We would like to thank the reviewers for their constructive comments on the manuscript “Clustering CAMELS using hydrological signatures with high spatial predictability”

(comments of the referees are printed in blue, responses of authors are held in black, added text to the manuscript is in italic)

Response to Reviewer #1 (Anonymous)

Jehn et al. classified the CAMELS catchments based on hydrological signatures, and subsequently investigated the link between catchment attributes and the classes. The conclusion of the study is that catchment behavior can mainly be attributed to climate in regions with homogeneous topography, but that this is more difficult in regions with heterogeneous topography. Unfortunately, my perception is that the conclusions of the study are based on a fallacy. The main problem can be found here: “If climate were the main driver, the clusters would be located along a climatic gradient. However, this is only true for the eastern half of the United States (for a climatic map of the United states see (Beck et al.,2018). In this part of the United States, the low relief allows large regions with a uniform climate, that only changes of larger scales.” If looking at the map in Beck et al. (2018), but also Peel et al. (HESS, 2007), or Knoben et al. (WRR, 2018), indeed the eastern part of the US shows large regions with uniform climate. But the maps also all show the large scattering in climates in the west: there is more spatial variation in climate in the western part of the US. This therefore seems no justification to state that climate is less relevant in regions with varying topography - there, the climate is just more variable too. This is also confirmed by the results of the study, where precipitation falling as snow is found as one of the main indicators in the west. “This implies that climate is a good indicator for the discharge characteristics as long as the topography is homogenous.” seems therefore a too strict and incorrect conclusion, that does not necessarily follows from the results / figures.

After receiving those very constructive comments on our first version of the manuscript, we did a mayor reanalysis of our data and provide additional comparisons and tests. This changed our conclusions and we now have rewritten substantial parts of the paper to accommodate this. Section 3.3 was in the focus of the reviewers concerns and is now changed to:

3.3 Exploration of the catchment clusters

The catchment attributes in the CAMELS and similar large scale datasets often show a pattern that resembles climatic zones (Addor et al., 2018; Coopersmith et al., 2012; Yaeger et al., 2012). The picture is less clear for the hydrological catchment clusters presented. This is directly observable in the spatial distribution of the clusters (Figure 3). Usually the 100th meridian is seen as the dividing climatic line in the US, splitting the country in a semi-arid west and a humid east.

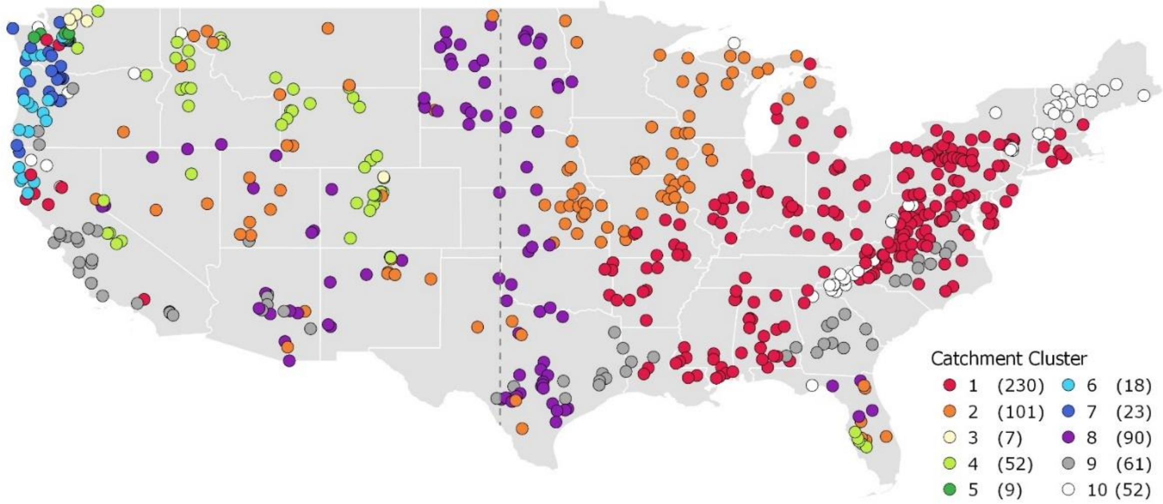

Figure 3: Locations of the clustered CAMELS catchments in the continental US. Dotted line marks the 100th meridian.

This split can also be found in some of the clusters depicted in Figure 3. Cluster 3, 4, 5, 6 and 7 are all located mainly in the West, while Cluster 1 and 10 are in the East. However, the remaining Clusters 2, 8 and 9 have roughly similar amounts of catchments in both regions. The catchments in the eastern half of the United States form large spatial patterns of similar behavior, while the catchments in the west are a lot patchier. The descriptions of the catchment clusters are summarized in Table 2. A further detailed description of the clusters can be found in the appendix, together with figures showing the distribution of hydrological signatures (Figure A2) and catchment attributes (Figure A3) in the clusters. A list of all catchments with index, position and cluster classification is given in the supplementary material.

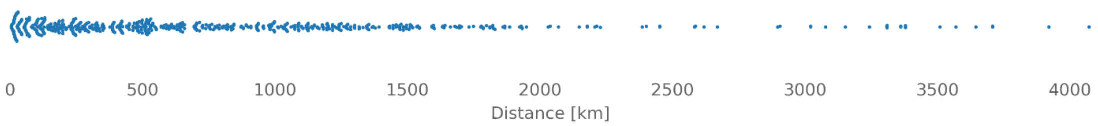

Figure 4: Swarm plot of the real world distances of all catchments to the most hydrologically similar catchment (based on their distance in the PCA space of the hydrological signatures).

In addition, similar catchments can be quite far away from each other (Figure 4). Sometimes, the catchment with the most similar signature was found as far as 4000 km away (almost the entire longitudinal distance of the continental US). This explains why spatial proximity seems to be important in some studies that look into explanations of catchment behavior (Andréassian et al., 2012; Sawicz et al., 2011), but not in others (Trancoso et al., 2017). This also indicates that clustering by using spatial proximity might only work in regions like the eastern US, where the behavior of rivers changes gradually. The finding that the most similar catchment (based on their hydrological signatures) can be far away, also explains the behavior of clusters that contain catchment quite distant from each other (e.g. Cluster 4). Even though the catchments might be far away from each other, the interplay of different catchment attributes and driving factors, including obviously different climates, can lead to similar (equifinal) discharge behavior.

The derived importance of the catchment attributes in the clusters is highly variable and partly differs from the order of importance in the overall dataset (compare Figure 1 and

Figure 5). For Cluster 1 (Southeastern and Central Plains), 6 (Marine West Coast Forests), 8 (Great Plains and Deserts) and 9 (Southern states) aridity has the clearest connection to the clusters. However, this is not the case for the remaining catchment clusters. For Cluster 3 (Northwestern Forested Mountains), 4 (Northwestern Forested Mountains and Florida) and 7 (Western Cordillera) the clearest connection is to the fraction of precipitation falling as snow. However, for Cluster 3, and 4 many other catchment attributes have a weighted R^2, which is almost as high as the one for the fraction of precipitation falling as snow.

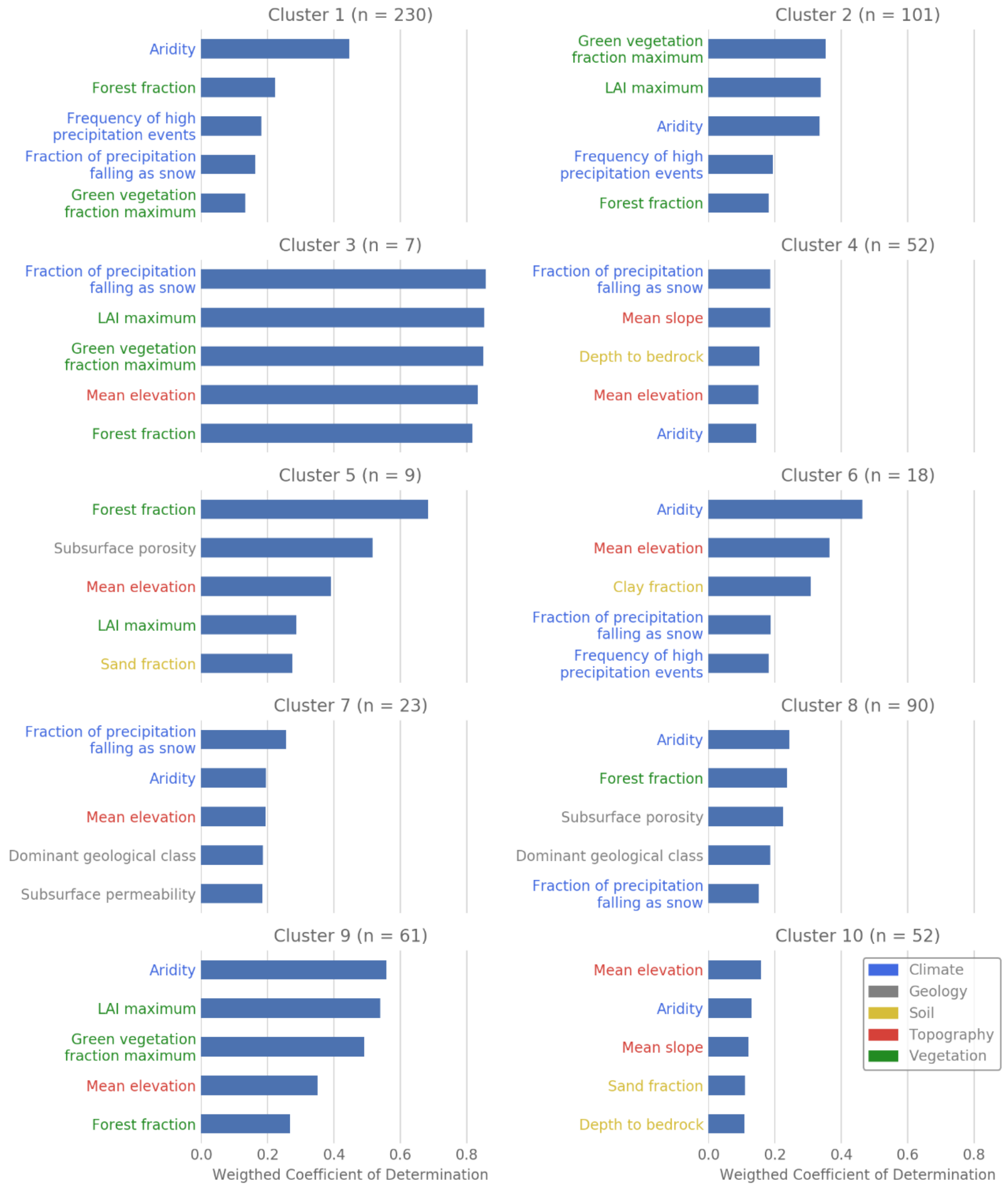

Figure 5: Importance of the catchment attributes evaluated by the quadratic regression. For the catchment clusters. Attributes colored according to their catchment attribute class.

In addition, all catchment attributes have a high weighted R^2 in Cluster 3, while the weighted R^2 is low for all catchment attributes in Cluster 4. For the remaining clusters, it is green vegetation maximum (Cluster 2, Central Plains), forest fraction (Cluster 5, Northern Marine West Coast Forest) and mean elevation (Cluster 10, Appalachian Mountains). Overall, the western clusters (west of the 100th meridian) have the highest weighted R^2 with the:

- *Fraction of precipitation falling as snow (Cluster 3, 4, 7)*
- *Forest fraction (Cluster 5)*
- *Aridity (Cluster 6)*

The eastern clusters (east of the 100th meridian) with the:

- *Aridity (Cluster 1)*
- *Mean elevation (Cluster 10)*

The clusters equally present in west and east with the:

- *Green vegetation fraction maximum (Cluster 2)*
- *Aridity (Cluster 8, 9)*

In the next step, we linked the abovementioned findings to the differences between the correlations of the catchment attributes with each other in the eastern and western parts of the continental US (Figure 6). While aridity is the most important catchment attribute, when looking at all catchments at the same time (Figure 1), this does not hold true for most of the single clusters (Figure 5). Yet, the factors with the highest weighted coefficient of correlation might simply be proxies for aridity. To test this, we scrutinized the correlation between the catchment attributes with each other, separated by East and West (Figure 6). The western US (Figure 6a) and eastern US (Figure 6b) show high differences in the way the catchment attributes correlate with each other (Figure 6c). The main differences are in the mean elevation, the fraction of precipitation falling as snow, and the LAI maximum. For example, in the western US the mean elevation has a high correlation ($r = 0.8$) with the fraction of precipitation falling as snow. In the eastern US however, this correlation is much smaller ($r = 0.4$). This is probably caused by the overall higher elevation in the western US. In addition, in the western US, the fraction of the precipitation falling as snow does not correlate with the aridity ($r = 0.1$), while the forest fraction does ($r = -0.8$). Thus, the forest fraction is linked very directly to the climate in this region. Therefore, aridity (and the highly correlated forest fraction) have the highest weighted R^2 in two out of the five clusters in the western US. Only two clusters are mostly located in the eastern US (Cluster 1 and 10). Here, aridity and the mean elevation have the highest weighted R^2 with the hydrological behavior. The mean elevation has a medium correlation with the aridity. Hence, the hydrological behavior in the eastern US is most highly correlated with aridity, which is not the case for the western US. There, the fraction of precipitation falling as snow is more prevalent. Those results imply that aridity is a good indicator for the discharge characteristics in the eastern US and only mediocre in the West.

Overall, we found that it is relatively easy to link the dominating catchment attributes to the hydrological behavior, in some regions of the US. However, it is more challenging in others. We link this to a less strong climatic signal in those regions. This hints that climate and catchment attributes are more intertwined in those areas and indicates regions where different types of hydrological model structures are needed. Furthermore, it indicates regions where hydrological predictions in ungauged basins (Hrachowitz et al., 2013) *can become very challenging, as the interplay of the available meteorological- and catchment-attributes data cannot sufficiently explain the hydrological characteristics.*

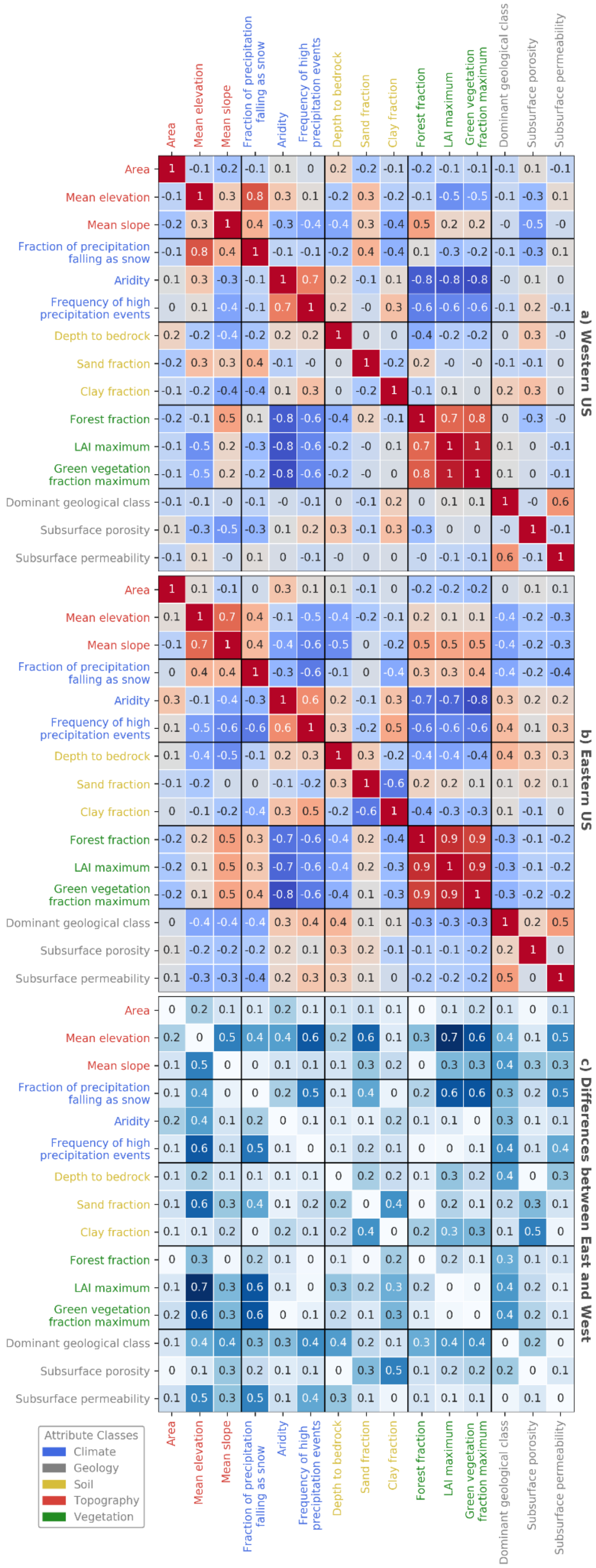

Figure 6: Correlation of all catchment attributes for western (a) and eastern (b) US and absolute differences in correlation between the eastern and western US. Eastern and western is defined by the 100th meridian.

Furthermore, I wonder to what extent ‘homogeneous topography’ can be found as criterion, when looking at the catchment scale, considering that most catchments in CAMELS are rather small.

We do not use this phrasing anymore in the revised version of the manuscript.

Besides my disagreement with the main conclusion, I consider the insights gained from the paper low compared to already available literature, especially considering Addor etal. (2018) and Knoben et al. (WRR, 2018). What did we learn from this study about the relation between attributes and signatures, or catchment clustering, that was unknown before? Especially given that I disagree with the main conclusion. If is it the method applied (PCA combined with clustering), then further elaborate on the methods and better explain everything that is done and how it differs from other studies. This also needs explanation why this method would provide insights that cannot / haven’t been obtained with other methods. I would like to encourage the authors to dive deeper into the material and expand the analysis, and have a critical look at their own conclusions.

To widen the scope of this study and to address the differences in clustering approaches that use hydrological behavior and climate respectively, we added a new section to discuss those topics:

3.5 Comparing catchment clusters based on hydrological behavior and climate

Besides hydrological behavior, climate is often used to sort catchments into similar groups (e.g. Berghuijs et al., 2014; Knoben et al., 2018). Therefore, we are interested if both approaches deliver comparable results. To evaluate this, we contrasted our results to the commonly used Koeppen-Geiger climate classification (Beck et al., 2018) (Figure 7) and recently published approach of Knoben et al. (2018), who sorted climate along three continuous axis of aridity, seasonality and fraction of precipitation falling as snow (Figure 8). The resulting clusters based on climate and hydrology should be the same, if climate is the dominating driver of hydrological behavior in every catchment. Yet, this is not the case for the Koeppen-Geiger classification. In every hydrological cluster are at least two different climates regarding the Koeppen-Geiger classification, ranging up to eight different climatic regions for Cluster 2 and 8 (those even include deserts and very cold regions). Thus, the Koeppen-Geiger classification seems unable to capture the essential drivers of hydrological behavior. A critique also raised in other studies (e.g. Haines et al. (1988); Knoben et al. (2018)).

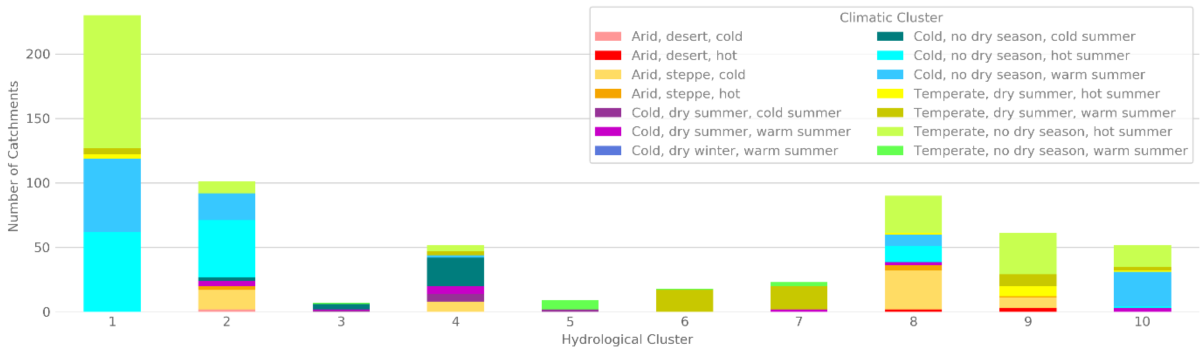

Figure 7: Membership of Koeppen-Geiger clusters (Beck et al. (2018)) in the hydrological clusters.

The picture is less clear concerning the climatic index space of Knoben et al. (2018) (Figure 8a). Due to the continuous nature of the approach of Knoben et al. (2018), there are no clear boundaries as in the Koeppen-Geiger classification. Still, there are some emerging patterns.

For example, according to the approach of Knoben et al. Cluster 1 is mainly defined by a relatively arid climate, with some seasonal variability and little to no snow. This is in line with our analysis of the most influential catchment attributes for this cluster, as we identified aridity as the main driver. Contrastingly, we could not identify a clear dominating catchment attribute, if we look at Cluster 4 (located in the Northwestern Forested Mountains and Florida) (Figure 5). Catchments with this hydrological behavior can be found in the space of the climatic indices of Knoben et al. with very different aridity, seasonality and fraction of the precipitation falling as snow. There seem to be regions were the forcing signal of the climate is transferred more directly to a streamflow response than in others. However, this does not mean that climate is unimportant in those regions. Either the climate forcing signal is changed more through other attributes of the catchment, or the mean values describing the climate do not properly reflect the variability of the climate in the single catchments. This leads to less clear correlation between the climate and the hydrological behavior. Interestingly, when we look at the single hydrological signatures in the climate index space (Figure 8b, A4) we see a very clear connection between the single hydrological signatures and the climate. This direct connection of the signatures used was also found by Addor et al. (2018). Our results and the comparison show that the complex hydrological behavior, captured in a range of hydrological signatures, does not simply follow the climate only, even though the individual signatures do. This is even more remarkable, as the signatures used are linked to climate directly. For example, the signature "mean half flow date" can be seen as a measure of seasonality. Still, all signatures combined seem to capture a dynamic, which is climatic in origin, but is shaped through the attributes of the catchments (like vegetation and soils (Berghuijs et al., 2014)). Therefore, to find truly similar catchments, using climate characteristics only, is probably not sufficient.

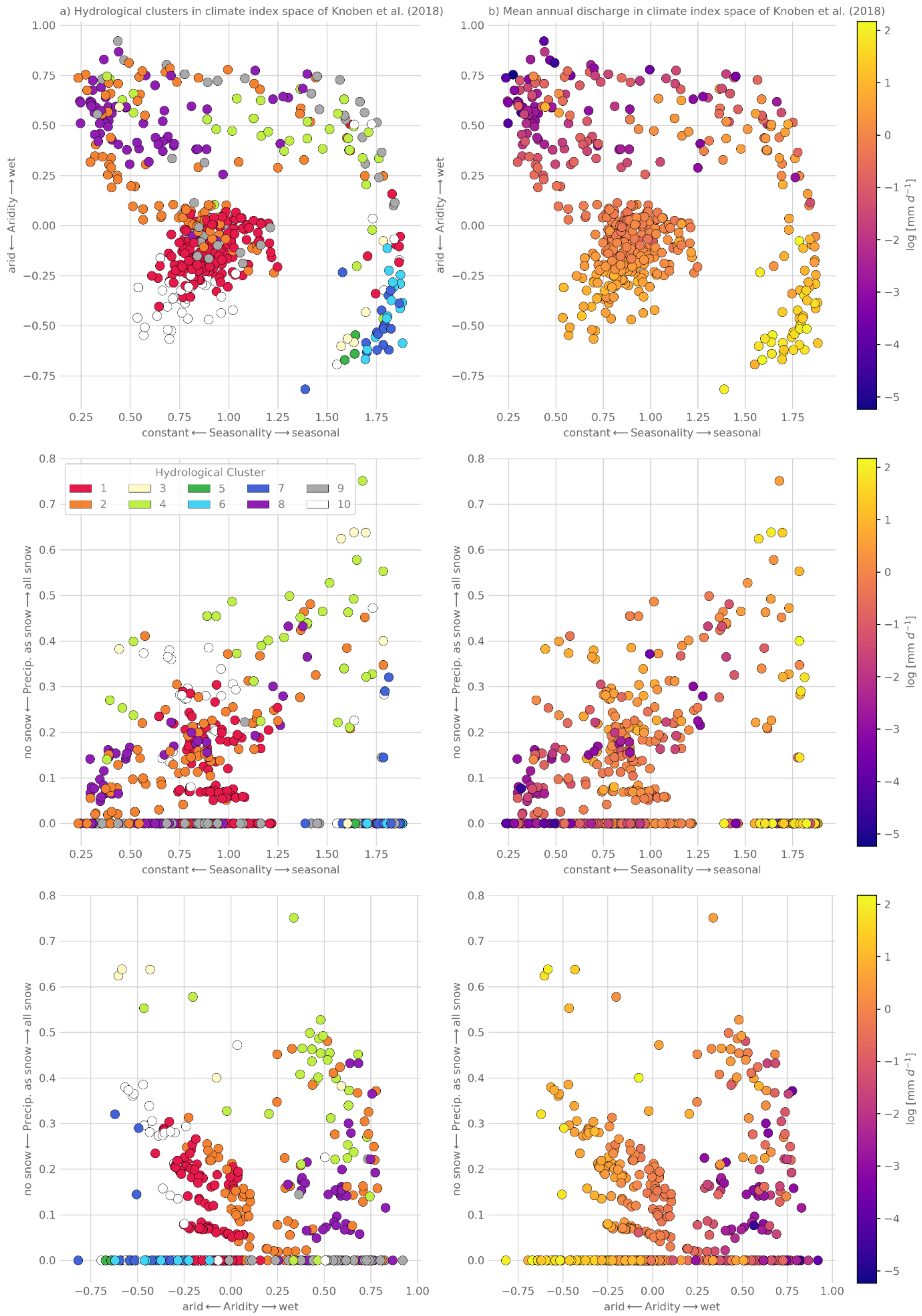

Figure 8: a) Comparison of the hydrological clustering of this study with the climate index space of Knoben et al. (2018). Single dots show the catchments and are colored by their hydrological clusters. b) Mean annual discharge for all catchments in the climate index space of Knoben et al. (2018). Single dots show the catchments and are colored according to the value of the mean annual discharge. The log of the mean annual discharge is used to show the relative differences between the catchments. For a depiction of all hydrological signatures used, see Figure A4.

To reflect the abovementioned changes, we have also rewritten the abstract and the summary and conclusion:

- Abstract

The behavior of every catchment is unique. Still, we seek for ways to classify them as this helps to improve hydrological theories. In this study, we use hydrological signatures that were recently identified as those with highest spatial predictability to clusters 643 catchments from the CAMELS data set. We analyze the connections between the resulting clusters and the catchment attributes and relate this to the co-variability of the catchment attributes. To explore whether the observed differences result from clustering catchments by either climate or hydrological behavior, we compare the hydrological clusters to climatic ones. We find that aridity is more important for hydrological behavior in the eastern US, while it is the amount of snow in the West. In the comparison of climatic and hydrological clusters, we see that the widely used Koeppen-Geiger climate classification is unsuitable to find hydrologically similar catchments. However, in comparison with a novel, hydrologically based continuous climate classifications, some clusters follow the climate classification very directly, whilst others do not. From those results, we conclude that the signal of the climatic forcing can be found more explicitly in the behavior of some catchments than in others. It remains unclear if this is caused by a higher intra-catchment variability of the climate or a higher influence of other catchment attributes, overlaying the climate signal. Our findings suggest that very different sets of catchment attributes and climate can cause very similar hydrological behavior of catchments - a sort of equifinality of the catchment response.

- Summary and conclusion

This study explored the influence of catchment attributes on the discharge characteristics in the CAMELS dataset. We found that over the whole dataset climate (especially aridity) is the most important factor for the discharge characteristics. This changes when we take a closer look at clusters that are derived from specific hydrological signatures. For the clusters in the eastern US, aridity is still the most important catchment attribute. In the western US however, the amount of snow is more important. In addition, in the western catchments the hydrological behavior is less correlated with the remaining catchment attributes. It seems like the clear climatic signal in the east is dampened in the west. This might be caused by a higher influence of other catchment attributes like elevation and vegetation. A similar effect can be found, when we compare how catchment align along hydrological and climatic axes. While some hydrological clusters align along a relatively narrow range of values of the climatic indices, others are found in very contrasting climates. Summarizing, there are differences of how directly the signal of forcing climate can be found again in the hydrological behavior. This explains why catchments often show a surprisingly similar behavior across many different climate and landscape properties (Troch et al., 2013) and why the most hydrologically similar catchment can be hundreds of kilometers away.

The aggregated data used in this study might level out the variability of the catchment attributes in the single catchment, but it also indicates that there is a kind of equifinality in the behavior of catchments. Different sets of intertwined climate forcing and catchment attributes could lead to a very similar overall behavior, not unlike to hydrological models that produce the same discharge with different sets of parameters.

We acknowledge that the results are dependent on the amount and size of the clusters, the catchment attributes considered and the hydrological signatures used. Still, we think that the CAMELS dataset offers an excellent overview of different kinds of catchments in contrasting climatic and topographic regions. Nevertheless, it seems that even a comprehensive dataset like CAMELS, does not allow an easy way to find a conclusive set of clusters for catchments. For future research, it might be a worthwhile pathway to include measures of spatial variability of the climate in the single catchments. This might help to prove, if a less clear *climatic signal is caused by intra-catchment variability of the climate or a larger influence of other catchment attributes.*