Manuscript title
A Network Approach for Multiscale Catchment Classification using Traits
Authors
Fabio Ciulla and Charuleka Varadharajan

RC2

COMMENT #1 - This article describes the application of a post-PCA clustering algorithm for classification, in this case for catchments. There is no strong argument that the technique is much better than other methods in this particular application, but the breadth, quality and density of the GAGES-II dataset make it an attractive test bed.

The authors do not apply any effort in showing the improvement their technique makes over others. For example, the justification for their network-based approach is a single paragraph and three numbers. In a more structured analysis, the differences between PCA only, and each of the three post-PCA clustering techniques, would be outlined and their differences tabulated with relevant measures (with an equivalent of Figure 3 for each). There would also be a baseline measure, the PCA or one clustering technique with a minimum number of clusters, and some limited exploration of the number of clusters (or the two free parameters mentioned).

AUTHOR RESPONSE #1: We acknowledge that the manuscript we submitted does not provide sufficient evidence that our proposed method based on networks and cosine similarity performs better than traditional unsupervised clustering algorithms. We thank the reviewer for this suggestion and in response performed a more comprehensive analysis comparing the performance of our method against benchmark hierarchical clustering and k-means approaches.

We reiterate that because of the lack of benchmark datasets, it is not straightforward to evaluate the performance of unsupervised methodologies because of the lack of classified target variables. In the absence of a benchmark dataset, we identified two metrics to evaluate the performance of the different methods.

The first metric, which we refer to as the “cluster similarly metric” has been already introduced in the paper (lines 459-472) and reflects the similarity in traits of the catchment clusters. Clusters can be represented by vectors using the average trait z-scores aggregated among the catchments belonging to each cluster. In this way each catchment cluster can be compared with others by calculating the pairwise cosine similarity. For each catchment cluster, the highest value of the similarity is used as a conservative measure of inter cluster similarity, for the purpose of assessing how far apart the clusters are from each other. The median of all the highest similarities represent how distinct the clusters produced by each algorithm are. A good algorithm should produce distinct clusters, so we aim to *minimize this metric*.

The second metric that we now calculated is the silhouette score (Rousseeuw, 1987), which measures how similar each element (i.e., a catchment) is to the cluster it belongs to with respect to the other clusters. The values of this metric range between -1 and 1, with higher values denoting that an element is well placed in its cluster compared to other clusters. The silhouette values are averaged for all the items in the dataset. A good clustering algorithm would produce *higher values of the silhouette score*.

We use these two metrics to compare our clustering approach based on networks and cosine similarity against the hierarchical clustering (in its common implementation using the ward criterion) and the k-means clustering algorithm. Additionally, we also compare our method against a version where the pairwise similarity between nodes is computed using the euclidean distance instead of the cosine distance. This is done to show the difference produced by the metric choice while keeping the rest of the workflow unmodified. Finally, to show the robustness of our approach (and in response to reviewer 2's comment #3), we extended the comparison by exploring the landscape of the two free parameters in our workflow, namely the number of reduced dimensions after the PCA (k) and the cluster granularity. This last quantity is governed by different parameters according to the clustering method used. Three different values of k are investigated; k=6 corresponding to 50% of retained information after PCA, k=20 (our choice in the study) corresponding to 72% of retained information, and k=90 corresponding to 95% of retained information. For each value of k we generated clusters with the different methods so that the number of clusters covering 95% of the dataset ranges between 20 and 120.

The results of this investigation are shown in Fig. 1, which displays the cluster similarity metric, and Fig. 2, showing the silhouette score. According to these two metrics the performance of our network based clustering (red and green points) is considerably superior to both k-means (yellow points) and hierarchical clustering (blue points) across the different values of k and cluster granularity. This is evident from the consistently low values of the median cluster similarity and higher values of silhouette scores. Also, the network generated using the cosine distance as a similarity metric (red points) performs better than its counterpart that uses the euclidean distance (green points). This confirms that the cosine similarity should be preferred as a metric distance in our investigation, where the dimensionality of the problem can be high and the directionality of the data carries valuable information. If invited to resubmit, we would include this additional analysis as part of the "Methods", "Results" and "Discussion" sections.

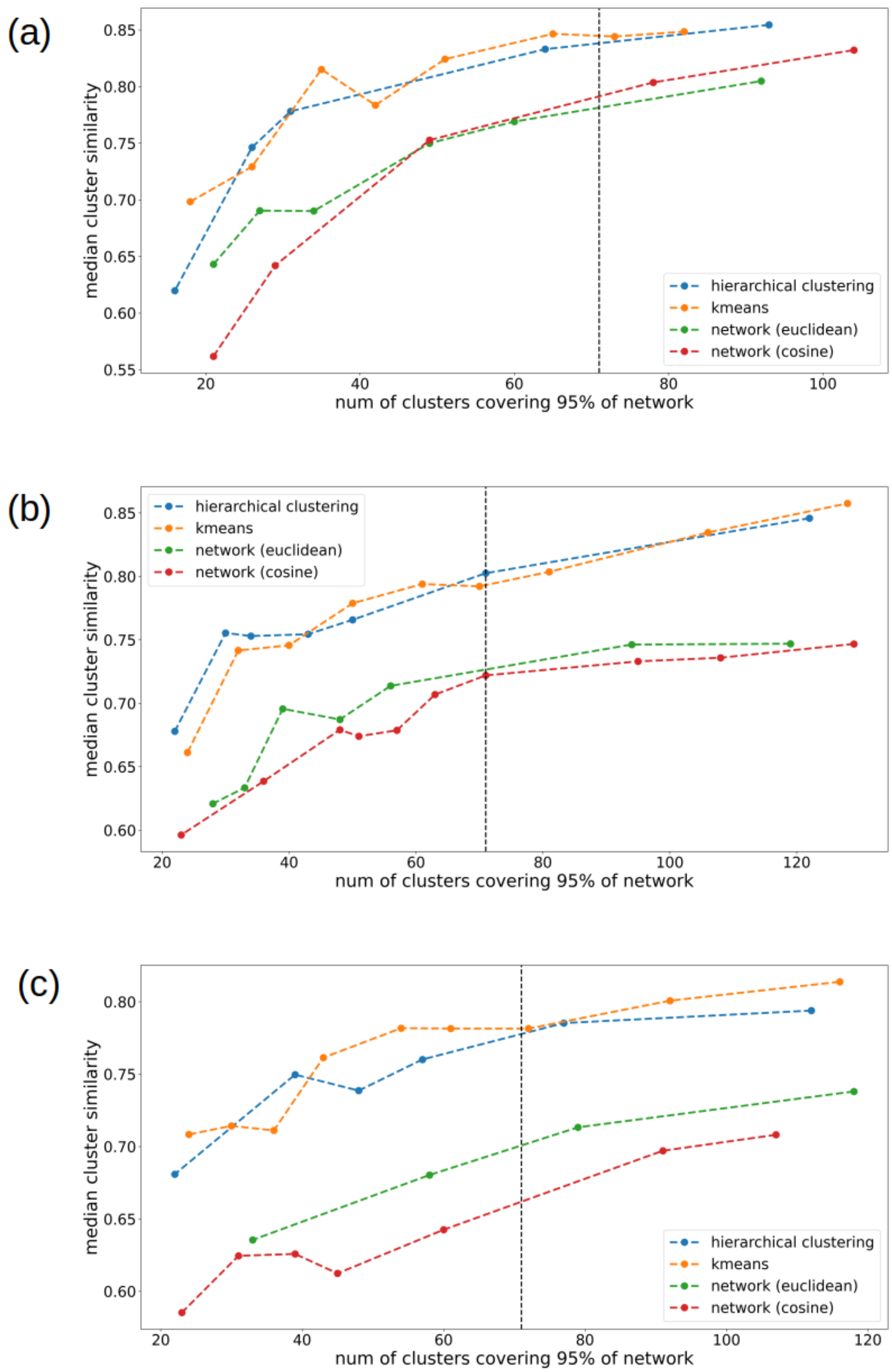

Figure 1. Median cluster similarity values for different clustering methods and similarity measures used in the network analysis. The number of reduced dimensions after PCA is equal to (a) 6, (b) 20 (used in the paper) and (c) 90 corresponding to 50%, 72% and 95% of retained information respectively. The vertical black dashed line in (b) refers

to the cluster granularity used in the paper. The colored dashed lines are shown for visualization of trends. Lower values of the median cluster similarity metric correspond to better clustering performance.

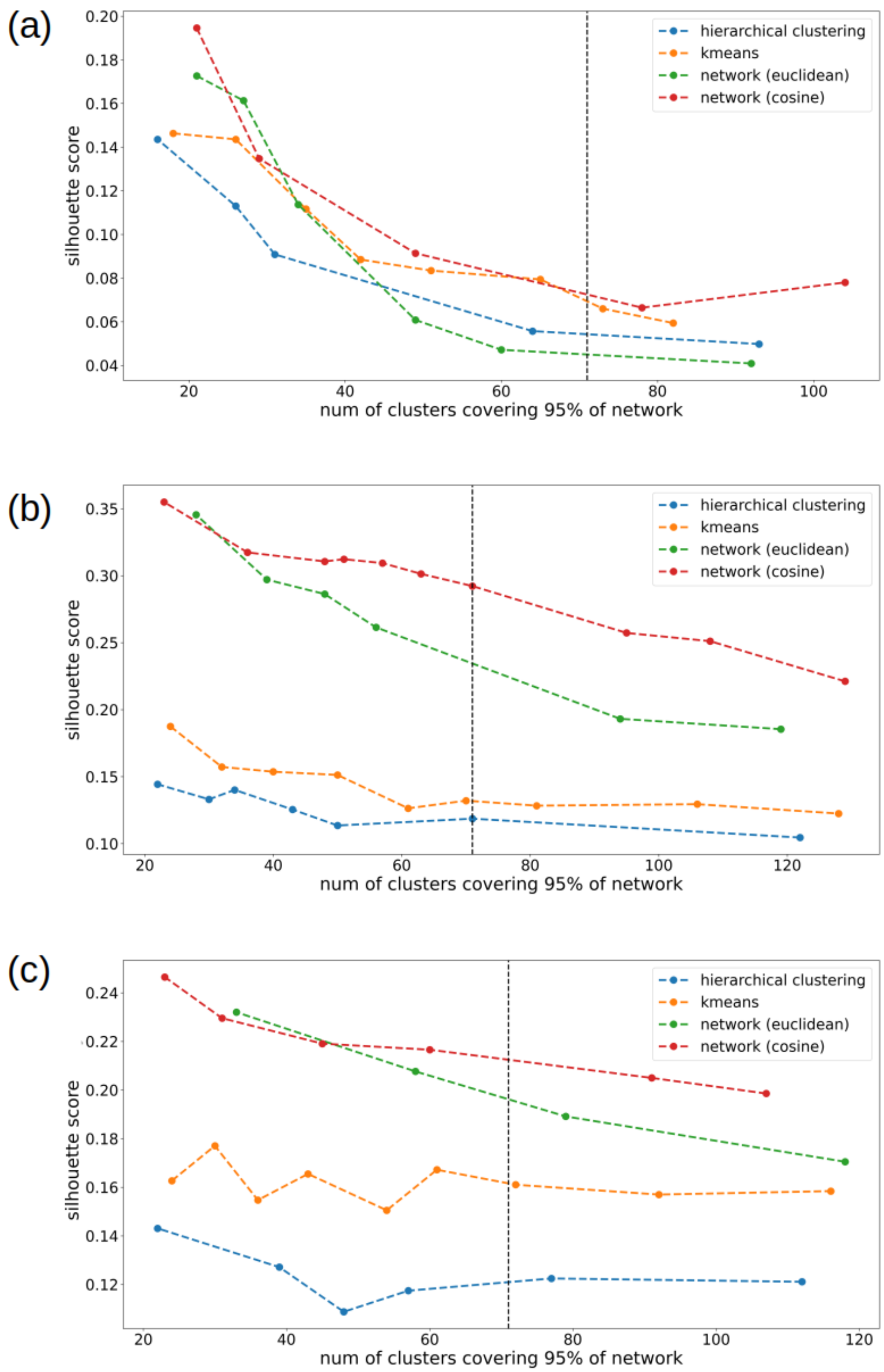

Figure 2. Silhouette scores for different clustering methods and similarity measures used in the network analysis. The number of reduced dimensions after PCA is equal to (a) 6, (b) 20 (used in the paper) and (c) 90 corresponding to 50%, 72% and 95% of retained information respectively. The vertical black dashed line in (b) refers to the cluster granularity used in the paper. The colored dashed lines are shown for visualization of trends. Higher values of the silhouette scores correspond to better clustering performance.

COMMENT #2 - It is not remarkable (line 579) that a classification method using indices and data from a database (of over 300 measures on over 9000 catchments) specifically designed to described gauged catchments for evaluating streamflow would result in a classification that was related to streamflow measures. It will be no surprise to hydrologists that high rainfall, high elevation, forested catchments behave hydrologically differently to flatter, lower rainfall, cropland areas, or that higher rainfall catchments with lots of urban areas get more flooding. What the results might show however is the bidirectionality such that starting from the stream flow indices we get catchment clusters, and that starting from catchment traits we can get groups of catchments with distinct flow behaviour.

AUTHOR RESPONSE #2: Thanks for this comment. Our original write up was intended to highlight that the approach produces intuitive results that are immediately obvious to all readers. Based on this comment, and those from reviewer #1, we have expanded the analysis to include some additional hydrological insights that can be gained with this methodology.

In particular we expanded our analysis of the examples in Section 4.5. In Figure 2, we show the spearman correlation coefficients (ρ) between the streamflow indices (y-axis) and catchment traits (x-axis) aggregated as a median on the trait categories generated in our method that reduces trait redundancy. We find that across the 9067 catchments, mean annual runoff (ma41) is not just positively correlated with traits related to precipitation, but also with the presence of mixed forests ρ=0.45) and to a lesser degree evergreen forests (ρ=0.25). The ma41 index is also negatively correlated with the pastures and grasslands trait category (ρ=-.40). This highlights the role of vegetation in mediating flows, and is somewhat counterintuitive given that in the absence of management, forested catchments with higher evapotranspiration would be expected to have lower flows compared to grasslands. As shown in Fig 13 of our paper, the catchments where there tends to be higher mean annual runoff is cluster 7, which is in the Pacific Northwest basin. The fh6 index indicating the mean number of moderate floods per year (>3 times median flows) is not just positively correlated with precipitation traits, but also with traits related to developed areas (ρ=0.40), croplands (ρ=0.32) and temperature (ρ=0.43). The fh6 index is inversely correlated with elevation (ρ=-0.45), presence of shrublands (ρ=-0.38), evergreen forests (ρ=-0.28), coarse soils & groundwater (ρ=-0.35). These relationships are consistent across other flood indices. For example, the fh7 index showing the propensity for heavy floods (7 times median flows) similarly has a moderate positive correlation with temperature (ρ=0.44) and overland flow (ρ=0.38), and a moderate negative correlation with elevation (ρ=-0.39) and coarse soils/groundwater (ρ=-0.43). This indicates how flooding is affected by the complex relationships between land use, vegetation, soil infiltration capacity and base flows.

The issue of bidirectionality is interesting but beyond the scope of this paper. We are working on building models to predict hydrological indices using trait clusters, and understanding the traits of signature-based classification as part of multiple follow-on studies.

COMMENT #3: What would also have been of interest is the places where the flow indices and clusters do not match well. For example, if there are two areas that are low slope, low elevation cropland that have distinctly different baseflow regime, one may be influenced by groundwater discharge or a factor not yet captured, and this would be useful additional data to know or require to be collected.

AUTHOR RESPONSE #3: Thanks for the suggestion and we agree that it is interesting to investigate subsets of catchments within a cluster where the flow indices do not match well. To investigate this aspect, we performed a new analysis that focuses on anomalies in the hydrologic indices within catchment clusters.

Here, for each streamflow index and for each cluster of catchments we selected catchment subsets that are considered outliers - i.e. where the index is either above the 90th percentile or below the 10th percentile of all indices in the cluster. We compare the z-scores of the traits associated with the catchment subsets relative to the entire catchment cluster to evaluate whether there are differences in traits that would explain the anomalous hydrological behavior. As suggested, we focus on catchments within a cluster that have distinct baseflow regimes, based on a baseflow index (ml17) in Olden and Poff (2003) that represents the 7-day minimum flows divided by mean annual daily flows. The results for anomalous catchments have higher than normal (>90th percentile) baseflow are shown in Figure 3, where the size and color of the bubbles are the relative z-scores of the trait categories.

We focus on a crop-dominated catchment cluster such as the one generally encompassing the Ohio Valley region (cluster 2), displayed in the third row of the bubble plot. This cluster is characterized by relatively low elevation, presence of croplands and fine soil as indicated by the higher z-scores of these trait categories relative to the rest of the CONUS catchments (Figure 4). Using our approach, we can identify the over and under expressed traits of the catchments with anomalously high baseflows in cluster 2 that generally has low elevation croplands. In Fig 3, we find there is a positive association of high baseflows with coarse soils (Z=0.98) and a negative one with fine soils (Z=-0.51), which is not surprising. In addition, there is an association of high baseflows with the “Non-cropland” trait category (third last column with green label, Z=0.85), which aggregates all non-agricultural landuse such as urban areas and forests. This indicates that within the context of a cropland-dominated cluster, the catchments that have relatively lower areas of croplands have higher baseflows. Interestingly, there is also a strong positive association of high baseflows with shrubland (Z=1.12) and a moderate negative association with temperature (Z=-0.65). One possible explanation for these results is that pumping groundwater for agriculture decreases the groundwater input into streams resulting in lower baseflows. This depletion of groundwater discharge into streams does not occur in shrublands or other areas without croplands.

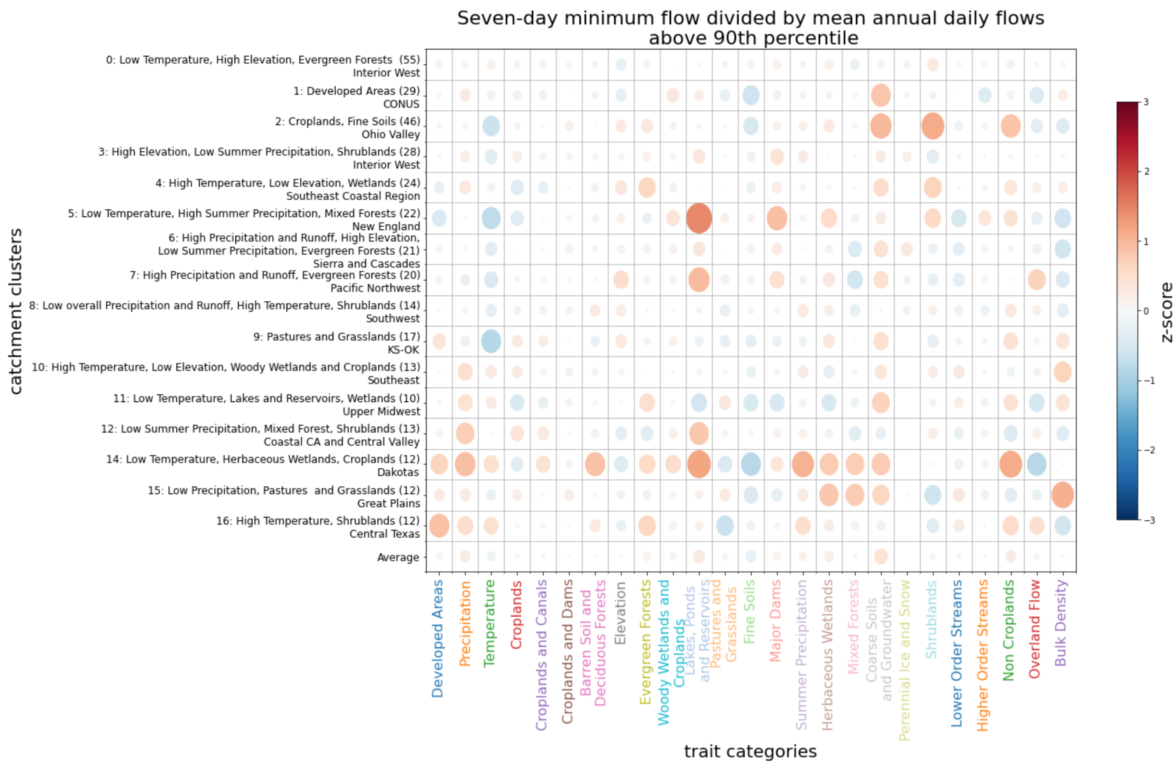

Figure 3. Bubble plot showing the z-scores of catchments where the baseflow is above the 90th percentile relative to the entire catchment cluster. The baseflow index is computed as the seven-day minimum flow divided by mean annual daily flows (averaged across all years). Bubble size is proportional to the absolute value of the z-score. Colors separate positive from negative values as indicated by the colorbar. Catchment clusters are displayed on the vertical axis using an identifier consistent with the one used in the original paper, a name describing their main characteristics, their approximate geographical area (if applicable), and the number of anomalous catchments above the 90th percentile shown in parenthesis in parenthesis. Only clusters with an anomalous set of catchments larger than 10 are included. Traits categories are displayed on the horizontal axis and are sorted in descending order, according to their size in terms of number of nodes in the traits network, and colored consistently with the trait clusters in said network. The last row of each plot refers to the average value of the trait z-scores of the clusters displayed in the plot and provides an idea of how much a trait category is over or under expressed across different clusters with different characteristics.

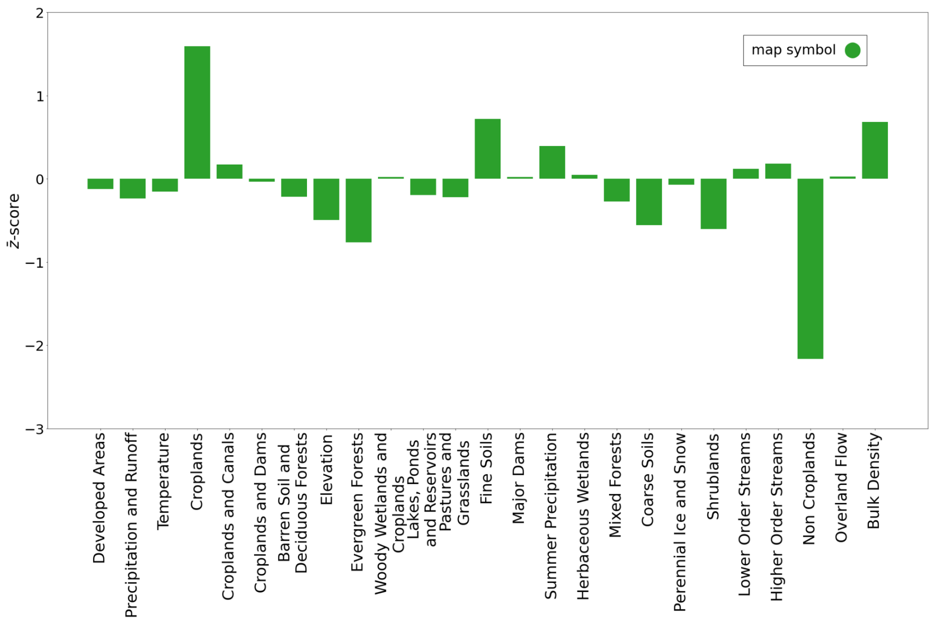

Figure 4. Bar chart of traits z-scores of for the catchment cluster 2, characterized by croplands and fine soils. The catchments in this cluster are generally located in the Ohio Valley region.

Another catchment cluster with a strong agricultural presence is cluster 14, generally located in North and South Dakota, which are characterized by low temperatures, herbaceous wetlands and croplands (Figure 5).

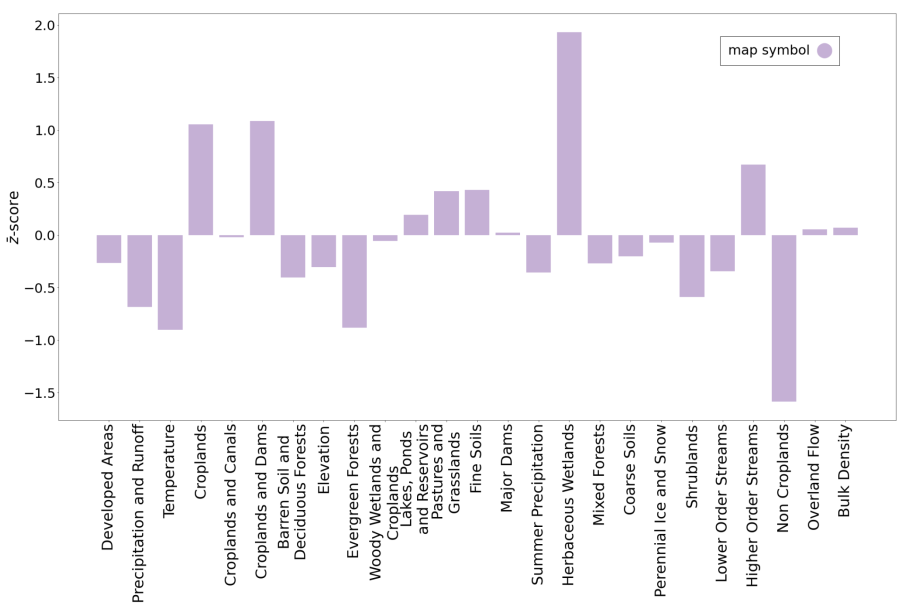

Figure 8. Bar chart of traits z-scores of for the catchment cluster 14 , characterized by low temperatures, croplands and wetlands. The catchments in this cluster are generally located in North and South Dakota.

Similar to cluster 2, there is a positive association of anomalously high baseflows with coarse soils (Z=0.76) and non-croplands (Z=1.10), and a negative association with fine soils (Z=-0.84). However, in comparison to cluster 2, several other factors have a positive association with high baseflows including precipitation/summer precipitation (Z=0.90, Z=1.04 respectively), the presence of lakes, ponds and reservoirs (Z=1.19), herbaceous wetland areas (Z=0.76), evergreen/mixed/deciduous forests (Z=0.58, Z=0.74, Z=0.88 respectively), and developed areas (Z=0.66). There is also a negative association with overland flows (Z=-0.82). This reveals that, in catchment cluster 14, anomalously high baseflows are more likely in the presence of bodies of surface water bodies such as lakes and wetlands that have the potential for increased surface-groundwater exchange. High baseflows also occur in forested areas potentially indicating that the partitioning of precipitation is weighted towards infiltration and recharge over evapotranspiration in these catchments.

Overall, averaged z-scores for all catchments in the CONUS (shown in the last row of Figure 3) indicates there is a moderate positive association of anomalously high base flows with the presence of lakes, ponds and reservoirs (11th column in light blue, Z=0.29), and with coarse soils and groundwater trait categories (18th column in gray, Z=0.40). Conversely, there is a negative link to fine soils (13th column in green, Z=-0.25). This indicates the potential for surface-groundwater exchange in regions where water bodies are present, and not surprisingly the importance of soil texture in mediating baseflow through infiltration and recharge.

COMMENT #4: The citing of references within the text is inconsistent and non-standard, while many of the listed references do not use capital letters where appropriate in journal names or proceedings.

AUTHOR RESPONSE #4: Thanks for pointing that out. We will correct the references if invited to submit a revised manuscript.

REFERENCES IN OUR RESPONSES:
Olden, J.D. and Poff, N.L., 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River research and applications, 19(2), pp.101-121.