation within all clusters was estimated to be 0.71 for catchment descriptors and 0.78 for flow signatures using catchment descriptors for classification, while it was 0.76 for catchment descriptors and 0.67 for flow signatures using CART.~~ Hence, the former discriminates classes more in terms of physiography (0.71 vs 0.76 for the CART classification) and the CART classification discriminates classes more in terms of flow signatures (0.67 vs 0.78).

3.2 Using regression analysis to understand controls on individual signatures

As explained in section 2.3, multiple regression models for signature prediction were developed both using the entire domain and within each group of the three classifications and their results were compared. The regression constants are given for each of the 480 calibrated linear models in section E of the supplementary material. This analysis step provides us with two insights: first, what are the dominant controls on individual signatures?; second, how predictable are individual signatures given available catchment/climate descriptors? Figure 4 shows that developing regressions for each of the classes derived leads to better predictive performance than developing an individual regression for each signature using all catchments at once.~~catchment classification did improve the overall prediction of flow signatures using regression models across the whole European domain.~~ This could be expected as using 10 models instead of only one increases the degree of freedom as the number of calibrated parameters increases. This result is consistent with previous findings (e.g. Almeida et al., 2016), which also found that single high performing regressions across large domains are difficult to achieve. On average, classification using catchment descriptors and CART improved the model performance by 14.7% and flow signatures by 33%. The latter yields the best results since this classification is based directly on the discriminating variables (flow signatures). There are few differences in terms of the performance of the models obtained using either the catchment descriptors or CART for classification, the later giving slightly better results for most of the variables (e.g. Q5, High Flow Discharge, high flow frequency variability, variability of reversals, flashiness, runoff ratio), but poorer results for base flow index and low flow frequency. The performance of the regression models for the different flow signatures will be further discussed in part 3.4.

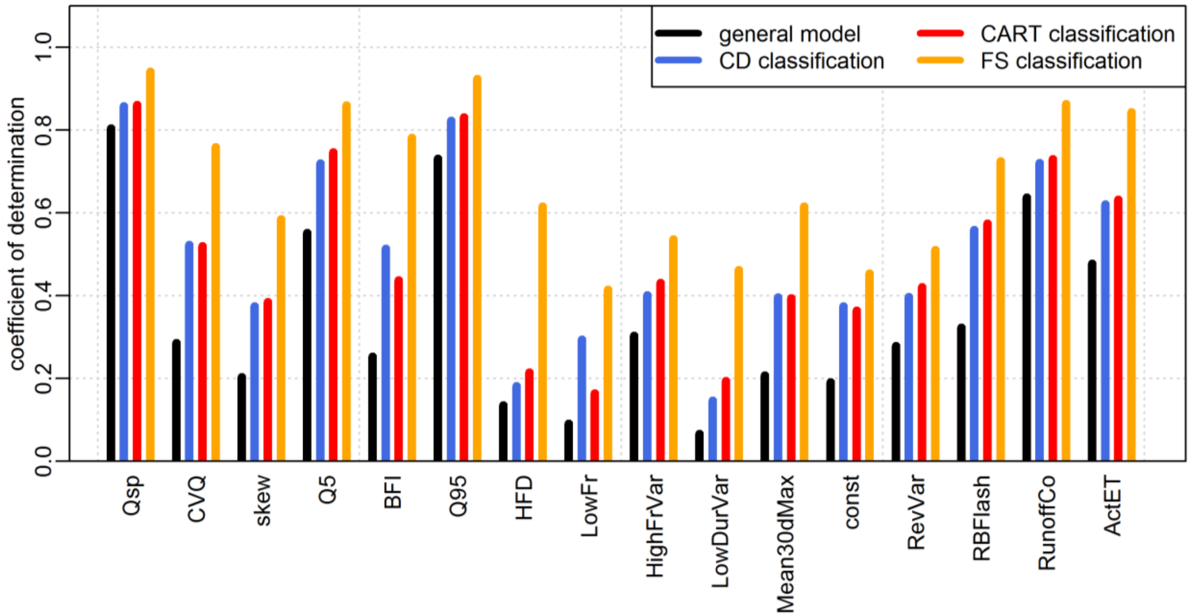


Figure 4. Performance of regression models when calibrated for each flow signature (Table 2) and applied over the whole domain with a general model or one per class, using catchment classification based on catchment descriptors (CD), flow signatures (FS) or regression tree (CART). Performance is evaluated over the whole set of flow gauges together even if different models are used in different classes.

The partial correlation analysis of the regression models shows that there are different controls for the different flow signatures (Fig. 5). The highlighted controls are rather similar ~~for~~ across the different classification methods, i.e. the patterns seen in all three plots are very similar (Fig. 5a-c). This suggests that the identification of controls is robust, while the performances of the different regressions vary. Climatic descriptors play the most important role for most of the flow signatures, especially those related to average and high flows. For the base-flow index, geology is more important and for the flashiness of flow, topography is the main control. Topography also plays an important role in low flows magnitude (Q5), being the main driver for this signature in some of the classes and for the global model. For most of the flow signatures, the second most important descriptor is generally land ~~use~~ cover (mean flow, high flows, runoff coefficient, ET, variability of reversals).

The importance of the different controls varies across the classes (length of the boxplots in Fig. 5) and the main drivers for a given variable can also differ between classes (not shown in the figure). ~~As an example, for evapotranspiration, land use is the main driver in classes 7, 8, 10 and 11 of the CART classification while climate plays the most important role in the other classes. Runoff ratio is mostly explained by soil type in class 11, by land use in classes 6 and 10 and by climate in the other classes (see detailed analysis in Section 3.3). High flow magnitude, also mainly driven by climate in most of the classes, is explained by topography in class 11 and by land use in class 5. It is interesting to note~~ For example, ~~that~~ climate is a strong driver for almost all signatures in class 4 (warm regions in southern Europe) while other drivers play an important role in other parts of Europe, for example in class 7 (topography, land ~~use~~ cover and geology are important), 9 (topography) 10 (topography and land ~~use~~cover). This shows that the drivers behind hydrological responses vary between European regions.

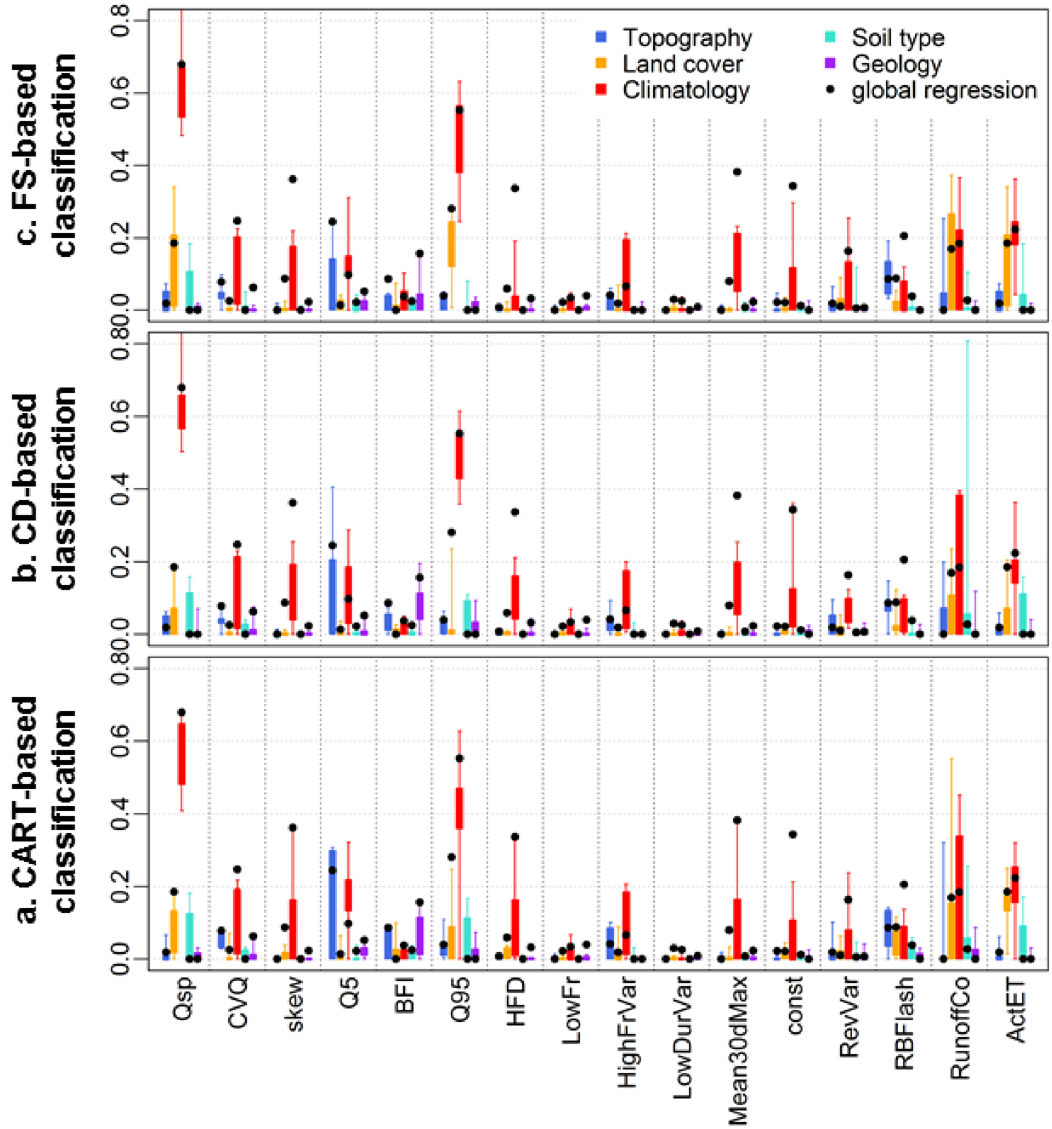


Figure 5. Partial R^2 of different type of descriptors (Table 1) used in the regression models for flow signatures (partial R^2 for the type of descriptors is the sum of partial R^2 of variables from that type used in the regression model). The boxplots show the range of values among the models calibrated in the different classes using the different catchment classification methods: a) flow signatures at flow gauges, b) catchment descriptors, and c) CART predictive regression tree. The black point gives the value for the general model calibrated over the whole domain.

The identified controls for the different flow signatures are generally consistent with the findings of ~~several recent~~previous studies conducted in different parts of the world. For instance, Longobardi and Villani (2008) and Bloomeld et al (2009) found a strong relationship between the base flow index and geology for the Mediterranean area and the Thames basin, respectively. Similarly, Holko et al (2011) found out that the flashiness index is correlated with geology, catchment area and elevation as well as percentages of agricultural and forest landuses for catchments in Austria and Slovakia. For catchments across the US, Yaeger et al (2012) found out that the upper tail of the flow duration curve is controlled more by precipitation intensity while the lower tail is more controlled by catchment landscape properties, such as soils, geology, etc. For the same US dataset, Sawciz et al. (2011) showed that runoff coefficient was dominated by aridity~~ of the climate~~, and that the baseflow index was controlled by soil and geological characteristics. The influence of topography on the magnitude of low flow was also found by Donnelly et al (2016) through a correlation analysis of a set of flow signatures and catchment descriptors across Europe.

3.3 Hydrological interpretation of classes using CART

The regression tree classification (CART) enabled us to~~ a better~~ understand~~ing of~~ the main controls driving the separation into classes (rather than individual signatures), as it predicts the classes of flow signature combinations ~~are predicted~~ from the available catchment descriptors. In the resulting tree (Fig. 6), the main variable separating the different classes is the Aridity Index (AI) with a separating value close to 1. This purely empirical finding is nice, because this value separates the energy-limited catchments (AI<1) from the moisture-limited catchments (AI>1). As expected for classification over such a large domain, we therefore find climate to the first order control. Mean temperature is the second separating variable; followed by variables describing soil types (peat, moraine), land ~~use~~ cover (agriculture, open without vegetation, wetland, forest), topography (area, mean elevation) and climate (precipitation seasonality index, mean precipitation). This indicates the order of importance of catchment descriptors that control flow signatures moves from climate to other descriptors.

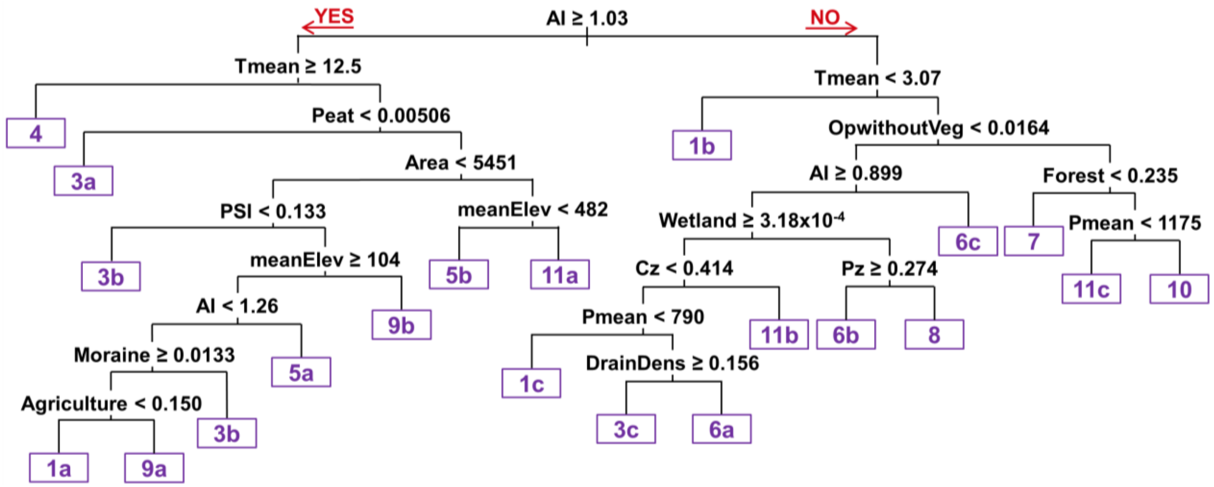


Figure 6: CART tree adjusted on the FS classification and used as a predictive tree for the "CART" classification.

Some of the differences between the hydrographs within catchment classes and across catchment classes can be seen in Fig.7, where we show examples of the observed time-series. We found the following characteristics, which are summarized in Table 3 and further supported by results figures in section C~~B~~ of the supplementary material:

Class 1 has a rather smooth flow, seasonal flow pattern with a very pronounced spring flood peak. The~~i~~se catchments are located in~~s~~ a cold northern part of Europe and some parts of the Alps and Caucasus, characterized by spring snowmelt with some dampening in lakes and wetlands.

Class 3 is a very large~~big~~ (about 1/3 of the catchments) miscellaneous class without any distinct character. As explained in section 3.1~~2~~, efforts to further classify catchments in this class (and more specifically in node 3a) did not succeed.

Class 4 is characterized by very spiky hydrographs with high peaks and low baseflow. The flow regime ~~is~~ exhibits high winter flows and low summer flows. Catchments are located in~~This is~~ the Mediterranean ~~area~~ region characterized by arid climate, flow seasonality and human impacts.

Class 5 shows ~~overall~~ relatively low flows with some influence of snow-melt (spring flood) for some catchments during some years. This is Northern part of central-eastern Europe characterized by low flashiness due to the large amount of water bodies, low topographic slopes and low elevation, which dampen the flow response.

Class 6 has very high peaks especially during winter and high flow periods in general. Overall, flashy flow with a tendency to lower flow during summer and geographically scattered humid areas all over Europe.

Class 7 shows in general high and flashy flows, for most catchments these are higher winter flows ~~but~~ though for some catchments summer high flows instead, due to snow and glaciers melt (this is the class with most glaciers, see~~-~~ Fig. ~~F~~ G of the supplementary material). This class encompasses wet and cold mountainous areas along the coasts in north-western Europe and some humid parts of the Alps.

Class 8 is characterized by peaky flow ~~all over the~~throughout the year with higher peaks in winter. This class consists of smaller headwater catchments in some warm and humid parts of central, south-western and north-eastern Europe.

Class 9 has rather low flow with a snow melt dominated spring flood. Low amplitude but frequent short term variability. These catchments are~~This is~~ mainly in flat lands around the Baltic and Northern Sea further characterized by forests, lakes and wetlands. Some catchments exhibit~~are characterized by~~ similar geological structures (Pz, pCmi, see Fig. K~~J~~ in the supplementary material).

Class 10 shows high flows with very high and frequent peaks, some tendency to ~~spring~~ peaks in spring, but also high flow during winter. Frequent short-term variability is common in these wet, high elevation and steep catchments across mountain ranges of Europe.

Class 11 is characterized by sustained ~~flow with~~ high baseflow and some tendency to spring season peaks in some catchments, but overall low seasonality of flow. These catchments are ~~below~~ close to mountains ~~and~~ or in lower parts of large river basins. We suspect some ~~artifacts~~ outliers in th~~e~~is class~~ification~~ when extrapolating the CART tree to the full European domain, as parts of the catchments in this class were not representative of the majority of~~to the~~ river gauges in the same class (see Fig. ~~L~~ M in the supplementary material, showing that the gauged catchments in node 11b ~~do not have the same characteristics as the rest of the catchments in the same node~~have different characteristics than those in nodes 11a and 11c).

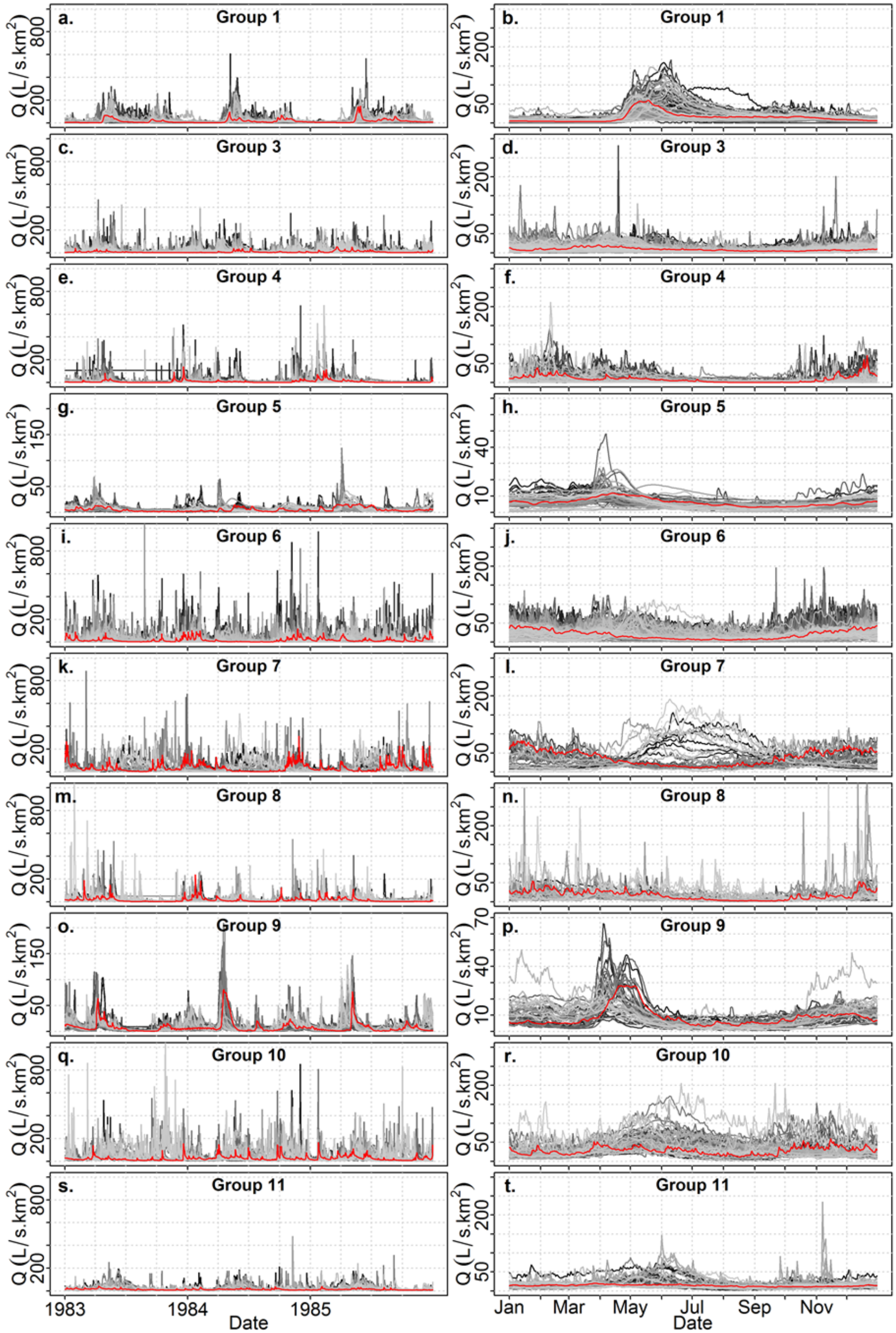


Figure 7. 3-years hydrographs (left) and average annual hydrographs based on > 5 years daily flows (right) at the stream gauges of the CART classification classes. Grey->black: all stream gauges belonging to the class; red: stream gauge where the flow signatures are closest to the class median flow signatures. Note that the scales are different for classes 5 and 9 and that this classification doesn't contain any class no. 2 as explained in part 3.2.

Table 3. Summary of findings when using the CART tree to classify catchments (CART classification shown in Fig. 3c) and extracting the main features for each cluster. Appointed flow signatures (Table 1) and catchment descriptors (Table 2) have median values in the 30% low/high percentile of the distribution over whole domain. Bold indicates median values in the 10% low/high percentile. Supporting figures with boxplots and matrices of flow signatures and catchment descriptors as well as detailed maps of spatial patterns are found in the supplementary material (resp. in sections ~~B~~C.1, ~~B~~C.2 and ~~B~~C.3).

Class	Sample size		Flow signatures (FS)		Catchment descriptors (CD)		Spatial Pattern	Dominant hydrological processes
	No. of catchm.	No. of gauges	FS low	FS high	CD low	CD high	(% of map area)	
1	6878	112	RBFlash, **ActET**	RunoffCo, HighFrVar, Mean30dMax RevVar	Urban, **Agriculture**, Pasture, Medium AI, DrainDens, Pmean, **Tmean**,	**Water**, Forest, **Wetland,** OpwithVeg,, Peat, **NoTexture**, **Moraine**, PSI, pCm, PzpCmm	N and center Scandinavia, W Iceland, Russia. (22.8%)	Snow dominated flow regime with significant snow melt during spring but rather even flow during the rest of the year due to dampening in lakes, wetlands and low actual evapotranspiration. Flow influenced by some hydropower regulation.
2	-	-		-	-	-	-	-
3	14282	536	-	-	-	Agriculture, Moraine, PzpCmm	Large coverage in Western, Central and Eastern Europe. (38.8%)	-
4	5112	91	Qsp, Q RunoffCo, **BFI**	**CVQ,** const, RBFlash, HFD, LowFr, **skew**, **Mean30dMax**	Forest, Pasture	Agriculture, **Irrigated**, Moraine, **Tmean**, PSI, **AI**, PzpCmm	Southern and Eastern part of Europe. (15.0%)	High ET and high human alteration of natural processes. Winter flow is dominated by precipitation while summer flow is limited by evapotranspiration.
5	1765	72	Qsp, CV(Q95, **RBFlash** RunoffCo, skew, HFI Mean30dMax	BFI, HighFrVar, LowDurVar, RevVar	meanElev, stdElev, meanSlope, Pmean	area, Water, Agriculture, Coarse, Peat, Moraine , AI, Cz, PzpCmm	Mainly Poland, Belarus, Lituania, some in S Sweden and Russia (5.6%)	Water flow is dampened by large river channels and water bodies and flat lands. Some influence of snowmelt driven flows. One sub-class (5b) is more controlled by water bodies and the other (5a) by surrounding flood plains.
6	3325	261	HighFrVar, RevVar	Qsp, Q95, RBFlash, RunoffCo	AI	Pasture, Moraine, Pmean, PzpCmm,	Rather scattered distribution: the Brittish Islands, S. Scandinavia, Russia, lower regions of mountainous areas. (6.3%)	Precipitation driven frequent peak flows. One sub-class with rapid response due small area and high slope (6b).

7	678	33	**ActET**, HighFrVar, LowDurVar, RevVar	**Qsp**, Q5, **Q95**, RBFlash, **RunoffCo**	Urban, Forest, **Agriculture,** Medium , DrainDens, Tmean, **AI**	stdElev, meanSlope, Wetland, Peat, **OpwithVeg**, **Pmean**, **OpwithoutVeg**, NoTexture , Shallow, Moraine, PzpCmm	SE Iceland, Scotland, W Norway, some in the Alps. (2.4%)	Low storage (in soil and water bodies) that generates quick response to rainfall. Most catchments have rainfall dominated flow but also some are snow and glaciers melt dominated.
8	670	63	BFI, HighFrVar	CVQ, RBFlash, ActET, skew, LowFr,	area, OpwithVeg, NoTexture	Pasture, Moraine, Pmean, Tmean, Mz, PzpCmm,	Close to class 6 regions in center of France, Carpathians and Russia. (1.6%)	Fast response to precipitation since they are small headwater catchments with low storage capacity.
9	969	52	Q5, RBFlash, ActET	HFD, LowFr, LowDurVar, Mean30dMax, RevVar	**meanElev**,, ,stdElev, **meanSlope,** Pasture, Pmean, Tmean	Water, Forest, Wetland, Peat, NoTexture, Moraine, PzpCmm,	Around Baltic Sea and along the Northern Sea and English Channel coast. (3.2%)	Snow dominated flow regime with significant snow melt during spring. Indications of short-term regulations. Continuous contribution through lateral flow leading to a more sustained flow.
10	762	79	CVQ, skew, HFD, HighFrVar, Mean30dMax, RevVar	**Qsp, Q5** , Q95, RunoffCo, BFI, const	Agriculture, Tmean, **AI**	**meanElev, stdElev** , **meanSlope**, **Pmean,** OpwithVeg, PzpCmm, **OpwithoutVeg**, Shallow, Moraine,	Mountainous regions of W Norway, Pyreneous, Alps, Bosnia, Montenegro, few in Carpathians and Scotland. (2.3%)	Regulated flow for hydropower production during winter but still with some tendency of spring flow.
11	774	67	**CVQ** , RBFlash, **skew** , **HFD**, **Mean30dMax**	Q5, BFI	-	area, meanElev, **stdElev,** meanSlope , Water, Irrigated, OpwithoutVeg, Coarse, Moraine, DrainDens, Cz, pCm, Pzi, PzpCmm	SE France, NE Italy, W Danmark, SE Norway, some in Sweden, large catchments of big rivers like Rhine and Danube. (2.0%)	Flow is governed by continuous supply from upstream storages either from large upstream areas or upstream mountains. (Note: Some catchments (e.g in Denmark) are not representative to the gauges in this class)

The hydrological interpretations of the detected spatial patterns (Table 3) pointed to climate~~ology~~ as the main control of the hydrological ~~processes~~ response in most classes (which is consistent with AI as the main control in Fig. 6). This is highlighted by the notable influence of rainfall-driven river flow in clusters No 6, 7, 8 (Western and Northern Europe) throughout the year, and during winter in 4 (Southern and Eastern Europe). The latter region is most obviously strongly affected by evapotranspiration, while snow-dominated regimes with a spring melt season are characteristic for clusters No 1, 7, 9 and to some extent also No 5 and 10. These clusters are found in the Northern and mountainous parts of Europe.

Regarding landscape influence, dampening effects of river flow response are found in clusters No 1 and 5, due to ~~passage through~~the presence of many waterbodies and vast flatland areas. Continuously strong baseflow ~~supply to river flow~~ is found in clusters No 9 and 11 through lateral flow, large catchment sizes~~contributing area~~ or upstream mountainous areas. On the other hand, clusters No 7, 8 and 6b show fast response and low storage capacity, which could be attributed to their thin soils, high slopes or small catchment sizes.

Impact from hydropower production was found in clusters No 1, 9, 10, which were all snow dominated but showed redistribution of water during the year due to regulation and in some cases influence of short-term regulation. It should be noted that this effect was visible although the gauges from most regulated rivers were already excluded from the study (section 2.1). Human alteration was also assumed to dominate the hydrological ~~processes~~ response in cluster No 4, where the hydrographs did not look natural and ~~irrigation is high~~irrigated areas are large (Southern and Eastern Europe).

Interestingly, ~~S~~some clusters were found to have similar flow signatures but for different reasons. For instance, the damping of peak flows in cluster No 5 could be caused by either the presence of water bodies (5b) or floodplains with a wider river channel (5a).

~~Some clusters were easier to distinguish with many different characteristic signatures or physiography (e.g No 1, 4, 7, 10), while others do not have particular signatures that stand out in terms of their magnitude from the rest (e.g. No 6 and 8). Again, it should be recognized that 1/3 of the catchments (the class no. 3) could not be interpreted hydrologically as they did not show similarities in flow signature values and shared only few catchment descriptors (within 30% percentile of agriculture, moraine and one geological feature, see Table 3).~~

The insight gained from this classification analysis varies across the different parts of the European continent as the classes correspond to different percentage of area (Table 3) and for some classes we learned more than for others. The classification highlighted distinct patterns for most of the classes, some of them showing several outstanding signatures or physiography (e.g No 1, 4, 7, 10), while others had signatures with more average magnitude (e.g. No 6 and 8). On the other hand, about 1/3 of the catchments, covering 39% of the studied area could not be interpreted hydrologically as they did not show similarities in flow signature values and showed only little similarity in catchment descriptors (within 30% percentile of agriculture, moraine and one geological feature, see Table 3). For this part of Europe, we need to search for other or more detailed data of catchment descriptors for understanding the physical controls.

Previous studies have noted that large-scale databases are connected with uncertainties and may sometimes even be disinformative at high resolution (Donnelly et al., 2012; Kauffeldt et al., 2013), which may be a reason for some of the weak statistical relationships and difficulties in catchment classification. European hydrology is also very much affected by human alteration, which is probably not fully covered by the descriptors. Hence, there is still need for further investigations to better understand hydrologic variability across Europe.

3.4 Application of the results: predicting flow signatures over Europe

Figure 8 shows the result of predicted flow signatures using the regression models calibrated within each class of the CART classification. As shown in Figure 4, the performances of these models are diverse: some flow signatures are well modelled (R^2 above 0.8 for mean specific flow and 95th quantile, above 0.7 for 5th quantile, runoff ratio, skewness of daily flow, mean 30-days maximum), but some other models perform very poorly (R^2 below 0.2 for low flow frequency and variability of low flow duration). It is well recognized that modelling low flows can be difficult (e.g. Nicolle et al., 2014; Donnelly et al., 2016; Zhan et al., 2016) and the correlation matrices (see supplementary material) showed that these two flow signatures were poorly correlated to catchment descriptors. This indicates that we currently lack understanding of process and physical controls to predict low flows.

The performances also vary from class to class (not shown here). Models are generally poor (most R^2 below 0.4, a few between 0.4 and 0.6) in class 3, which is a very large and miscellaneous class, but also for classes 6 and 8 which bring together mostly humid catchments, rather scattered over the continent. On the other hand, the best performances are observed in classes 7, 10 and 11, containing a majority of mountainous or close to mountains catchments. Good performances were also observed for at least some of the flow signatures in classes 1, 4 and 5 covering both northern Europe and arid Mediterranean regions.

Figure 8 shows that some negative values appear when applying the calibrated regression models to predict flow signatures. This is explained by the larger range of values of the predicting variables in the whole domain than in the subset of 1366 catchments with flow stations. For example, the predicted values for the 5th quantile of daily flow are negative in 2607 catchments (over the 35215 modelled), most of them belonging to classes 3 and 4. In class 4, the regression for Q5 uses percentage of forest (positive coefficient) and mean temperature (negative coefficient) as the first two predictors. Some negative values appear when the model is applied to catchments with a low percentage of forest and a high mean temperature.

These mitigated results emphasize the ~~experimental~~ empirical nature of these regression models (without process controls) and that they should not be applied outside of the observed ranges of catchment descriptors. However, these regression models help us improving our understanding of European hydrological processes and identifying the dominant controls of the flow signatures in different parts of Europe (see section 3.2). This understanding can be useful when building models that include physical reasoning.

One implication of the identified spatial pattern of flow characteristics and their dominant physiographic controls is that one can delineate regions of particular flow characteristics, for which part of the hydrograph is important. This could be related to the season or component of the hydrograph where the flow is more sensitive to the controlling physiographic attributes. In addition to establishing empirical relationships between the flow signatures and catchment ~~physiographic attributes~~descriptors, like we did in this work, this has a potential application in improving dynamical rainfall runoff models across Europe. Design and results of process-based models should be coherent to empirical findings and when applied on the large-scale, they should thus be evaluated against empirical observations of large-scale spatial patterns, like the ones we provided in this paper.

Furthermore, our results ~~can~~ could be applied to ~~directly~~ improve ~~process-based~~ hydrological models, as patterns of flow signatures are used for defining regions globally for regional model calibration (Beck et al., 2016). We showed that ~~model~~ regression predictions are improved by 15% when establishing ~~models~~ regressions for separate classes of catchment with similar signatures and controls (see section 3.2). This knowledge ~~is highly~~could be valuable when estimating parameter values for continental-scale hydrological models. Currently, there is an emerging need for parameter estimation also in ungauged basins from several modelling communities (Archfield et al., 2015). For instance traditional catchment models have recently been applied on a pan-European scale, e.g. SWAT (Abbaspour et al., 2015) and HYPE (Donnelly et al, 2016). Accordingly, global hydrological models are starting to develop rigorous calibration procedures (e.g. Müller Schmied et al., 2014). The new empirical knowledge we gained in this work could, for instance, be incorporated in the ~~modelled~~ processes description of such models. Processes that control the part of the hydrograph that is sensitive to given physiographic attributes can be parameterized and calibrated separately as functions of the physiographic attributes for the different catchment classes ~~separately~~ (Hundecha et al., 2016). This ~~will~~ could ultimately improve the ~~dynamic models~~ predictive ability of dynamic models in ungauged basins~~, while at the same time enabling prediction in ungauged catchments since the model parameters are functions of the controlling catchment physiographic attributes instead of gauged flow~~.

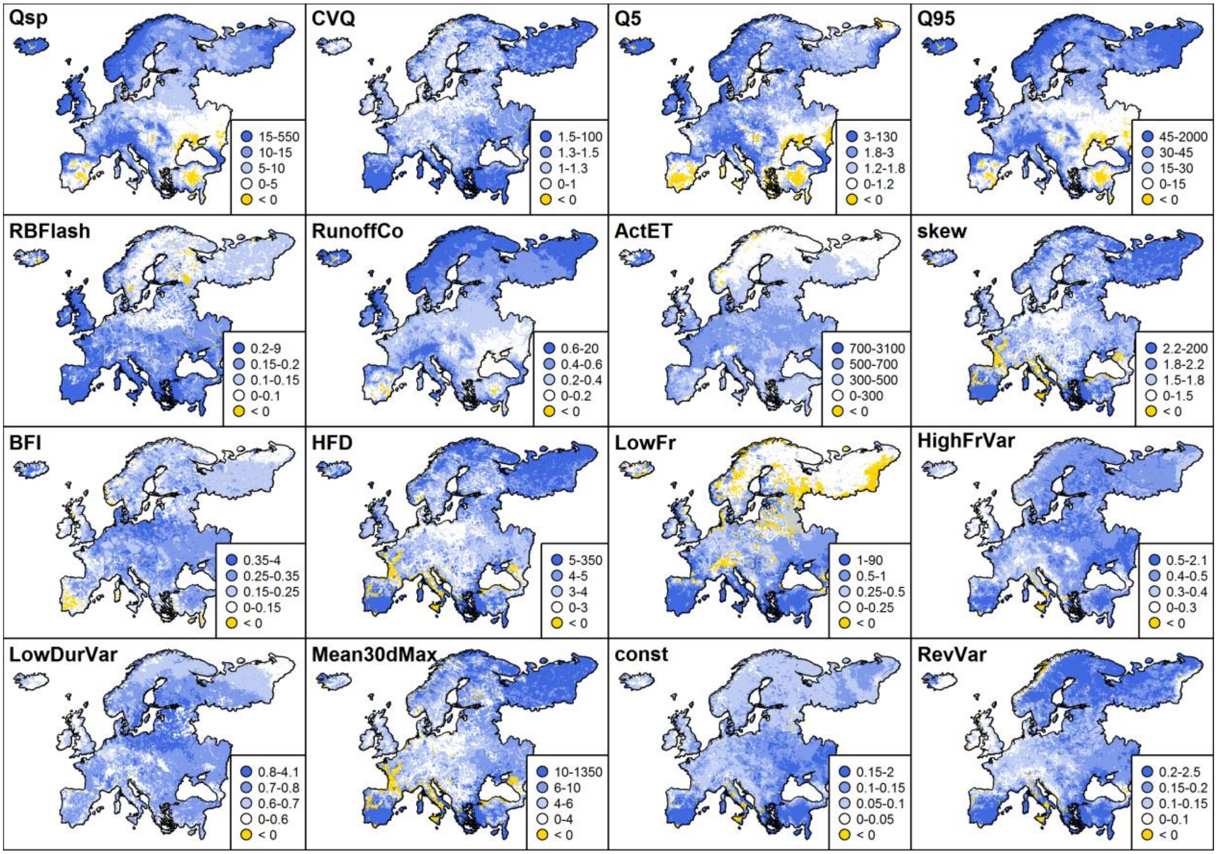


Figure 8. Predicted flow signatures using the regression models calibrated within classes of the CART classification (Fig 3c). Note that the color intervals are adapted to each signature and do not have a constant size for a given signature: for a better readability they are based on the quartiles of the signature distribution. The coefficients of determination of these models are shown in Figure 4.

Conclusions

We set out to better understand hydrological patterns and their controls across the European continent by exploring similarities in flow signatures and physiography. Using open datasets and statistical analysis we found it possible to attribute dominant flow generating processes to specific geographical domains. From the analysis of catchment classification using similarities in 16 flow signatures and 35 catchment descriptors across Europe, we can conclude that:

- Physiography is significantly correlated to flow signatures at this large scale and catchment classification improves predictions of hydrologic variability across Europe (15 to 33% - depending on the classification used - improvement in regression model skills). Different physiographical variables control different flow signatures, though~~;~~ climatic variables play the most important role for most of the flow signatures (12 out of 16). ~~but~~ T~~t~~opography is more important for flashiness and low flow magnitude while geology is the main control for base flow index. All studied flow signatures were significantly correlated with at least one catchment descriptor.

- ~~Different classification methods (e.g. based on physiographic characteristics *vs* flow signatures) can lead to very different patterns, emphasizing the importance of the choice of the methodology according to the use of the classification. However, c~~Classes obtained ~~based~~ by clustering of~~n~~ flow signatures can be predicted ~~using~~ from ~~physiographic characteristics~~catchment descriptors. ~~with o~~On average, 60% of the catchments were correctly classified ~~catchments~~ in each class. In total, Europe~~an catchments can be described through~~ could be divided into ten hydrological classes with both similar flow signatures and physiography. The most important physiographic~~y~~ characteristic for predicting classes is the aridity index ~~(AI)~~, which separates the energy-limited catchments from the moisture-limited catchments. ~~Thereafter follows~~Further explanatory variables ~~describing~~ include soil type~~s~~, land ~~use~~cover, topography and other aspects of the climate/weather. The CART analyses also separated different explanatory variables for the same class of catchments. For example, the damped peak response for one class was explained by the presence of large waterbodies for some catchments, while large flatland areas explained it for others catchments in the same class.

- Interpretation of dominant flow-generating processes and catchment behavior (such as rainfall response, snow-melt, evapotranspiration, dampening, storage capacity, human alterations) could explain the hydrologic variability across Europe to a large extent (61% of the studied domain area). Distinct patterns with characterized flow signatures and processes appeared for some European regions (e.g. Northern Europe, arid Mediterranean regions, mountainous areas), providing a useful information for predictions in ungauged catchments in these areas. ~~However~~On the other hand, flow signatures from 1/3 of the catchments (mainly situated in central Europe) could not be classified or understood~~were not possible to classify or understand~~ based on the ~~physiographical variables~~catchment descriptors ~~used in~~available for this analysis. Th~~e~~is~~e~~ limitations of our large-scale study call~~s~~ for more detailed analysis with additional data in these areas.

- ~~The links we found~~Links between ~~the~~ flow characteristics and physiography could potentially be used in spatial mapping of flow signatures (for instance mean specific flow, 5th and 95th quantiles, runoff ratio, skewness of daily flow, mean 30-days maximum) ~~also~~ for ungauged basins, which might be used in hydrological modeling in the future. ~~Moreover, the findings have potential to constrain and derive parameters for process-based models to increase predictability in dynamic modelling. Using many gauges from catchments with similar dominant processes for flow generation gives more robust parameter values, so therefore, t~~The ten classes of similar catchments may facilitate model parameter estimation in pan-European hydrological models.

- Open data sources enable new forms of comparative science and show large potential for research to generate new knowledge and hydrological insights encompassing variable environmental conditions. However, for Europe there is a lack of homogenous datasets for human impact on flows, such as local water management, abstractions and regulation schemes. There is thus still a need for opening up more public sector data for re-use and, especially, for compiling large-scale databases on the global or continental scales across administrative boarders.

Acknowledgment

This study was performed within the EU FP7-funded project SWITCH-ON (grant agreement No 603587), which explore the potential of Open Data for comparative hydrology and collaborative research, as well as promote Open Science for transparency and reproducibility. All data, scripts and protocols are available in the SWITCH-ON Virtual Water-Science Laboratory at http://www.water-switch-on.eu for review. We would like to thank the Global Runoff Data Center (GRDC) for compiling, maintaining and sharing time-series of river flow as monitored by national institutes. Much of the data used in this study was available from input files of the E-HYPE model; the authors therefore also wish to thank staff at the Hydrological Research unit at SMHI for previous efforts on data compilation, especially we would like to acknowledge the work by Kristina Isberg and Jörgen Rosberg.

References

Abbaspour, K.C., Rouholahnejad, E., -Vaghefi, S. -Srinivasan, R. Yang, H. and Kløve, B.: A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model. Journal of Hydrology (524):733–752 . doi:10.1016/j.jhydrol.2015.03.027, 2015.

Almeida, S., Le Vine, N., McIntyre, N., Wagener, T., and Buytaert, W.: Accounting for dependencies in regionalized signatures for predictions in ungauged catchments. Hydrology and earth system sciences, 20(2), 887-901, 2016.

Archfield, S. A., Archfield, Stacey A., Clark, M., Arheimer, B., Hay, L. E., McMillan, H., Kiang, J.E., Seibert, J., Hakala, K., Bock, A., Wagener, T., Farmer, W.H., Andréassian, V., Attinger, S., Viglione, A., Knight, R., Markstrom, S., Over, T.: Accelerating advances in continental domain hydrologic modeling, Water Resour. Res., 51, 10078–10091, 2015

Arheimer, B., Andersson, L. and Lepistö, A.: Variations of nitrogen concentrations in forest streams - influences of flow, seasonality and catchment characteristics. Journal of Hydrology 179:281-304, 1996.

Arheimer, B. and Lidén, R.: Nitrogen and phosphorus concentrations from agricultural catchments - influence of spatial and temporal variables. Journal of Hydrology 227:140-159, 2000.

Atkinson, E. J., & Therneau, T. M.: An introduction to recursive partitioning using the RPART routines. Rochester: Mayo Foundation, 2000.

Bartholomé, E., & Belward, A. S.: GLC2000: a new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing*, *26*(9), 1959-1977, 2005.

Beck, H. E., A. I. J. M. van Dijk, A. de Roo, D. G. Miralles, T. R. McVicar, J. Schellekens, and L. A. Bruijnzeel (2016), Global-scale regionalization of hydrologic model parameters, Water Resour. Res., 52, 3599–3622, doi:10.1002/2015WR018247.

Berghuijs, WR, RA Woods, M Hrachowitz.: A precipitation shift from snow towards rain leads to a decrease in streamflow. Nat. Clim. Change 4 (7), 583-586, 2014.

Beven, K.: Uniqueness of place and processes representations in hydrological modelling. Hydrology and Earth System Sciences, 4(2), 203-213, 2000.

Bloomeld, J. P., Allen, D. J., and Griths, K. J.: -Examining geological controls on base flow index (BFI) using regression analysis: -an illustration from the Thames Basin, UK. J. Hydrol. 373(1): 164-176, 2009.

Breiman, L., Friedman, J., -Stone, C.J., and Olshen, R.A.: Classification and Regression Trees. The Wadsworth and Brooks-Cole statistics-probability series. Taylor & Francis, 368 pages. ISBN: 0412048418, 9780412048418, 1984.

Brooks, P.D., Chorover, J., Fan, Y., Godsey, S.E., Maxwell, R.M., McNamara, J.P. and Tague, C.: Hydrological partitioning in the critical zone: Recent advances and opportunities for developing transferable understanding of water cycle dynamics. Water Resources Research, DOI:10.1002/2015WR017039, 2015.

Brown AE, Western A, McMahon T, Zhang L.: Impact of forest cover changes on annual streamflow and flow duration curves. Journal of Hydrology. Elsevier Science, Vol. 483, 2013.

Brown, A. E., Zhang, L., McMahon, T. A., Western, A. W., and Vertessy, R. A.: A review of paired catchment studies for determining changes in water yield resulting from alterations in vegetation. J. Hydrol. 310: 28 - 61, 2005.

Bruijnzeel, L.A.: Hydrological functions of tropical forests: not seeing the soil for the trees? Agriculture, Ecosystems, and Environment 104 (1): 185-228, 2004.

Buttle, J.: Mapping first-order controls on streamflow from drainage basins: the T3 template. Hydrological Processes, 20, 3415-3422, 2006.

Buytaert, W., Iñiguez, V., and De Bièvre, B.: The effects of afforestation and cultivation on water yield in the Andean Páramo. Forest Ecology and Management 251 (1-2): 22-30, 2007.

Colwell, R. K.: Predictability, Constancy, and Contingency of Periodic Phenomena, Ecology 55(5):1148–1153, 1974.

CORINE Land Cover dataset: http://www.eea.europa.eu/publications/COR0-landcover

Deshmukh, A., and R. Singh, Physio-climatic controls on vulnerability of watersheds to climate and land use change across the U. S., Water Resour. Res., 52, 8775–8793, 2016.

Donnelly, C, Andersson, J.C.M. and Arheimer, B.: Using flow signatures and catchment similarities to evaluate a multi-basin model (E-HYPE) across Europe. Hydr. Sciences Journal 61(2):255-273, 2016

Donnelly, C., Rosberg, J. and Isberg, K.: A validation of river routing networks for catchment modelling from small to large scales. Hydrology Research, 44(5), 917–925, 2012.

DSMW-fao-Unesco database: http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/faounesco-soil-map-of-the-world/en

Euser, T., Winsemius, H.C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S. and Savenije, H.H.G.: A framework to assess the realism of model structures using hydrological signatures. Hydrology and Earth System Sciences Disc. 9, 12989-13036, 2012.

Falkenmark, M. and Chapman, T.: Comparative hydrology: an ecological approach to land and water resources, UNESCO, Paris, 1989.

Hartigan, J. A., and Wong, M. A. Algorithm AS 136: A K-Means Clustering Algorithm. Journal of the Royal Statistical Society, Series C 28 (1): 100–108, 1979.

Holko, L., Parajka, J., Kostka, Z., Škoda, P., and Blösch, G.: Flashiness of mountain streams in Slovakia and Austria. J. Hydrol. 405: 392–401, 2011.

Hrachowitz, M., Savenije, H.H.G., Blöschl, G., McDonnell, J.J., Sivapalan, M., Pomeroy, J.W., Arheimer, B., Blume, T., Clark, M.P., Ehret, U., Fenicia, F., Freer, J.E., Gelfan, A., Gupta, H.V., Hughes, D.A., Hut, R.W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P.A., Uhlenbrook, S., Wagener, T., Winsemius, H.C., Woods, R.A., Zehe, E. and Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB) - a review. Hydrological Sciences Journal, 58(6):1198-1255, 2013.

Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., Freer, J., Savenije, H.H.G. and Gascuel-Odoux, C.: Process Consistency in Models: the Importance of System Signatures, Expert Knowledge and Process Complexity. Water Resources Research -50:7445-7469, 2014.

Hundecha, Y., and Bárdossy, A. Modeling the effect of land use changes on runoff generation of a river basin through parameter regionalization of a watershed model, J. Hydrol., 292, 281–295, 2004.

Hundecha, Y., Arheimer B., Donnelly, C. and Pechlivanidis, I.: A regional parameter estimation scheme for a pan-European multi-basin model, J. Hydrol.: Regional Studies -6, 90-111, DOI: 10.1016/j.ejrh.2016.04.002, 2016.

Jensen, M.E., Haise, H.R.: Estimating evapotranspiration from solar radiation. Proceedings of the American Society of Civil Engineers, Journal of the Irrigation and Drainage Division 89, 15–41, 1963.

Kauffeldt, A., Halldin, S., Rodhe, A., Xu, C.-Y., and Westerberg, I. K.: Disinformative data in large-scale hydrological modelling, Hydrol. Earth Syst. Sci., 17, 2845-2857, doi:10.5194/hess-17-2845-2013, 2013.

Lehner, B. and Döll, P.: Development and validation of a global database of lakes, reservoirs and wetlands. J. Hydrol. 296, 1–22, 2004.

Lehner, B., Verdin, K., Jarvis, A.: New global hydrography derived from spaceborne elevation data. Eos, Trans., AGU 89(10), 93–94, 2008.

Leibowitz, S.G. et al.: Hydrologic Landscape Characterization for the Pacific Northwest, USA. JAWRA, 52(2), 473-493. Doi:10.1111/1752-1688.12402, 2016.

Ley, R., Casper, M. C., Hellebrand, H., & Merz, R.: Catchment classification by runoff behaviour with self-organizing maps (SOM). Hydrology and Earth System Sciences, 15(9), 2947-2962, 2011.

Longobardi, A. and Villani, P.: Base flow index regionalization analysis in a Mediterranean area and data scarcity context: role of the catchment permeability index. J. Hydrol. 355 (1-4): 63-75, 2008.

McDonnell, J. J. and Woods, R. A.: On the need for catchment classification, J. Hydrol., 299, 2–3, 2004.

Müller Schmied H, Eisner S, Franz D, Wattenbach M, Portmann FT, Flörke M, Döll P: Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration. Hydrol. Earth Syst. Sci. 18:3511-3538, 2014.

Nash, J.E.: A unit hydrograph study with particular reference to British catchments. Proceedings of the Institution of Civil Engineers 1960 17:3, 249-282, 1960.

Nicolle, P., Pushpalatha, R., Perrin, C., François, D., Thiéry, D., Mathevet, T., Le Lay, M., Besson, F., Soubeyroux, J.-M., Viel, C., Regimbeau, F., Andréassian, V., Maugis, P., Augeard, B., and Morice, E.: Benchmarking hydrological models for low-flow simulation and forecasting on French catchments, Hydrol. Earth Syst. Sci., 18, 2829-2857, 2014.

Olden, J. D. and Poff, N. L.: Redundancy and the choice of hydrologic indices for characterizing streamflow regimes, River Res. Applic., vol. 19, no. 2, pp. 101–121, 2003.

Olden, J.D., Kennard, M.J. and B.J. Pusey. A framework for hydrologic classification with a review of methodologies and applications in ecohydrology. Ecohydrology 5: 503-518, 2012.

Panagos P. The European soil database GEO: connexion, 5 (7), pp. 32-33, 2006.

Patil, S., & Stieglitz, M.: Controls on hydrologic similarity: role of nearby gauged catchments for prediction at an ungauged catchment. Hydrology and Earth System Sciences, 16(2), 551-562, 2012.

Pawlewicz, M. J., Steinshouer, D. W. and Gautier, D. L. Map showing geology, oil and gas fields, and geologic provinces of Europe including Turkey, Open File Report 97-470I, 1997.

Pechlivanidis, I. G. and Arheimer, B.: Large-scale hydrological modelling by using modified PUB recommendations: the India-HYPE case, Hydrol. Earth Syst. Sci., 19, 4559-4579, doi:10.5194/hess-19-4559-2015, 2015.

Pollastro, R.M., Karshbaum, A.S., Viger, R.J.: Maps showing geology, oil and gas fields and geologic provinces of the Arabian Peninsula, Open-File Report 97-470B version 2.0, 1999.

Rice, J.S., Emanuel, R.E., Vose, J.M. and Nelson, S.A.C.: Continental U.S. streamflow trends from 1940 to 2009 and their relationships with watershed spatial characteristics. Water Resources Research. DOI:10.1002/2014WR016367, 2015.

Sawicz, K., Kelleher, C., Wagener, T., Sivapalan, M., Troch, P.A. and Carrillo, G.: Understanding hydrologic change through catchment classification. Hydrology and Earth System Sciences, 18, 273–285, 2014.

Sawicz, K.A., Wagener, T., Sivapalan, M., Troch, P.A. and Carrillo, G.: Catchment classification: Empirical analysis of hydrologic similarity based on catchment function. Hydrology and Earth System Sciences, 15, 2895-2911, 2011.

Sefton, C.E.M and Howarth, S.M: Relationships between dynamic response characteristics and physical descriptors of catchments in England and Wales. Relationships between dynamic response characteristics and physical descriptors of catchments in England and Wales. Volume 211, Issues 1–4, November 1998, Pages 1–16. doi:10.1016/S0022-1694(98)00163-2, 1998.

Siebert, S., Döll, P., Hoogeveen, J., Faures, J.-M., Frenken, K., Feick, S.: Development and validation of the global map of irrigation areas. Hydrol. Earth Syst. Sci. 9, 535–547, 2005.

Singh, R., S.A. Archfield, and T. Wagener, 2014, Identifying dominant controls on hydrologic model parameter transfer from gauged to ungauged basins - a comparative hydrology approach, Journal of Hydrology, doi: 10.1016/j.jhydrol.2014.06.030, 2014.

Singh, S.K., McMillan, H., Bardossy, A. and Chebana, F.: Non-parametric catchment clustering using the data depth function. Hydrological Sciences Journal, 2016.

Sivapalan, M.: Pattern, process and function: Elements of a unified theory of hydrology at the catchment scale, in: Encyclopedia of Hydrological Sciences, edited by: Anderson, M., London, John Wiley, 193–219, 2005.

Székely, G. J. and Rizzo, M. L.: Brownian distance covariance, Ann. Appl. Stat., vol. 3(4): 1236–1265, 2009.

Trancoso, R., Larsen, J.R., McAlpine, C., McVicar, T.R. and Phinn, S., Linking the Budyko framework and the Dunne diagram. Journal of Hydrology, 535, 581-597. doi:10.1016/j.jhydrol.2016.02.017, 2016.

USGS (US Geological Survey): Hydro1k Elevation Derivative Database. http://edc.usgs.gov/products/elevation/gtopo30/hydro/index.html

Wagener, T., Sivapalan, M., Troch, P. A., and Woods, R. A.: Catch- ment Classification and Hydrologic Similarity, Geog. Comp., 1/4, 901–931, 2007.

Wagener, T., Sivapalan, M. and McGlynn, B.: Catchment Classification and Catchment Services – Towards a new Paradigm for Catchment Hydrology Driven by Societal Needs. In Anderson, M.G. (ed.) Encyclopedia of hydrological sciences. John Wiley & Sons Ltd., Chichester, UK, 2008.

Ward Jr., J. H.: Hierarchical grouping to optimize an objective function,. J. Am. Stat. Assoc., 58, 236–244, 1963.

Weedon, G. P. et al.: The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. Water Resour. Res. doi:10.1002/2014WR015638, 2014.

Winter, T. C.: The concept of hydrologic landscapes, J. Am. Water Resour. Ass., 37, 335–349, 2001.

World Map Of Carbonate Soil Outcrops: http://web.env.auckland.ac.nz/our_research/karst/

Wriedt, G., Van der Velde, M., Aloe, A., Bouraoui, F.: A European irrigation map for spatially distributed agricultural modelling. Agricultural Water Management 96, 771-789, doi:10.1016/j.agwat.2008.10.012, 2009.

Yaeger, M., Coopersmith, E., Ye, S., –Cheng, L., Viglione, A., –and Sivapalan, M.: Exploring the physical controls of regional patterns of flow duration curves – Part 4: A synthesis of empirical analysis, process modeling and catchment classification. Hydrol. Earth Syst. Sci., 16: 4483–4498. doi:10.5194/hess-16-4483-2012, 2012.

Ye, S., Yaeger, M., Coopersmith, E., Cheng, L. and Sivapalan, M.: Exploring the physical controls of regional patterns of Flow Duration Curves—Part 2. Role of seasonality, the regime curve and associated process controls. Hydrol. Earth Syst. Sci., 16, 4447–4465, 2012.

Zhang, D. Chen, X., Yao, H. and Lin, B. Improved calibration scheme of SWAT by separating wet and dry seasons, Ecological Modelling 10(301):54-61, 2016.