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 #2 (Andrew Newman)

General comments: This paper examines the CAMELS catchments and clusters them using hydrologic signatures that have been previously found to have high spatial predictability. Overall this study is somewhat unsatisfying. Little new physical insight is gained in understanding how we determine similarity across catchments. The results do agree with past studies, which is a good test of the previous work. However, what does this specific study bring to us? Previous results discussed here, e.g. Addor et al. (2018) (Fig. 4), and Newman et al. (2015) (Fig. 12) found the same results. Aridity is the primary driver of basin behavior given the catchment scale attributes used, followed by other climate indices (e.g. snow). Finally, one of the primary conclusions drawn from the clustering results needs to be reexamined (specific comment #2).

First of all, we would like to thank the Reviewer herein to provide his very constructive comments. We tried to pick up all points, which lead to a revised version of this manuscript, which provides from our point of view now a clearer insight into the gained understanding and the novelty of this research.

Specific comments:
1) Is 10 clusters necessary? Why does 10 make this study similar to others? Did those studies arbitrarily pick 10 also? I wonder if a similar cluster selection method is better, rather than the same number of clusters. A more detailed justification in the methods section is necessary.

To further elaborate on our choice of 10 clusters, we added a more detailed explanation of our decision in section 2.3:
From those studies, Kuentz et al. (2018) provides the largest set with over 35,000 catchments. They also clustered their catchments in a PCA space of a range of hydrological signatures. To select the number of clusters, they used the elbow method (and two other methods to validate their results) and found that ten or eleven clusters (depending on the method) were most appropriate for their data. Due to the similarity in the clustered data and the larger database of Kuentz et al. (2018), we also used ten clusters.

2) Many of the attributes have high co-variability. For example, elevation and temperature/fraction of snowfall, elevation and mean slope, forest fraction and elevation (in the western US) are likely candidates. Addor et al. (2018) discusses this briefly, but much more could be done here. It would be good to understand this co-variability and modify the discussion accordingly, particularly the conclusions on lines 173-180. Spatial proximity or the attributes defined as climate by the authors are bad predictors in areas with heterogeneous topography precisely because topography and climate are intertwined. That does not mean that climate is a poor predictor of catchment behavior in those same regions.

In light of the reanalysis of our data, we have mostly rewritten section 3.3 and added a discussion of the co-variability of the catchment attributes. This also changes our discussion of the connection between the topography and the climate.

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.

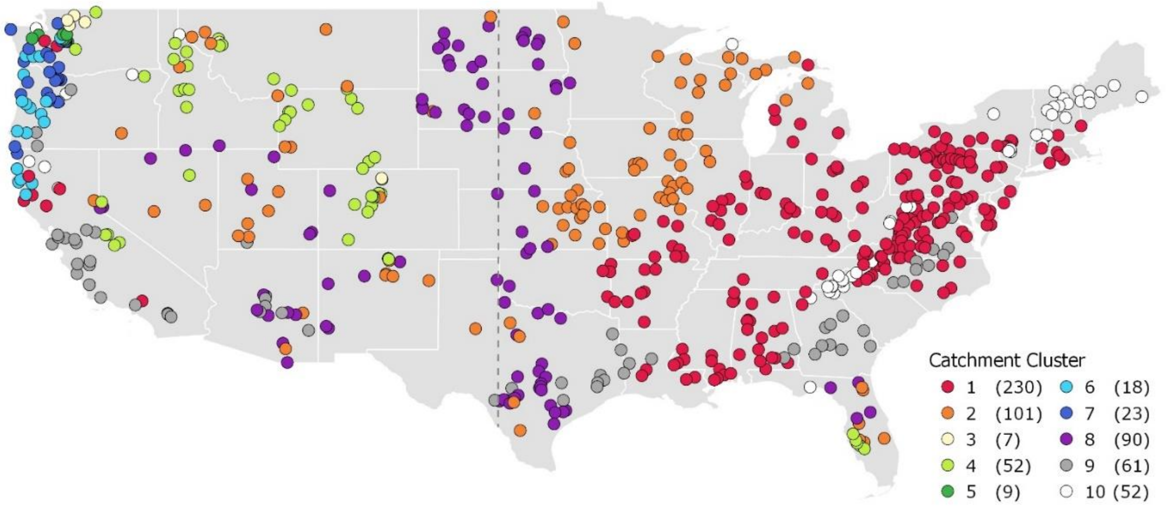

Figure 3: Locations of the clustered CAMELS catchments in the continental US. Dotted line marks the 100th meridian.

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.

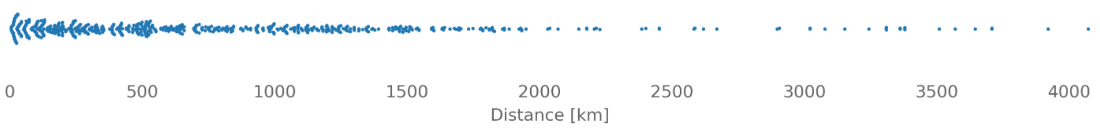

Figure 4: Swarm plot of the real world distances of all catchments to the most hydrologically similar catchment (based on their distance in the PCA space of the hydrological signatures).

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.

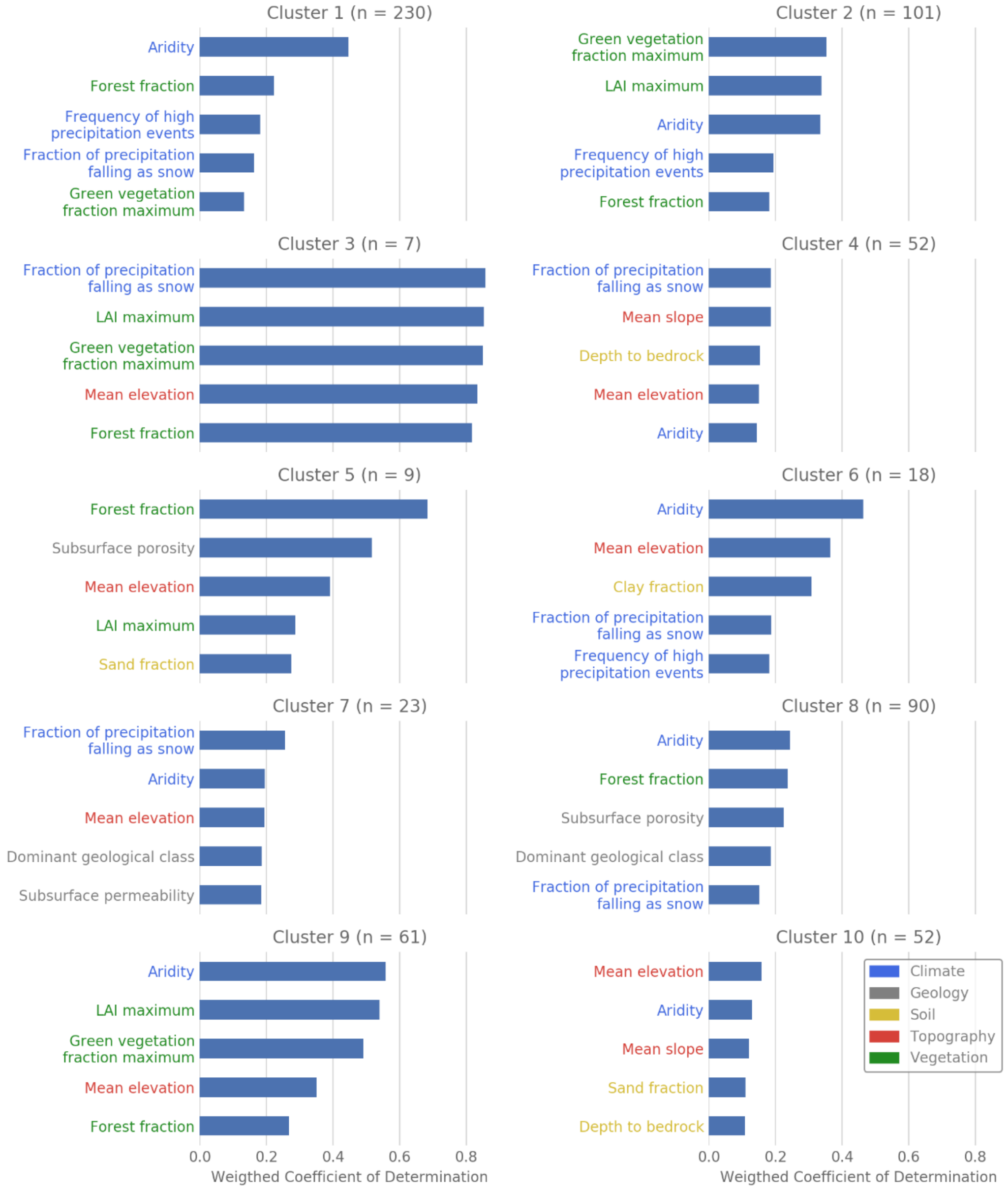

Figure 5: Importance of the catchment attributes evaluated by the quadratic regression. For the catchment clusters. Attributes colored according to their catchment attribute class.

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.*

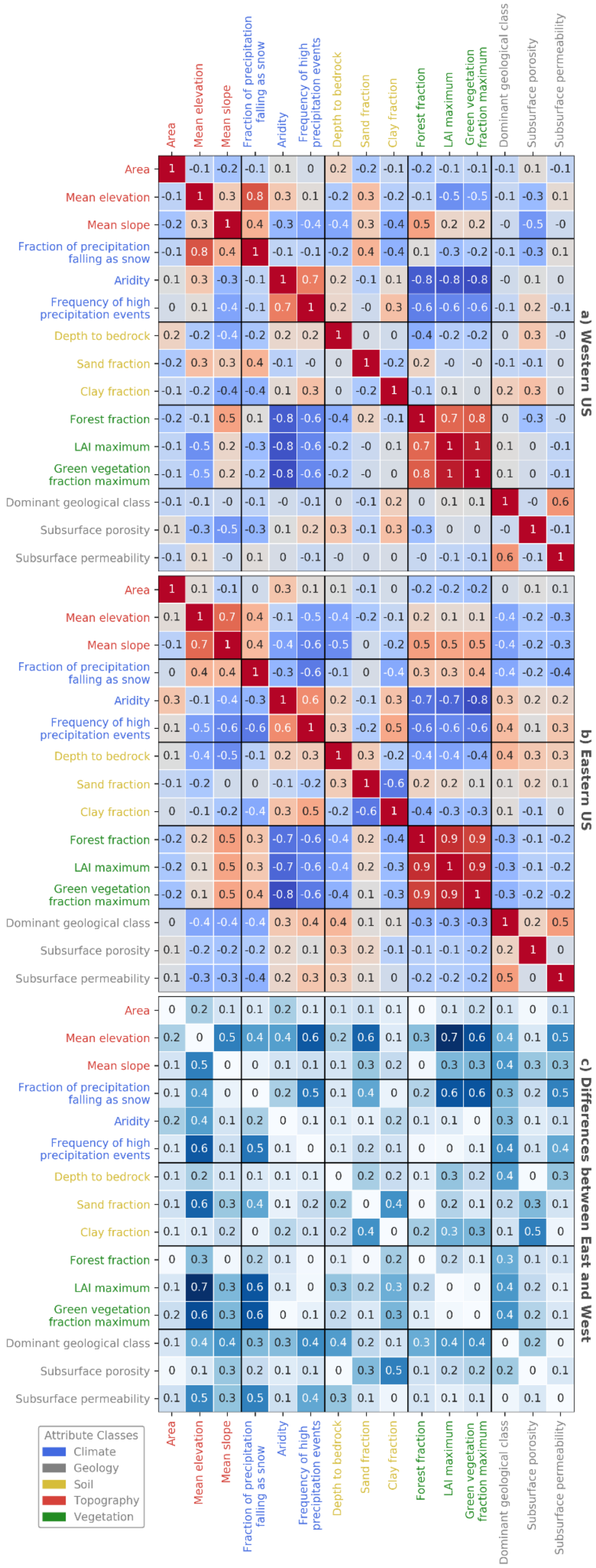

Figure 6: Correlation of all catchment attributes for western (a) and eastern (b) US and absolute differences in correlation between the eastern and western US. Eastern and western is defined by the 100th meridian.

3) Could more explanation be given as to how the clusters contain basins from very different locations (e.g. cluster 4)? There is some discussion in the appendix, which is good, but this cluster highlights limitations in our current clustering methods or application of those methods. How could other hydrologic signatures be used to provide more discriminatory power? Is predictability in space the best metric to determine which signatures to use in a study like this?

This is now discussed in section 3.3:

[…]

This indicates that clustering by using spatial proximity might only work in regions like in 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 far away from each other (e.g. Cluster 4). The catchments might be far away from each other, but the interplay of different catchment attributes and driving factors can lead to similar discharge behavior

[…]

Also, it seems like more discussion on the issues/benefits of using this method (clustering on principle components) using already aggregated data (signatures and catchment averaged attributes) would be useful. This could help the community learn more from these various clustering studies. The authors already provide a summary discussion relating these results to other studies, so I do not feel like this is out of scope.

We added a short discussion of this to section 3.4:
In addition, this study shows that using clusters derived from principal components of hydrological signatures create meaningful groups of catchments with similar attributes (Figure A2, A3). Those clusters also show distinct spatial patterns (Figure 3). Similar results were also found in other studies that used the same method (Kuentz et al., 2017; McManamay et al., 2014), but based them on partly different hydrological signatures. Therefore, the principal components of hydrological signatures can be used as a measure of similarity between catchments. They represent the "essence" of all hydrological signatures used. Our results also show that it is difficult to link those catchment clusters to simple averaged measures of catchment attributes. While some clusters have very clear connections to the attributes, others have no catchment attribute that could easily explain the behavior of the catchments. This hints, that some catchments are easier to explain (in a hydrological sense) than others. Those difficulties might be an artifact of the averaged catchment attributes or be caused by complex catchment reaction, forced by intertwined climate and catchment attributes. Which in turn, might indicate an equifinality of catchment response.

Minor comments:

The sentence starting on line 55 and ending on line 59 is a very long run-on sentence. It is hard to follow and should be reworked. I suggest checking the manuscript for other instances of run-on sentences.

Changed as proposed.

Figures: 1) Please consider increasing the contrast in the cluster colors in Figures 1and 3. Specifically clusters 1-3, and 4-6 are hard to visually separate.

Changed as proposed. We changed all figures with clusters to a more easily distinguishable color scheme. We also changed Figure A2 and A3 from swarm plots to violin plots, to make them easier to interpret.