A NOVEL TOPOGRAPHY-BASED CLASSIFICATION FOR MOUNTAIN BASINS: UTILITY AND LIMITATIONS OF UNSUPERVISED MACHINE LEARNING

By

Brian Pfaff

A thesis submitted to the Faculty and the Board of Trustees of the Colorado School of Mines in partial fulfillment of the requirements for the degree of Master of Science (Hydrology).

Golden, Colorado

Date _____

Signed: _____
Brian Pfaff

Signed: _____
Dr. Adrienne Marshall
Thesis Advisor

Golden, Colorado

Date _____

Signed: _____
Dr. David Benson
Professor and Director
Hydrologic Science and Engineering Program

ABSTRACT

Unsupervised machine learning algorithms are commonly used data analytic methods with applications spanning many disciplines. In the field of hydrology, K-means and similar clustering methods have been shown to be useful discerning differences in hydrologic signatures between catchments. Specifically, these approaches apply a clustering algorithm to hydrologic data, including hydroclimatic data, land cover data, and topographic data. Topographic attributes alone have also been clustered to establish distinct distributions of mountain ranges to understand species response to climate change and habitat availability, but similar approaches haven't been applied to hydrology to help understand categories of hydrologic response to climate change. Here, I show that calculated distributions of elevation, slope, and aspect for 604 delineated catchments in the USGS GAGES II dataset along with elevation data originating from the USGS National Elevation Dataset can be used to partition catchments into two clusters. Two clusters were determined to be optimal for this dataset according to the elbow method and average silhouette method. The notion that two clusters is optimal suggests there is minimal differences between clusters when simplifying the topographies of catchments to only distributions of elevation, slope, and aspect. Modeling experiments were conducted using HEC-HMS to assess hydrologic response of clustered catchments to the same hydroclimatic forcing parameters. A sensitivity analysis of models run for the basins closest to the center of Cluster 1 and Cluster 2 indicated Cluster 1 Center shows more variation in the timing of peak flow while Cluster 2 Center shows a shorter time to peak flow (TP), time to baseflow from peak flow (TB), and a faster recession. A Kolmogorov-Smirnov (KS) test indicated that the summary statistics TP, TB, and R^2 are statistically significant and do differentiate Cluster 1 Center and Cluster 2 Center. However, when modeling randomly selected basins in each cluster, the lack of visual difference between hydrographs and a KS test indicated no statistical

significance for TP, TB, and slope of recession rate (b). These results together suggest no hydrological difference between Cluster 1 and Cluster 2. The linear regression fit of recession rate against discharge performed well for Cluster 1 Center as indicated by a high R^2 (0.86) but performed poorly for Cluster 2 Center and for the randomly selected basins. These results demonstrate pitfalls of the K-means clustering algorithm that must be considered. This study is somewhat unique in that I use an external validation method to determine the extent to which the K-means clustering produced meaningful categories. This external validation method indicated no differences between clusters and that the K-means algorithm did not create meaningful classifications. I propose a validation method is needed such as hydrologic modeling of the clustered catchments to assess the integrity of the K-means results and to ensure classifications are distinct and meaningful. Without proper validation techniques, the results of K-means algorithms and broader unsupervised clustering methods should not be assumed to be correctly partitioned.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Adrienne Marshall, for the opportunity to undertake the research described in this thesis, for her ongoing academic/professional guidance, and for challenging me to develop new technical skills and expand my scientific horizons. All of her assistance is invaluable as I move onto my professional endeavors. I would also like to thank my committee members, Dr. Kamini Singha and Dr. Danica Roth for their unique perspectives and additional guidance that culminated in this research. The various professors and faculty members who have who have shared their knowledge and professional experiences in and outside of classes. Classmates and other students at Colorado School of Mines, both current and previous, who have provided feedback on research, clarified topics in the classroom, and have been alongside me to navigate graduate school. Lastly, my fiancé, family, and friends who have provided unwavering support and motivation throughout my entire time at Mines.

CHAPTER 1

GENERAL INTRODUCTION

1.1 Previous Clustering Studies in Hydrology

Often referred to as the water towers of the world, mountain watersheds provide up to 90% of water supply worldwide in the form of streamflow or groundwater that replenish reservoirs and aquifers ultimately for human consumption (Viviroli, 2007). It is widely known that an increasing temperature advances snowmelt timing, with peak melt occurring weeks or months earlier than in previous years (Stewart, 2005). But even in the absence of temperature increase, topographic characteristics such as slope, elevation, curvature, and drainage area affect how water is distributed and moves around a catchment (Prancevic, 2019). Hydrologic characteristics of watersheds, specifically stream flow fluxes, are often of concern for water resource planning over large and small spatial scales. While studying regional hydrological trends is useful for large scale water resource planning, understanding these trends at the basin scale is useful to identify the topographic mechanisms that control streamflow generation. Basin scale trends can be studied using classification systems, which allows for a heuristic understanding of basin behavior in response to climate change, particularly for ungauged basins.

Catchments vary in scale, topography, and hydrologic/climatic composition and reflect these differences in unique hydrological signatures. Despite this apparent heterogeneity, numerous previous studies have tried to differentiate and classify catchments according to these unique hydrological signatures in a uniform manner such that basin characteristics are discernable. Geospatial analyses have benefitted from the increased availability of hydrologic, climatic, and high-resolution topographic datasets for catchments around the world at various

scales over the last several decades. This increased data availability has made clustering algorithms a commonly used analysis tool to assess similarities or differences between groups of catchments. The use cases for clustering have been demonstrated in various aspect of hydrology. For example, several studies have used hydrologic response to differentiate catchments to develop new generalization theories and to provide a method for assessing hydrological implications of climate and land use (Sawics, 2011, 2014; Olden 2011; Köplin 2012). Regional classifications according to stream flow time series (Corduas, 2011) in addition to a classification according to how climate/catchment attributes affect the spatial distributions of flood generating processes (Stein, 2021) both prove to be valuable clustering approaches that further our understanding of hydrological theories. For example, Corduas (2011), proposes the temporal dynamics of streamflow time series can be compared through estimating linear model parameters and regionally classified. Specifically, the Mahalanobis distance which is an index that measures the dissimilarity of linear model parameters for independent time series was clustered to establish regional classifications of watersheds according these differing Mahalanobis distances (Corduas, 2011). Stein (2021) used a statistics-based approach to first determine types of flood processes from the CAMELS dataset including snow/rain flood, snowmelt flood, excess rain flood, short rain flood, and long rain flood. This was included along with physical catchment parameters including catchment area, catchment shape, catchment size, soil type, geology, and vegetation in a machine learning classification methodology which found that climatic attributes, namely fraction of snow, aridity, and precipitation seasonality have the strongest influence on flood generating processes within space (Stein, 2021). Sawicz (2011) determined hydrological similarities between 280 catchments in the Eastern US based on how the catchments respond to precipitation inputs using publicly available precipitation, temperature, and streamflow data.

Using a Bayesian clustering model where number of clusters was identified through the clustering process, they identified 9 homogenous clusters for which they compared the climate and landscape similarities across the study area. Sawicz (2014) expanded upon their previous work to classify 314 catchments across the US based on hydro-meteorological observations using the previously used Bayesian clustering model in addition to a regression tree. By establishing a baseline period of 1948 to 1958 from which to derive the clusters, they were able to track how catchments change clusters through time. They concluded that the catchment changing of clusters is most strongly controlled by changes in the water balance of the catchments. These studies make use of clustering methods by ultimately clustering hydrologic signatures such as streamflow time series or response to precipitation input.

Although topography has not been outright ignored in existing clustering models, it does take a back seat to climatic forcings and hydrological responses and is often not directly included in the clustering analysis itself. It is commonly found that climatic variables are often the most important driving factors for hydrological behavior (Sawicz, 2011; Jehn, 2020) and as such, are the first variable considered in a clustering analysis. But it is also noted that this does not always hold true for all regions and all scales (Jehn, 2020). There have been some applications that directly include topography in the clustering methods, namely studies involving the creation of Hydrologic Response Units (HRUs). HRUs are geographic units representing a unique combination of slope, land cover, and soil type such that each HRU has acceptably similar hydrological behavior regarding climate inputs and water runoff. The purpose of HRU creation is to simplify modeling and reduce computational complexity relative to a fully distributed hydrologic model. It is noted that topography defines an HRU at the most basic level of the delineated watershed (Devito, 2005). Aytac (2020) performed a K-means clustering analysis on

HRUs for 33 catchments in Turkey to create a classification schema of the catchments which can be transferred and applied to ungauged basins. Aytac (2020) also notes that the results of their K-means clustering approach cannot be validated as there is inherent inexactness of K-means clustering. Topography - specifically slope, is required for creation of these HRUs, and is generally estimated from a Digital Elevation Model (DEM).

Topographic differences have been examined over large spatial scales including classifying mountain ranges into distinct distributions (Elsen & Tingley, 2015). Elsen & Tingley (2015) determined that surface area does not always decrease monotonically as elevation increases. By analyzing the available surface area at various elevations for 182 of the world's mountain ranges, they found that 68% have topographic distributions in which the elevational availability of surface area does not decrease monotonically. They argued that 68% of mountain ranges exhibit 4 distinct topographic patterns from which they conclude that the notion that species might encounter less surface area as they migrate upslope into higher elevations might be incorrect based on their four distinct topographic classifications. Although Elsen & Tingley (2015) classified mountain ranges into 4 categories, this classification was developed heuristically and there was no validation of their results. They did use an ANOVA to demonstrate that their identified mountain classifications did not differ within their respective geographic areas however this did not address the similarity/dissimilarity between classifications.

Typically studies in which classifications are developed using unsupervised machine learning algorithms do not use associated validation measures to assess if the classifications truly represent distinct groupings. This is not commonly done because it is an inherent part of unsupervised clustering algorithms that the results are inexact (Aytac, 2020). However, this logic

assumes that the classifications cannot be validated as the data needed for validation was already considered as part of the classification process. To validate if the classifications are indeed distinct would require the use of external data or application of a 3rd party validation technique which uses the original data in a differing manner from the clustering process. While statistical tests are often used to assess the dissimilarity of classifications on a statistical basis, there is often no further validation efforts that determine if classifications are truly different.

1.2 Utility of Classifying Topographic Distributions

The methodological approach of Elsen & Tingley (2015) answers an ecological question (related to changes in habitat size) using their clustered topographic distributions. However, in the realm of hydrology, a comparable methodology has not been used to try to discern major topographical differences between catchments by developing a unified classification scheme in addition to assessing hydrologic performance of this classified scheme under idealized conditions. By purposefully neglecting climatic conditions, we can study how topography alone impacts runoff and streamflow generation. This is important because topography could be a basis for understanding how basins might respond to changing climate given hydrological processes that are topographically driven. Recent work has shown the importance of topography as slope controls the rate of infiltration; shallow slopes promote groundwater infiltration whereas steep slopes promote streamflow (Carroll, 2019). In addition, watersheds that are more dependent on groundwater recharge could be buffered against climatic extremes (Taylor, 2012). For example, basins with more low-slope terrain in lower elevations might tend to have more groundwater infiltration and therefore be less sensitive to climate change. Alternatively, basins with steeper slopes could have more rapid runoff and therefore be more sensitive to climate change. It is

established that areal extent of winter wet-day temperatures conducive to snowfall over the western United States is likely to decrease by 30% by the mid-21st century (Klos, 2014). In addition, areas in the rain-snow transition zone, defined as the wintertime precipitation regimes where precipitation dominantly falls as either rain or snow are changing as wintertime precipitation increasingly falls as rain and not as snow (Knowles, 2006). Areas previously suited for snow could change in the future, meaning that streamflow runoff characteristics and timing could change as well. By understanding how topography alone dictates the movement of water within catchments, catchment response to changing hydroclimatic conditions can be quantified and tracked. Furthermore, within a given climatic region, the topography of neighboring basins could help predict differing catchment responses to climate change. By better understanding the coupling of topography and streamflow runoff/generation using a classification framework, basin response to changing hydroclimatic forcings can be measured used to understand how certain regions might be dominantly affected compared to others.

1.3 Research Objectives and Hypotheses

Here, I propose that clustering catchments according to the topographic attributes: elevation, slope, and aspect will result is an idealized distribution of catchments that are related only in their topographic signatures. Experimental hydrologic models can be then conducted on the clustered basins in which basin topographies are subject to the same vegetation, land use, and forcing parameters and the resultant hydrographs can be studied. This study consists of the following research objectives:

1. Can the topography of basins be simplified to only elevation, slope, and aspect and be used to classify basins in a meaningful manner?

2. Given these classifications, how does the shape and elevation profile of catchments affect the timing and magnitude of runoff and streamflow generation? Do runoff patterns indicate that the clusters represent distinct categories of basin topography?

Hypothetical precipitation events are commonly incorporated in sensitivity analyses and vulnerability studies such as how hydrologic variables change in response to climate change scenarios (Bao, 2011). By developing conceptual models where a hypothetical precipitation distribution is dispersed uniformly across all basins, these experiments should show how various topographies elicit a hydrologic response under the same forcing conditions. From the defined research objectives, I hypothesize the basin clustering will yield a meaningful set of discrete topographic basin types where classifications will significantly impact runoff characteristics in modeled experiments. For example, I expect magnitude and timing of flow to ultimately be a function of basin relief and length of stream channels with the basin.

CHAPTER 2

K-MEANS CLUSTERING OF TOPOGRAPHIC VARIABLES AND HYDROLOGIC MODELING OF SELECT CLUSTERED BASINS

2.1 Methods

The overall research methodology consists of two phases: Phase 1 will be to study and classify topographic distributions for delineated catchments across the western United States using unsupervised machine learning. Phase 2 will be to develop hydrologic models of selected clustered basins identified in Phase 1 and ultimately see if the clustered topographies yield significantly different hydrograph signatures.

2.2 Data Acquisition

The aim of Phase 1 is to develop a data-driven topographic watershed classification system across the western US. Specifically, Phase 1 consists of data processing and analysis of a publicly available dataset using the R programming language. GAGES II (Falcone, 2011) was chosen as it is a publicly available geospatial dataset created and maintained by the USGS. It consists of stream gauges for the entire United States that have 20 + years of continuous records in addition to the delineated drainage basins that drain to those gauges. Furthermore, GAGES II is a robust dataset with various use cases including correlating snow melt with annual streamflow (Hammond, 2018). However, for the scope of this project, only basin geometry is of interest. In addition, only catchments west of the Continental Divide were included in this analysis. GAGES II differentiates between reference gages and non-reference gages where reference gages are the watersheds that are the least disturbed based on their indicators. For this project, I am only

concerned with reference gages; this results in 604 watersheds. Figure 2.1 shows the locations of the watersheds selected for analysis.

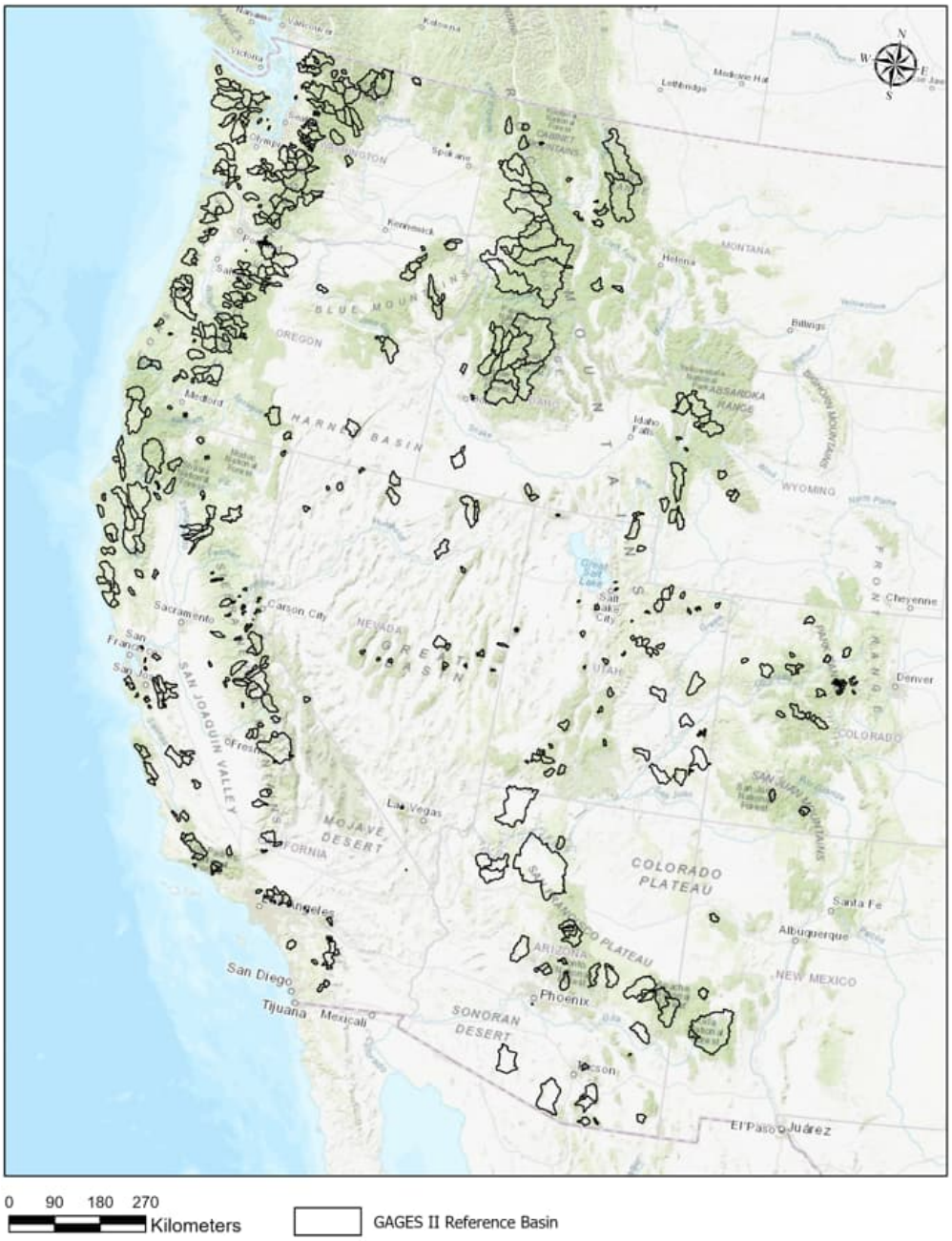

Figure 2.1: Map showing locations of GAGES-II reference basins

Once the GAGES II dataset was filtered to only include reference gages and the watersheds west of the Continental Divide, I acquired a DEM raster for the entire study area.

Data was obtained using the Elevatr package in R, which downloads elevation data from Amazon Web Services Mapzen Terrain Tiles (Hollister, 2021). The Elevatr package requires the user to specify a zoom level which corresponds to the resolution of the elevation raster. Different zoom levels yield elevation data from different sources. For the contiguous United States, data is sourced from the National Elevation Dataset (NED) created and maintained by the USGS. DEMs were downloaded at a 57 m resolution. After the DEM was downloaded for the entire study area, the DEM was then clipped and masked to each individual catchment, resulting in 604 individual DEMs. In addition, slope and aspect rasters were also created for each catchment during this step. The data for each respective topographic criteria was then merged into 3 master datasets each grouped by Station ID which corresponds to the unique Station ID of each reference gage in the original dataset. The Elevation dataset was then normalized to a 0-1 scale and aspects were defined within 45° of each cardinal direction. After reclassifying both the elevation and aspect datasets, the fraction of pixels was determined in 0.1- and 10-unit increment thresholds for elevation and slope respectively and in thresholds of North, South, East, and West for aspect. It was necessary to determine fraction of pixels in each threshold for each topographic criterion because we ultimately provide these fractions to the K-means algorithm to assign the basin to a particular cluster. For elevation and slope, the thresholds were split into 10 equal-width bins for simplicity and to uniformly discretize the input datasets so basins can be compared according to the topographic criterions. These thresholds were calculated within each watershed, so the elevation, slope, and aspect values represented how these topographic criterions were distributed within the watershed, rather than separating watersheds into relatively high- or low-elevation watersheds.

Across all pixels included in the dataset, the distributions of elevation are roughly normally distributed, with the most elevation values around 1800 m, and the highest elevation values around 3500 m (Figure 2.2 A). Figure 2.2 B shows the total dataset contains the highest density of slope values around 0 degrees, meaning most of the basins contain low angle topography. Figure 2.2 C shows that all aspect values are approximately equally distributed throughout the dataset.

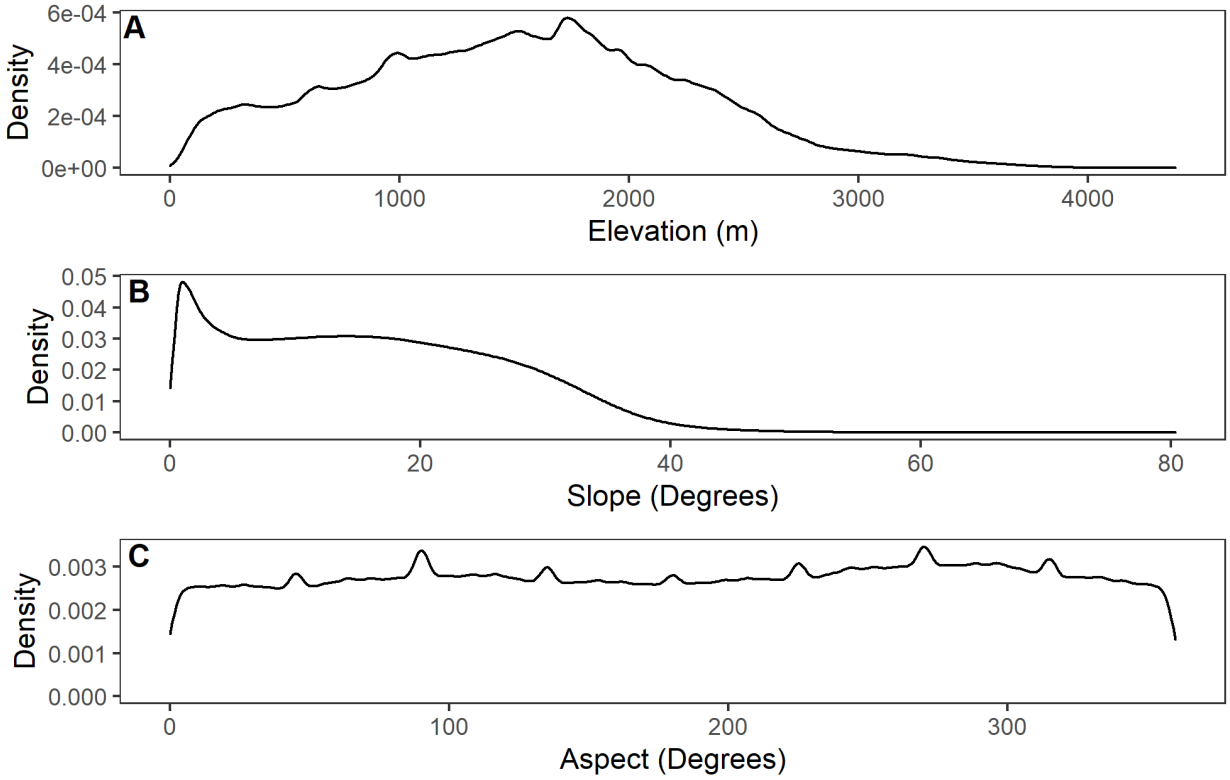

Figure 2.1: A) Density plot of elevation values for all pixels in 604 basin dataset. B) Density plot of calculated slope values for all pixels in 604 basin dataset. C) Density plot for calculated aspect values for all pixels in 604 basin dataset.

2.3 Phase 1: K-Means Clustering

Following the data gathering/analysis process, the datasets consisting of the fractions of pixels in each threshold for elevation, slope, and aspect were combined into one and then scaled to have a mean of 0 and standard deviation of 1 before being fed into a K-means clustering

algorithm. K-means clustering is an unsupervised machine learning technique to partition *n* observations into *k* clusters so that the intra-cluster variation is minimized (Macqueen, 1967). There are several K-means algorithms available; the algorithm used for this project, as well as the most used, is the Hartigan-Wong algorithm (Hartigan, Wong 1979). This algorithm defines the total within cluster variation as the sum of squared Euclidean distances between each observation and the cluster centroid. Operationally, K-means algorithms require the number of clusters to be pre-specified by the user. To determine the optimal number of clusters, two methods were used and compared: The Elbow Method and the Average Silhouette Method. Both methods require the use of clustering visualization packages in R, specifically, the packages "Cluster" (Maechler, 2021) and "Factoextra" (Kassambara and Mundt, 2020). The Elbow Method computes the K-means algorithm for a range of different *k* values, then for each *k*, calculates the total intra-cluster sum of squares (wss). Then *k* is plotted vs wss: as *k* increases, wss decreases. The first instance of a drastic decrease of wss will form an elbow and is considered the optimal number of clusters as the optimal number of clusters is the value of *k* where wss is minimized while also not overfitting the data into more clusters (Bholowalia, 2014) (Figure 2.3).

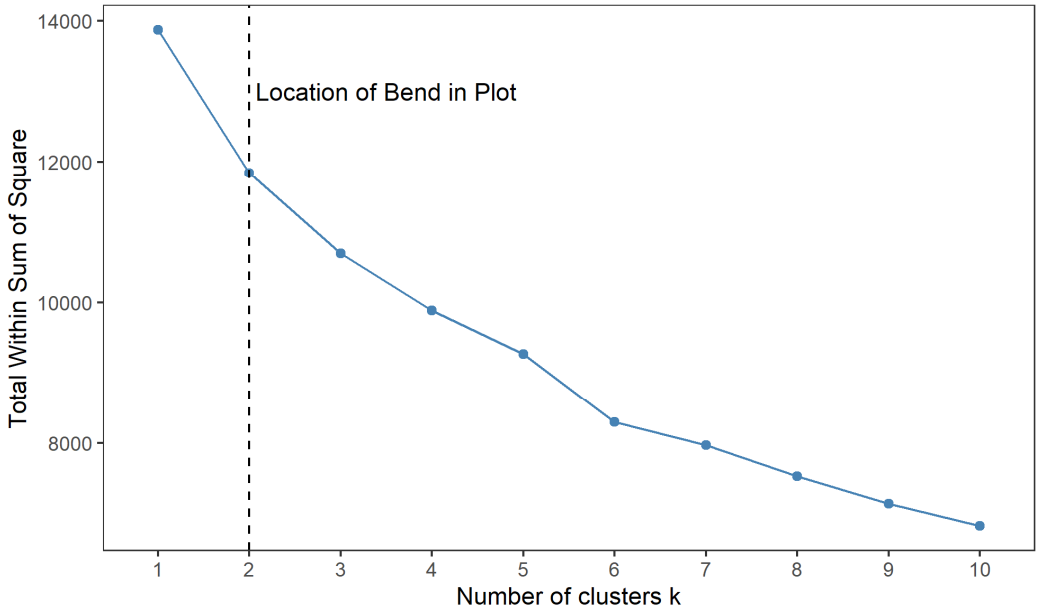

Figure 2.2: Number of clusters determined for K-means clustering using the Elbow Method

Silhouette plots were also created to show how well each observation fits within its cluster. The silhouette coefficient shows the degrees of cohesion or separation of the clusters according to the following equation:

$$s(n) = \frac{b(n)-a(n)}{\max\{a(n),b(n)\}} \quad (2.1)$$

Where s(n) is the average silhouette width at data point *n*, a(n) is the mean distance between *n* and all other data points in the cluster to which *n* belongs, and b(n) is the minimum mean distance from *n* to all other clusters to which *n* doesn't belong. S_{mean} can then be calculated for every *k*:

$$S_{mean} = mean\{s(n)\} \quad (2.2)$$

The optimal *k* is the one with the highest silhouette coefficient (Saputra, 2020), (Figure 2.4B). S(n) can also be plotted against each *k* (Figure 2.4A) to assess the total fits of data points in each cluster as opposed to just an average. Silhouette coefficients range on a scale of -1 to 1, with 1

being the best score meaning every observation is very compact within its cluster and far away from others. -1 indicates the opposite, that each observation does not fit very well in its cluster. 0 indicates an overlapping of clusters. While the selection of the elbow from the elbow method can appear arbitrary, the silhouette method uses a separate metric in which it seeks to determine the number of clusters in which the silhouette coefficient s(n) is maximized, which also occurs for two clusters. Therefore, both the Elbow Method and Average Silhouette Method agree that two clusters is the optimal number of clusters for this dataset.

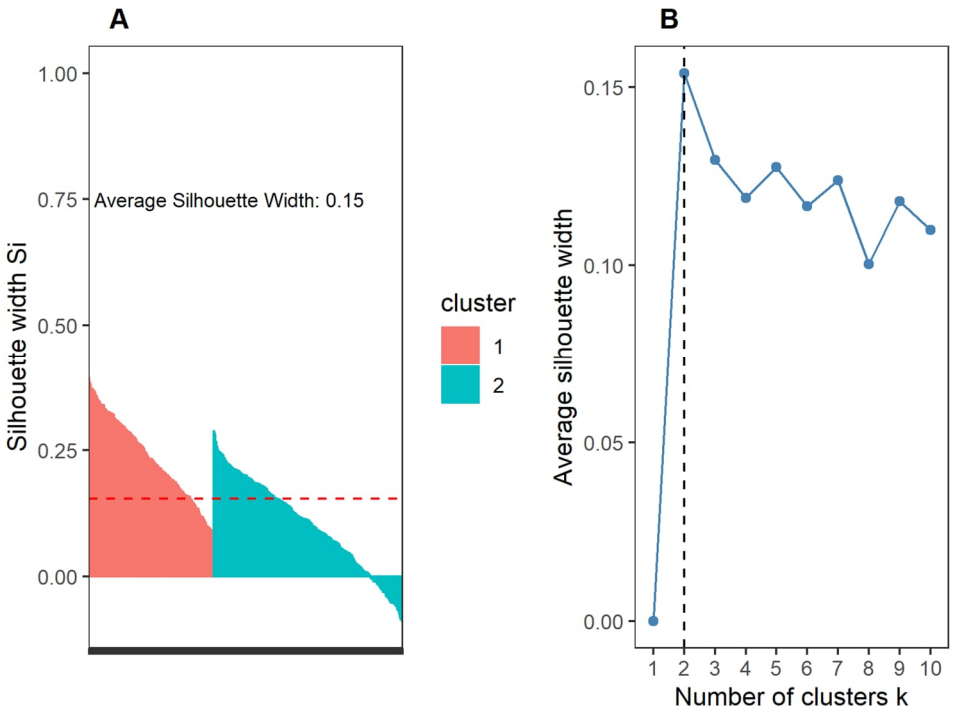

Figure 2.3: A) Silhouette plot showing average silhouette thickness and how well values fit in their respective clusters. B) Silhouette showing the average silhouette score for various numbers of clusters.

Once the number of clusters was determined, the Hartigan-Wong algorithm was run in R for 2 clusters. The Hartigan-Wong algorithm calculates the cluster centroids as the mean of datapoints *n* in that cluster where *wss* is minimized from centroid to datapoint *n*. Therefore, the

centroid is the most representative of that cluster as it is what is *wss* is being minimized in relation to. Using the K-means results, the center basins were identified as the basins where the Euclidean distance from the cluster centroid across all clustering input parameters is minimized, and therefore are the most representative basins of that cluster. As these center basins are determined to be the most representative of their respective clusters, their topographies are of special interest as they have a unique combination of elevation, slope, and aspect which are closest to the centroids which define the cluster. Figure 3.1 shows the locations of the identified center basins.

2.4 Phase 2: Hydrologic Modeling of Cluster 1 and Cluster 2 Center Basins

Phase 2 involved hydrologic modeling of the clustered basins identified in Phase 1. The goals of this modeling effort were to (1) see how much of an effect modeling decisions have on differences between the cluster hydrologic responses for the Center Basins, and (2), to see whether there are systematic differences between basins in each cluster. Modeling decisions were made to test the validity of the clustered basins using a hydrologic framework and aimed for consistency of comparison among basins, rather than for reality. Model calibration or validation techniques were not used as the modeling decisions are arbitrary and results were not intended to be used for real-world hydrologic predictions. The Hydrologic Modelling System (HEC-HMS) is a semi-physically-based modeling program developed by the US Army Corps of Engineers to simulate the complete hydrologic processes of watershed systems (USACE, 2000). It was chosen for this project due to its physical basis, past familiarity, and a straightforward graphical user interface. HEC-HMS requires an input terrain in the form of a DEM from which all stream reaches and subbasins are delineated. Hydrological models were set up for the basins

representing the cluster centers identified from Phase 1 as these basins are the most characteristic of their clusters. HEC-HMS delineates subbasins according to a stream drainage threshold and then uses those subbasins in an HRU-style modeling setup. To be consistent between basins and model instantiations, a stream drainage threshold of 2 km^2 was used. This stream drainage threshold was specifically chosen as it must be smaller than the smallest basin size. The delineation process consists of 3 steps: 1) Fill the sinks in the DEM, 2) Create flow accumulation and flow direction rasters, and 3) Identify streams from flow accumulation and flow direction rasters based on the stream drainage threshold. Once the subbasins are delineated and stream reaches are created, parameters can be checked and changed accordingly.

For the purposes of this project, a simple model set up was used with only loss and transform methods applied to each subbasin. In physical hydrology, loss methods are used to represent the physical calculations that describe infiltration, surface runoff, and subsurface flow within each subbasin (USACE, 2002). The SCS Curve number was chosen as the loss method for simple implementation and ease of tracking across basins and models. The SCS Curve number model estimates precipitation excess as a function of precipitation, soil cover, land use, and antecedent moisture (USACE, 2002). The curve number itself was developed as an empirical hydrological parameter by the United States Department of Agriculture Natural Resources Conservation Service (USDA NRCS) and is a function of soil and landcover type. These values are commonly found in a curve number lookup table with values ranging from 39 to 98 on a non-continuous scale corresponding to the combinations of soil type and land cover type. For predictive hydrologic modeling, assignment of curve numbers is done following an HRU approach where landcover and soil type observational datasets are combined to create an HRU for which a curve number can be assigned. However, as the modeling effort followed a “virtual

basin" approach – that is, a modeling setup that is strictly experimental and not necessarily completely realistic, a curve number of 62 was arbitrarily chosen. This curve number was specifically chosen as selecting a curve number of 39 or 98 would be on either extreme end of the curve number spectrum. Since the loss parameter will be applied uniformly to all delineated subbasins and across models, it was a valid assumption to select a curve number of 62 as it represents a more "average" curve number.

While the loss method represents the physical interaction of surface water runoff, infiltration, and subsurface processes, the actual surface water runoff calculations for each subbasin are performed according to a transform method (USACE, 2002). SCS Unit Hydrograph was chosen for transform method, lag times for each reach were calculated according to Watt and Chow 1985:

$$t_L = 0.000326\left(\frac{L}{\sqrt{S}}\right)^{0.79} \quad (2.3)$$

where t_L is the lag time [hr], L is the length of the longest flow path [km], and S is the slope of the longest flow path [unitless].

The HEC-HMS delineation process produces delineated subbasins in addition to reaches that drain those subbasins. For calculations to be performed, a stream routing method must be selected. HEC-HMS has 9 stream routing methods by default. For this study, the Muskingham-Cunge routing method was chosen as it has a long history of successful use in flood modeling and prediction scenarios. Muskingham-Cunge built upon previous methods identified in the Muskingham Method (Muskingham, 1959) by using a combination of the continuity equation and a simplified version of the momentum equation (USACE, 2002). In addition, routing parameters can be easily measured or calculated rather than having to be determined from historical flood data (Cunge, 1969). A sample table of parameter values is shown in Table 2.1.

While many parameters are held constant between models, Length (M) and Slope (M/M) are dependent on individual reach geometries and therefore differ from reach to reach but are calculated using the same equation.

Table 2.1: Sample routing parameters chosen for Muskingham-Cunge Method for one reach. Length (M) and Slope (M/M) are reach-specific parameters, bold text indicates parameter values that remain constant for all reaches

Initial Type	Discharge = Inflow
Length (M)*	3123.82
Slope (M/M)*	0.0429
Manning's number*	0.04
Space-Time Method	Auto DX Auto DT
Index Method	Flow
Index Flow (M3/S)*	100
Shape	Trapezoid
Bottom Width (M)*	5
Side Slope (xH:1V)*	0.33
Invert (M)	

HEC-HMS also requires control specifications to run a simulation. The control specifications are shown below in Figure 2.5. The start date and end dates are arbitrary as we are not calibrating to real world gauge data where exact dates would be of concern. The total duration for each storm event is 36 hours, and 1 hour time intervals were chosen for model simplicity.

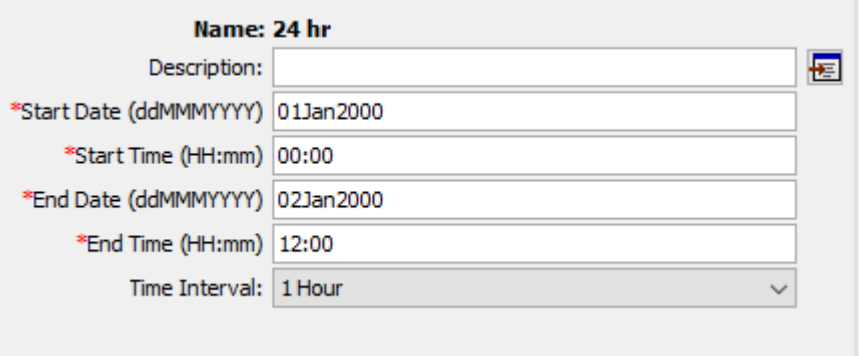

Figure 2.4: Control specifications used in HEC-HMS model

Precipitation conditions are needed for the delineated subbasins to supply water and ultimately drive the model. There are numerous possible meteorological model configurations to choose from in HEC-HMS, but for the purpose of this project, a hypothetical storm was chosen as it allows precipitation to easily be kept consistent between models and real-world precipitation distributions are not a concern for this project. The hypothetical storm drops an arbitrary amount of rain over user-defined areas or over the entire model domain. To maintain consistency, 50.8 mm of rain was dropped equally over each part of each basin. Once all parameters were calculated and model components were constructed, the model was run. Figure 3.6 below shows the hydrograph results of this model run with individual hydrographs labeled by the parameter changed in that model run. This first model run for each cluster is referred to as baseline.

2.5 Phase 2: Sensitivity Analysis of Cluster 1 and 2 Center Basins

Once the basin models and components were constructed and successfully run, sensitivity analyses were performed to see how these models performed with changing various parameters. The goal of the sensitivity analysis was to understand if the change in hydrograph characteristics resulted from changing input parameters between model runs or if changes evident are due to inherent differences in the underlying topographies. The parameters changed were lag time, impervious percentage, and the amount of rain dropped on each basin. A summary of model parameters and associated values is shown in Table 2.2.

Table 2.2: Model run names and parameters changed between runs

Model Run	Rain Amount (mm)	Curve Number	Impervious %	Lag Time
Baseline	50.8	62	10	Standard
Run 2	50.8	62	10	2x Standard
Run 3	50.8	62	10	0.5x Standard
Run 4	101.6	62	10	Standard
Run 5	50.8	62	30	Standard

The model runs correspond to different hydrographs created as outputs from HEC-HMS. The hydrograph outputs are shown in Figure 3.6 and discussed in the results section. The modeling of the center basins will be referred to as the Center Basin Test.

2.6 Phase 2: Determination of Summary Statistics

To make meaningful comparisons between clusters beyond visual interpretations, summary statistics were calculated for each cluster. Summary statistics are beneficial as they provide a quantitative measure that can be used to assess the similarities and/or differences between clusters. The following statistics were calculated for each cluster and model run: Time to Peak Flow (TP), Time to Baseflow from Peak (TB), Intercept of the Recession Line, Slope of the Recession Line (b), and R^2 of recession regression.

The summary statistics Intercept of the Recession Line, Slope of the Recession Line (b), and R^2 of recession regression were all determined by applying a linear regression to streamflow (Q) and recession rate (-dQ/dt). The apparent visual differing streamflow recessions according to the hydrographs for the basins closest to the centroids of Cluster 1 and Cluster 2 (Figure 3.6) suggested these two center basins might be different in terms of their recession characteristics and therefore should be quantified according to recession metrics. Recession events refer to the falling limb of the hydrograph, that is after the peak flow as streamflow (Q) returns to baseflow conditions. These recession periods contain temporal information on how watersheds move water after precipitation events and are therefore useful to assess hydrologic differences between watersheds in terms of their recession behavior. Previous studies (Kirchner, 2009; Stoelzle, 2012) have used streamflow recession analysis to compare streams and hydrological characteristics of watersheds. Streamflow recession analysis assumes a water balance equation

where (Q) is solely responsible for changes in storage as precipitation and evapotranspiration are negligibly small compared to (Q) (Stoelzle, 2012). This is a reasonable assumption for this modeling exercise as the model duration is 36 hours meaning water lost to evapotranspiration is likely much smaller than water leaving the system due to surface runoff. Furthermore, this assumption holds for storm-scale events, but wouldn't be appropriate for longer scales. Following the work of Kirchner (2009), (-dQ/dt) was defined as the difference in discharge between two successive hours following the peak (Kirchner, 2009), according to the equation:

$$-\frac{dQ}{dt} = \frac{Q_{t-\Delta t} - Q_t}{\Delta t} \tag{2.4}$$

These recession rates are then averaged between the two hours as to not introduce any artificial correlation between discharge (Q) and (-dQ/dt) as these values will span multiple orders of magnitude (Kirchner, 2009), according to the equation:

$$-\frac{dQ}{dt} = \frac{Q_{t-\Delta t} + Q_t}{2} \tag{2.5}$$

Once (-dQ/dt) was determined, it was plotted as a function of time to visualize changes in recession rate over time. A commonly used streamflow recession model, including the one used by Kirchner (2009) uses power law functions following the form of the differential equation:

$$\frac{dQ}{dt} = -aq^b \tag{2.6}$$

to explain the nonlinear relationship between discharge and streamflow during recession events (Boussinesq 1877; Dralle, 2015; Dralle, 2017). To quantitatively summarize recession rate as a function of discharge, (-dQ/dt) was plotted against (Q) in log-log space. Plotting (Q) against (-dQ/dt) on a log-log plot supports a roughly linear relationship and allows for a linear regression to be performed. A linear regression was performed solely for the purpose of obtaining the slope value (b) from the regression equation (EQ 2.6) as this value will be used to describe the

recession of that model run for that cluster. Furthermore, the linear regression approach is used to provide a quantitative description of the differences between model runs that can be seen qualitatively in the hydrographs. The fit of the linear models was also inspected by plotting residuals and evaluating R^2.

2.7 Phase 2: Hydrologic Modeling Randomly Selected Basins

The same modeling experiment was conducted on randomly selected basins in Clusters 1 and 2 to see if hydrograph responses to a storm event vary systematically between clusters. Eight basins were randomly selected from each cluster, and subsequent HEC-HMS models were set up for each basin using the baseline parameters outlined above. Conducting a sensitivity analysis of the randomly selected basins was not necessary as the Cluster 1 Center and Cluster 2 Center model results indicated how the models would perform when changing various hydrological inputs. The same summary statistics were also calculated for these randomly selected basins. Modeling of the randomly selected basins in each cluster will be referred to as the Random Basin Test moving forward.

2.8 Phase 2: Statistical Significance of Summary Statistics

To assess whether the K-means clustering could be validated by quantifying the hydrological differences between catchments in each cluster it is necessary to see if the statistics calculated are from different distributions or from the same distribution. If they are from the same probability distribution, then the summary statistics would be suggesting that there is not a large difference in hydrological signatures between basins. A Kolmogorov-Smirnov (KS) test is a non-parametric statistical test and was used as to not have to rely on data normality. The KS

test was conducted with an α of 0.05. The 2 sample KS test states the null hypothesis is that the samples are drawn from the same continuous distribution. A KS test works by comparing cumulative distribution functions of the two datasets and creating a test statistic that describes the difference between these two distributions (Kroll, 2015). If the KS test can reject the null hypothesis, that would indicate that the summary statistics are quantitively describing different distributions, and that there is a numerical difference between Cluster 1 and Cluster 2 according to these summary statistics. The KS test was run on the summary statistics TP, TB, b, and R^2 for both the Center Basin Test and the Random Basin Test.

CHAPTER 3

RESULTS OF K-MEANS CLUSTERING ANALYSIS AND HYDROLOGIC MODELING OF CENTER BASINS AND RANDOMLY SELECTED BASINS

3.1 Results of Phase 1: K-means Clustering Results

Running the K-means clustering algorithm for 2 clusters resulted in 240 basins in Cluster 1, and 364 basins in Cluster 2. A map of the clustered basins is shown in Figure 3.1 in addition to the center basins identified from the clustering effort. The center basins identified from the K-means clustering will be referred to as Cluster 1 Center and Cluster 2 Center moving forward. Cluster 1 Center is 1080.38 km^2 and is in north-northwest Oregon while Cluster 2 Center is 32.18 km^2 and is in western Montana. Interpretation of the map of clustered basins (Figure 3.1) suggests the spatial distributions of basins assigned to each cluster appear relatively uniform across the entire study area. However, the largest basins by area in the dataset tend to be in Cluster 2 and are in Idaho and Montana. The distributions of slope and aspect suggest for each basin indicate Cluster 1 Center has more complex topography, with more many ridges and valleys compared to Cluster 2 Center (Figure 3.2).

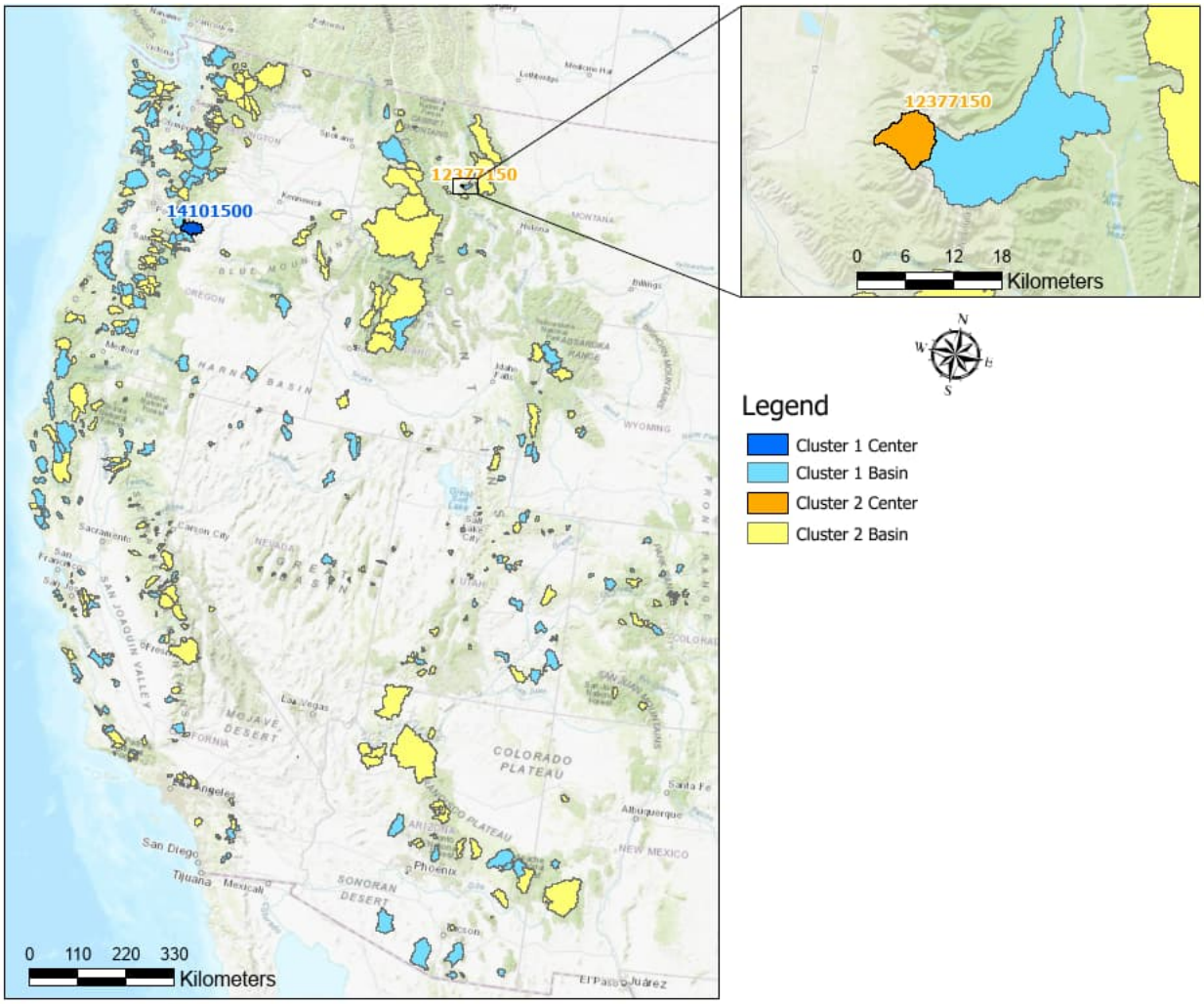

Figure 3.1: Map of GAGES-II watersheds and the results of Phase 1 K-means clustering analysis and center basins identified. The center basins are labeled by their station ID.

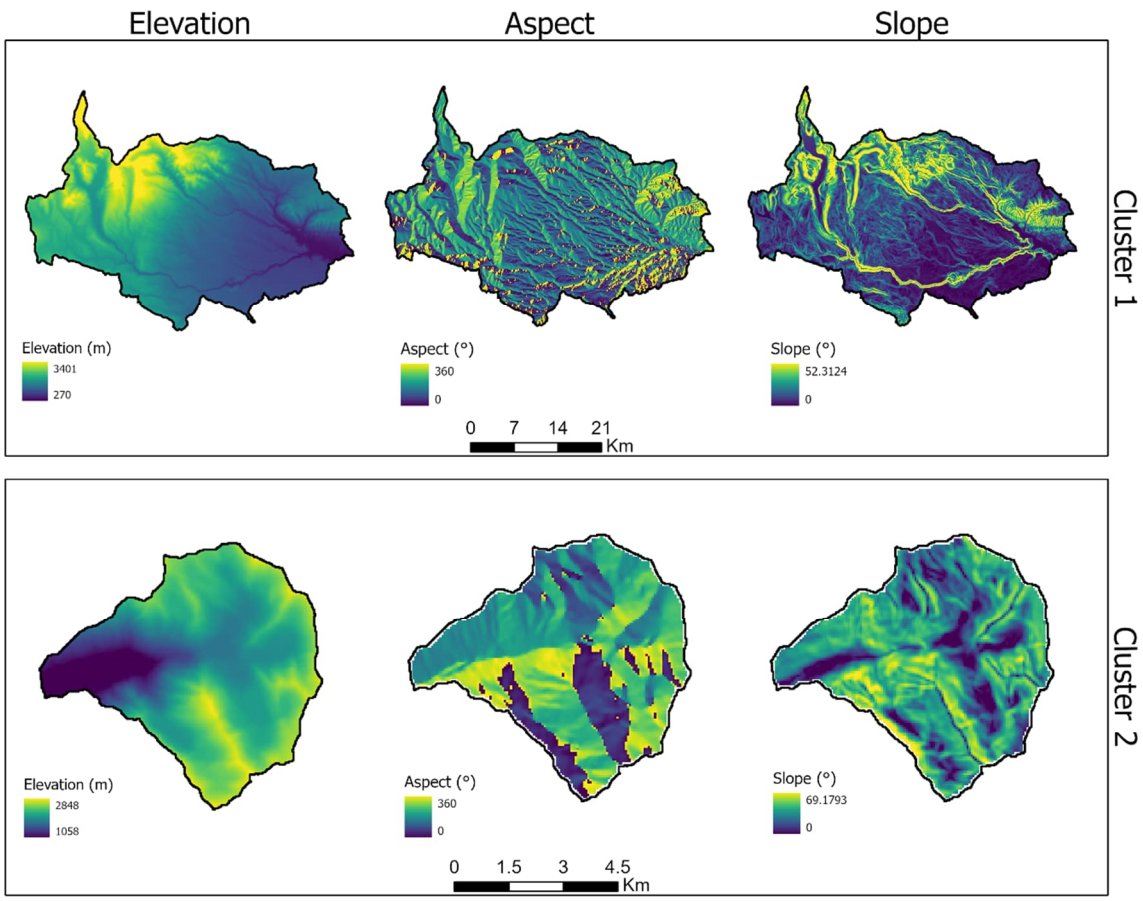

Figure 3.2: Elevation, aspect, and slope rasters for the center basins identified from the K-means clustering effort. Note the difference scales between Cluster 1 Center and Cluster 2 Center.

Distributions of elevation, slope, and aspect within each cluster show Cluster 1 contains many more pixels in the elevation range of 0-1000 m whereas Cluster 2 contains more pixels in the elevation range of 1000 m to 2500 m, and both clusters have few pixels with elevations greater than 3000 m (Figure 3.3A). The K-means clustering algorithm was blind to elevation, and only saw the percent of pixels within each elevation threshold. The slope distributions between clusters are consistent, however Cluster 2 has more pixels with slopes in the range of 20° to 40° (Figure 3.3B). No meaningful patterns are discernable comparing the distributions of aspects between clusters.

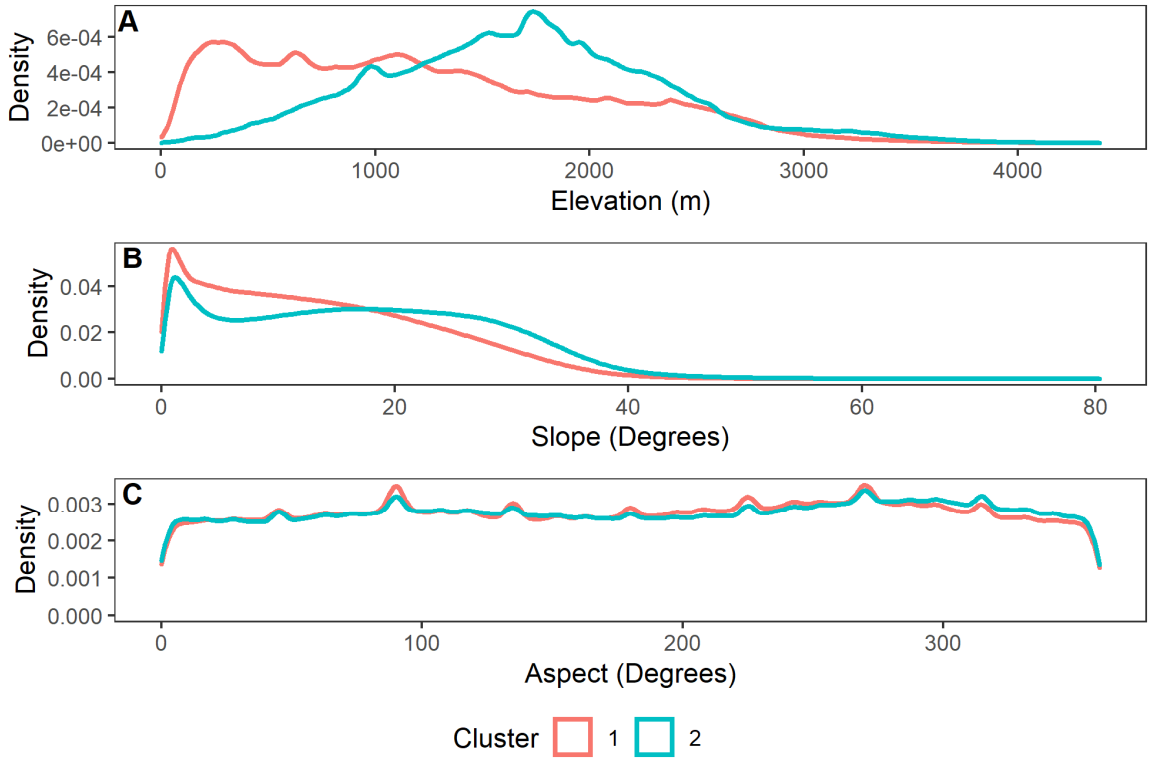

Figure 3.3: A) Density plot of elevation values colored by cluster. B) Density plot of calculated slope values colored by cluster. C) Density plot for calculated aspect values colored by cluster.

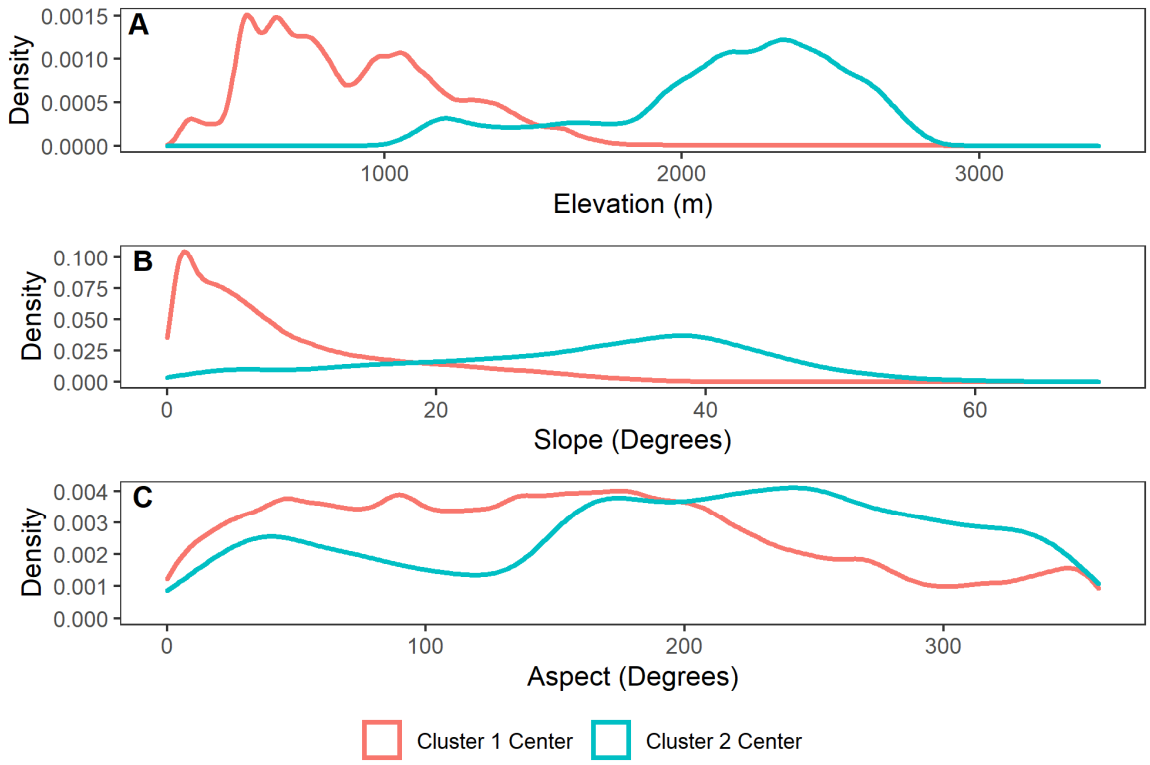

Figure 3.4: A) Density plot of elevation values for center basins colored according to Cluster. B) Density plot of calculated slope values for center basins colored according to Cluster. C) Density plot for calculated aspect values for center basins colored accord.

Comparing the topographic distributions of Cluster 1 Center and Cluster 2 Center, Cluster 1 Center has a greater density of pixels in the 0-1200 m band whereas Cluster 2 Center has a greater density of pixels in the 1900-2900 m band (Figure 3.4 A). Comparing the slope distributions (Figure 3.4 B), Cluster 1 Center has a higher density of lower angle slopes (0°-20°) compared to Cluster 2 Center, which has a higher density of slopes in the 20° to 45° range. These densities can be seen visually (Figure 3.2), with Cluster 1 Center having a basin that is dominated by lower elevations and smaller slopes across the entire catchment. The aspects density plots (Figure 3.4 C) suggests that Cluster 1 Center has more easterly aspects, while Cluster 2 Center has more westerly aspects. This is also supported by their positions on the map (Figure 3.1).

Of the 604 basins in the GAGES-II reference basins dataset, Cluster 1 has a higher average fraction of pixels within 0-30% of the total elevation range while Cluster 2 has a higher fraction in the range of 40-100% of the total elevation range (Figure 3.5 A). This difference of average normalized elevations is also reflected in the distributions of absolute elevations for each cluster (Figure 3.4 A). The fraction of pixels within the 0°-10° slope threshold is slightly higher for Cluster 1, however the average slope fraction for Cluster 2 within this threshold is within 1 standard deviation (Figure 3.5 B). Cluster 2 tends to have a higher fraction of pixels with steeper slopes (>20°) (Figure 3.5 B). The average fraction of pixels is equal across all aspects, with standard deviations being roughly equal (Figure 3.5 C). The fractions of each discretized topographic criterion suggest aspect is not as useful for discretely classifying basins compared to normalized elevation thresholds and slope thresholds.

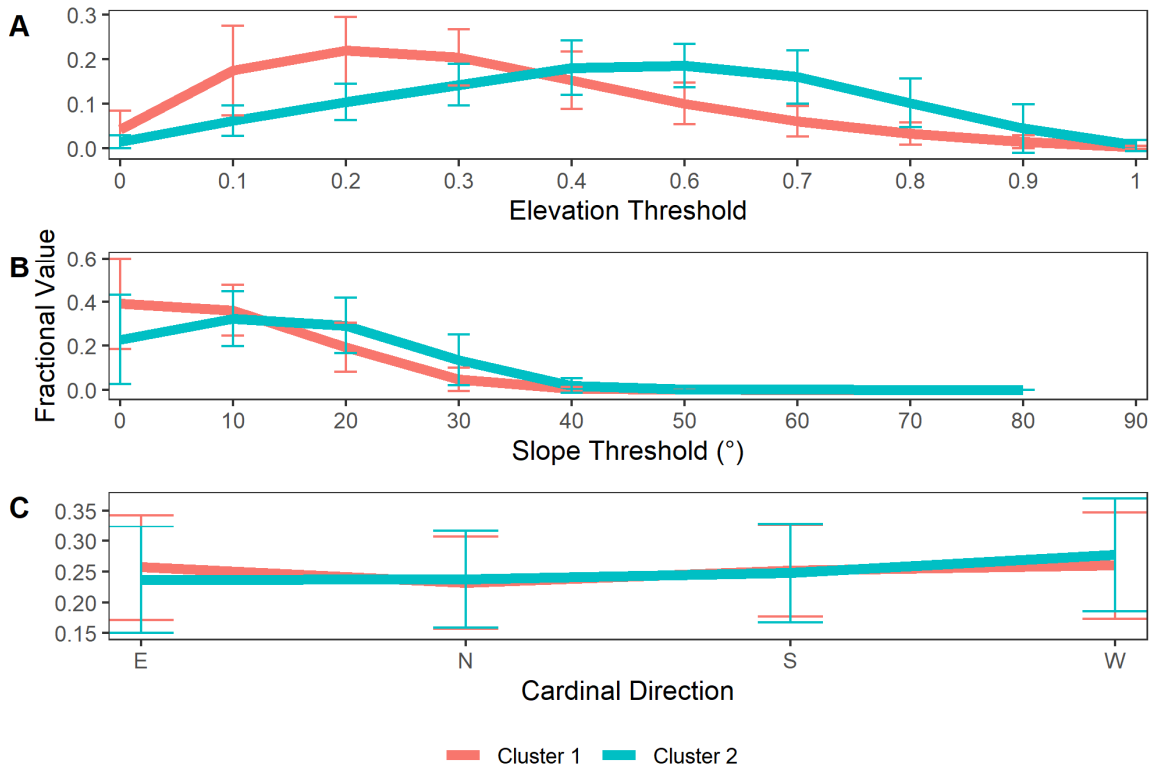

Figure 3.5: A) Fraction of average normalized elevation values for all basins colored according to cluster. B) Fraction of average slope values for all basins colored according to cluster C) Fraction of average aspect values for all basins colored according to cluster. For all 3 plots, the mean of each cluster is shown with standard deviations for each threshold value.

3.2 Results of Phase 2: Hydrologic Modeling of Cluster 1 & Cluster 2 Center Basins

Hydrographs from Cluster 1 show more variation in the timing of peak flow in response to modeling decision variability than in Cluster 2. Changing the parameters in Cluster 1 results in both a change in the magnitude of peak flow, as well as a change in the timing (Figure 3.6). When making the same changes for Cluster 2, the trends observed in Cluster 1 are not observed in Cluster 2. In Cluster 2, we only see changes in magnitude of peak flow but not in the timing. This is quantified by TP (Table 3.1). In Cluster 1, TP changes based on modeling decisions but is unchanged in Cluster 2. However, in Cluster 2, TB is much faster in Cluster 2 than in Cluster 1 (Table 3.1). TP and TB suggest that flow in Cluster 2 “flashes” through the system as opposed to in Cluster 1.

Figure 3.6 shows the hydrographs for the Baseline and subsequent model runs for the center basins of both clusters. The baseline scenario refers to the baseline models for both Cluster 1 Center and Cluster 2 Center. For Cluster 1 Center Run 2, doubling the lag time lowers the peak flow by ~21% relative to the baseline run while also retarding the timing of peak flow by 1 hour (Table 3.1). For Cluster 2 Center Run 2, doubling the lag time lowers the peak flow by ~10% relative to the baseline run, but it does not change the timing of peak flow, it just changes the magnitude of the peak. Lag times were halved for the Run 3, resulting in an increase of the peak flow by ~12% relative to the baseline run while also advancing the timing of peak flow by 1 hour for Cluster 1 Center. For Cluster 2 Center, neither the magnitude nor timing of peak flow changed from the baseline configuration. Doubling precipitation amount for Run 4 resulted in an increase in the peak flow by ~390% relative to the baseline run while also advancing the timing of peak flow by 1 hour for Cluster 1 Center. In Cluster 2 Center, the peak flow increased by ~365% relative to the baseline run while the timing of peak flow did not change. Impervious percentage was changed from 10% to 30% for Run 5 resulting in an increase of peak flow of ~176% relative to the baseline scenario while advancing the timing of peak flow by 1 hour for Cluster 1 Center. For Cluster 2 Center, the peak flow increased by ~198% relative to the baseline scenario while the timing of peak flow did not change.

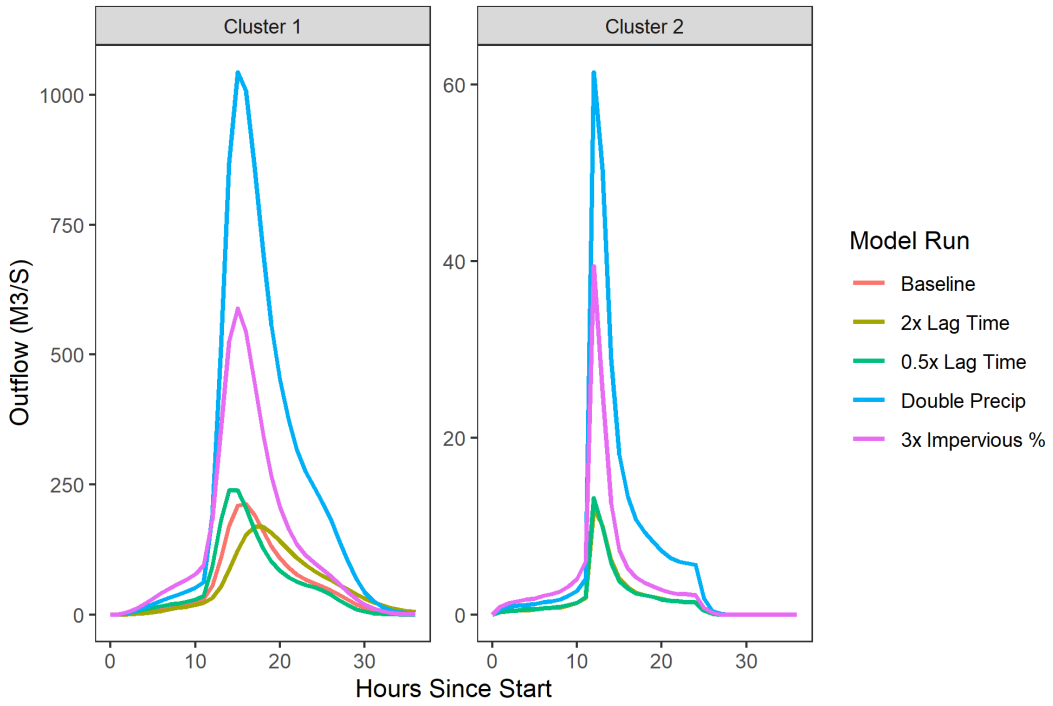

Figure 3.6: Hydrographs for Clusters 1 Center and Cluster 2 Center for the baseline scenario and subsequent model runs. Note the variable y-axes.

Based on graphical interpretation, Cluster 2 model runs have a much faster rate of recession than those in Cluster 1. The flows in Cluster 2 reach the peak flow and recedes back to baseflow or no flow much faster than in Cluster 1 where a gradual recession is seen. Figure 3.7 shows recession rates for each model run plotted versus time. The magnitude of recession rate is directly correlated to the peak flow, where higher peak flows equate to higher recession rates (Double Precip and 3x Impervious % Figure 3.7). This is quantified by an average correlation coefficient across model runs for (Q) and (-dQ/dt) of 0.92.

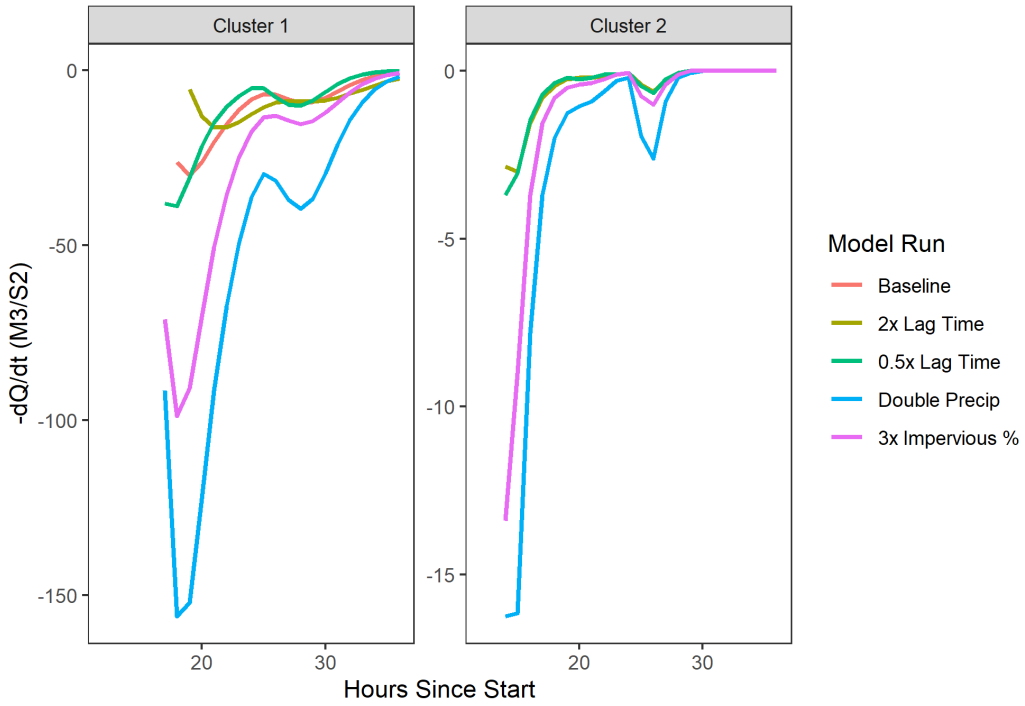

Figure 3.7: Recession rate -dQ/dt plotted against time for the baseline scenario and subsequent model runs for both Cluster 1 Center and Cluster 2 Center. Note the variable y-axes.

From looking at the shape of these hydrographs (Figure 3.6), it appears that Cluster 1 basins have slower recession rates than Cluster 2 basins. If Cluster 1 has slower recession rates, this should be evident by a shallower slope of the recession line. Running linear regressions for each model shows the slopes of recession lines for the various model runs are steeper in Cluster 1 than those in Cluster 2 (Figure 3.8). It is also noted both visually and from the R^2 values in Table 3.1 that the linear model better fits the data in Cluster 1 than in Cluster 2. The R^2 values are much closer to 1 in Cluster 1 and much closer to 0 in Cluster 2, suggesting a linear model does not best fit the data.

Table 3.1: Calculated summary statistics for Clusters 1 and 2 Center Basin Model Test

Cluster	Case	Time to Peak (Hr)	Time to Baseflow from Peak (Hr)	Intercept of Regression Line	Slope of Regression Line (b)	r squared
Cluster 1	Baseline	16	20	0.07	0.60	0.90
	Run 2	17	19	0.21	0.43	0.69
	Run 3	15	21	0.13	0.58	0.86
	Run 4	15	21	0.27	0.63	0.92
	Run 5	15	21	0.09	0.68	0.93
	Average	15.6	20.4	0.15	0.58	0.86
Cluster 2	Baseline	12	15	-0.47	0.37	0.00
	Run 2	12	15	-0.53	0.60	0.09
	Run 3	12	15	-0.47	0.37	0.00
	Run 4	12	16	0.04	0.31	0.05
	Run 5	12	15	-0.40	0.69	0.11
	Average	12	15.2	-0.37	0.47	0.05

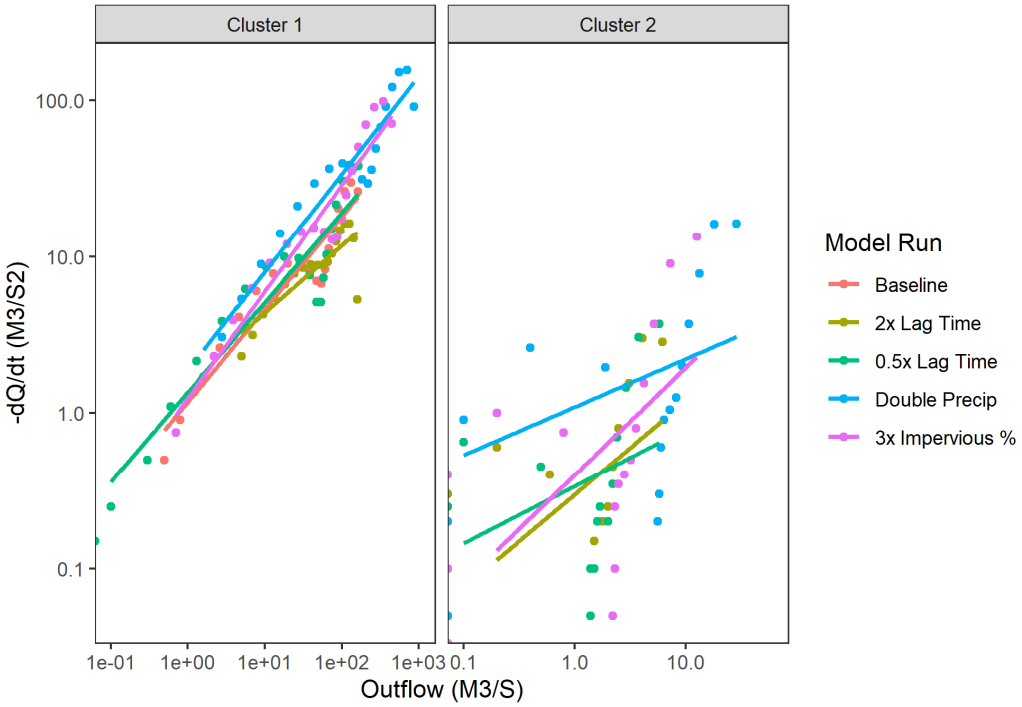

Figure 3.8: Linear regression of the Outflow (Q) against Recession rate (-dQ/dt) for the baseline scenario and subsequent model run for Cluster 1 Center and Cluster 2 Center.

Differences in topography between these two cluster center basins have a larger impact on the resulting hydrograph than the modeling decisions. The resulting p values from the KS test for TP, TB, b, and R^2 are 0.01, 0.01, 0.32, and 0.01 respectively. With an α of 0.05, we can reject the null hypothesis for TP, TB, and R^2, meaning that these summary statistics are describing different distributions, and that there is a statistical significance to these statistics which are being used to suggest differences between hydrographs from the two basins closest to the center of Clusters 1 and 2, regardless of modeling uncertainty. The p value of 0.32 for b is larger than an α of 0.05, suggesting that there is no statistical difference between the slopes from either center basin.

The R^2 values (Table 3.1) and visual fit of the recession lines (Figure 3.8) for the various model runs are not a good fit, especially for Cluster 2. This poor fit indicates that the data is not very well fit by a linear model. If the data is not well fit by a linear model, then it likely violates some or all the criteria necessary to fit a linear model, in particular, that residuals are homoscedastic. Figure 3.9 compares the fits of residuals for two model runs chosen to illustrate the differences in residuals between a good fitting model and a poor fitting model: Cluster 2 Center Run 4 and Cluster 1 Center Run 2. Cluster 2 Center Run 4 has an R^2 value of 0.049, suggesting a linear model does not fit the data well. Looking at the Residuals vs Fitted plot (Figure 3.9 A), the data appears heteroscedastic, with the standard deviation of residuals increasing with larger fitted values. Comparing this to Cluster 1 Center Run 2, which is better fit by a linear model as indicated by an R^2 value of 0.69, the Residuals vs Fitted plot (Figure 3.9 B) suggests less heteroscedasticity than for Run 4 Cluster 2.

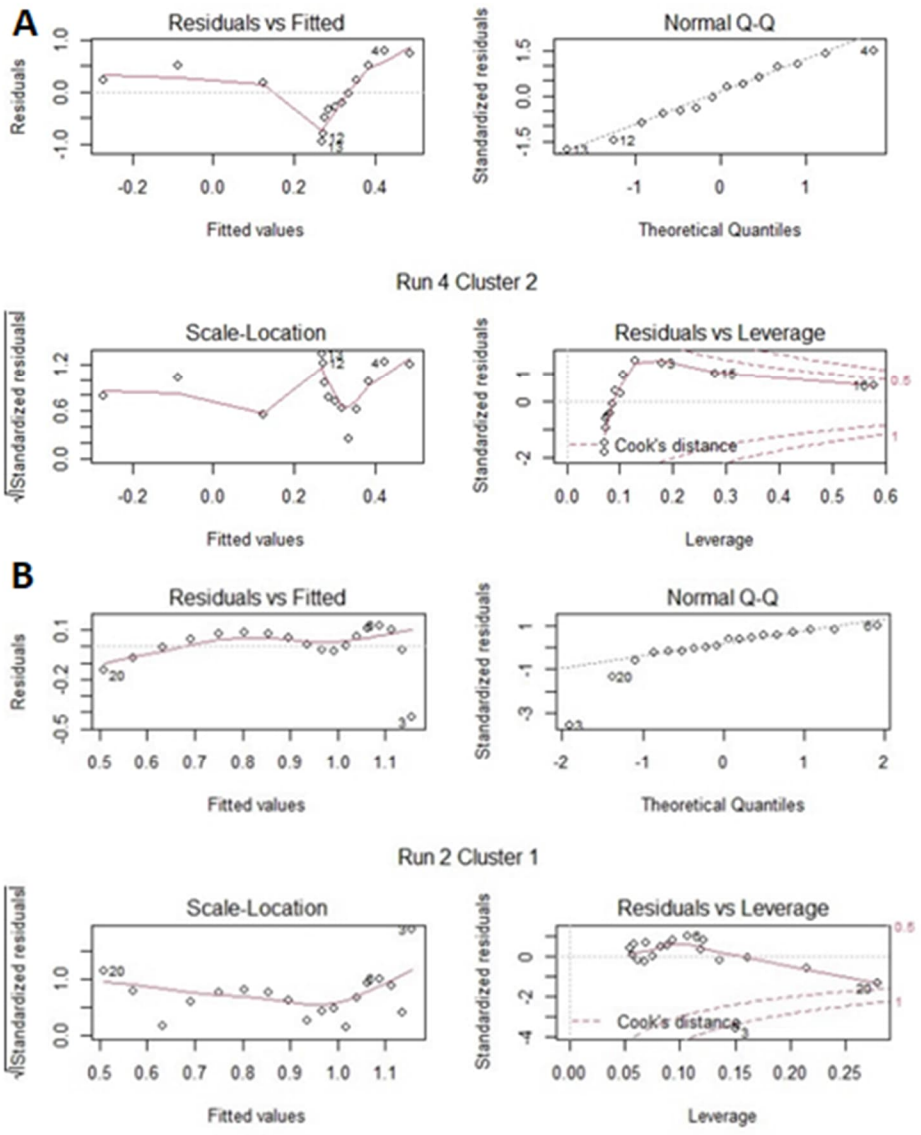

Figure 3.9: A) Plots assessing distribution of residuals for Cluster 2 Center Run 4 B) Plots assessing distribution of residuals for Cluster 1 Center Run 2

3.3 Results of Phase 2: Hydrologic Modeling of Randomly Selected Basins in Clusters 1 and 2

Figure 3.10 shows the locations of the randomly selected basins, basins range in area from 14 km^2 to 310 km^2 and roughly follow the trends of the geographic distribution of clustered basins resulting from the K-means analysis. More basins in Cluster 1 were selected on the west

coast region whereas more basins from Cluster 2 were selected in the Inter Mountain West and Rocky Mountain regions.

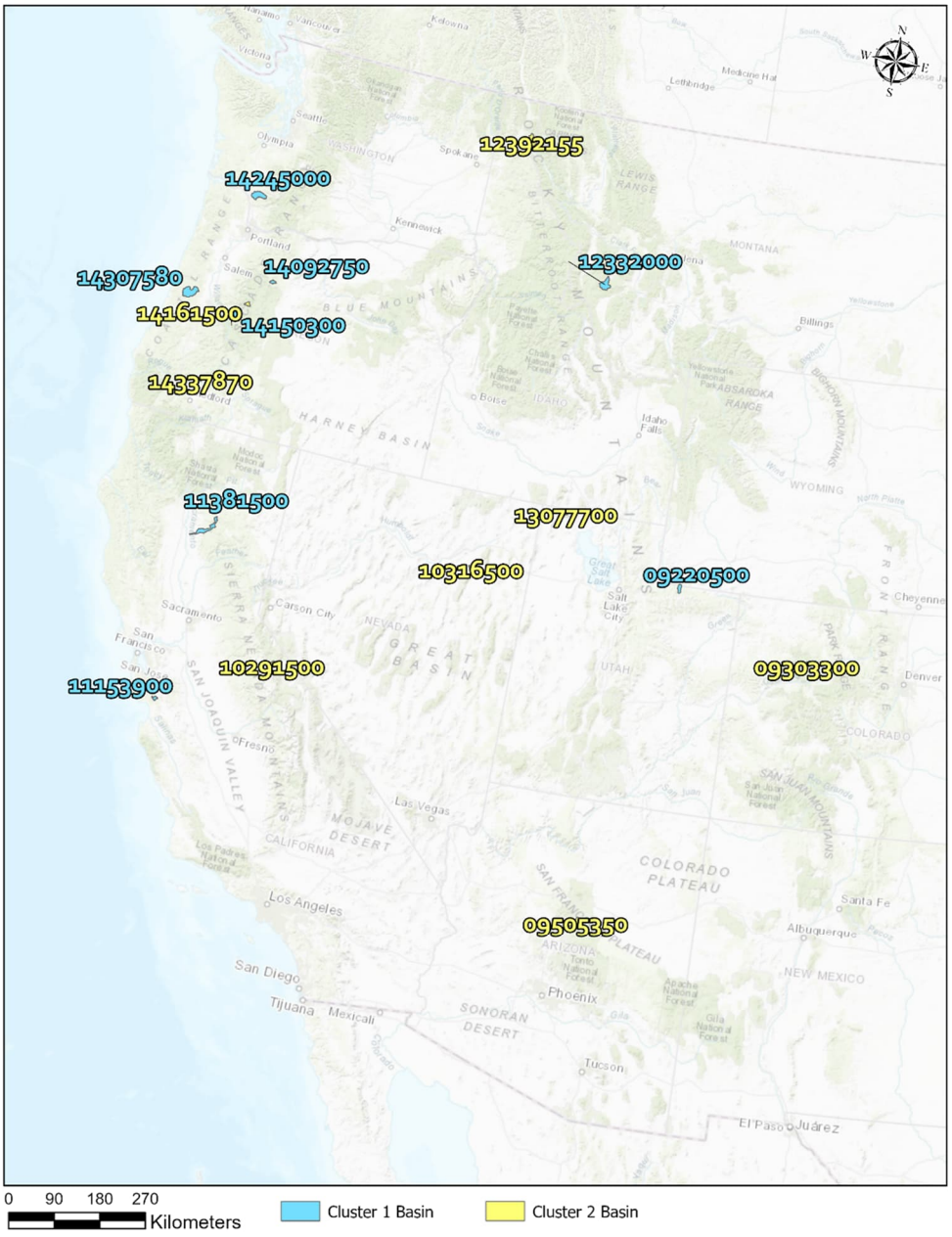

Figure 3.10: Map of randomly selected GAGES-II reference basins. The basins are labeled by Station ID.

The hydrographs from these model runs (Figure 3.11) also suggest basins were sampled from an equal distribution of areas as the peak flows are roughly the same between clusters. Some of the same trends identified from the cluster centers are also evident, such as the apparent faster rates of recession in Cluster 2 and faster time to baseflow (TB) in Cluster 2 (Figure 3.12). The average TB (Table 3.2) supports this notion that flow in Cluster 2 returns to baseflow faster than in Cluster 1. Also evident from both the hydrograph (Figure 3.11) and TP (Table 3.2) is that the basins in Cluster 2 reaches peak flow faster than those in Cluster 1. The summary statistics TP and TB support the trend seen in the center basin models that basins in Cluster 2 appear to have hydrograph characteristics in which water moves through the system faster than the in the basins in Cluster 1. A KS test comparing the TP and TB distributions across clusters indicated that these differences were not statistically significant ($p > 0.05$).

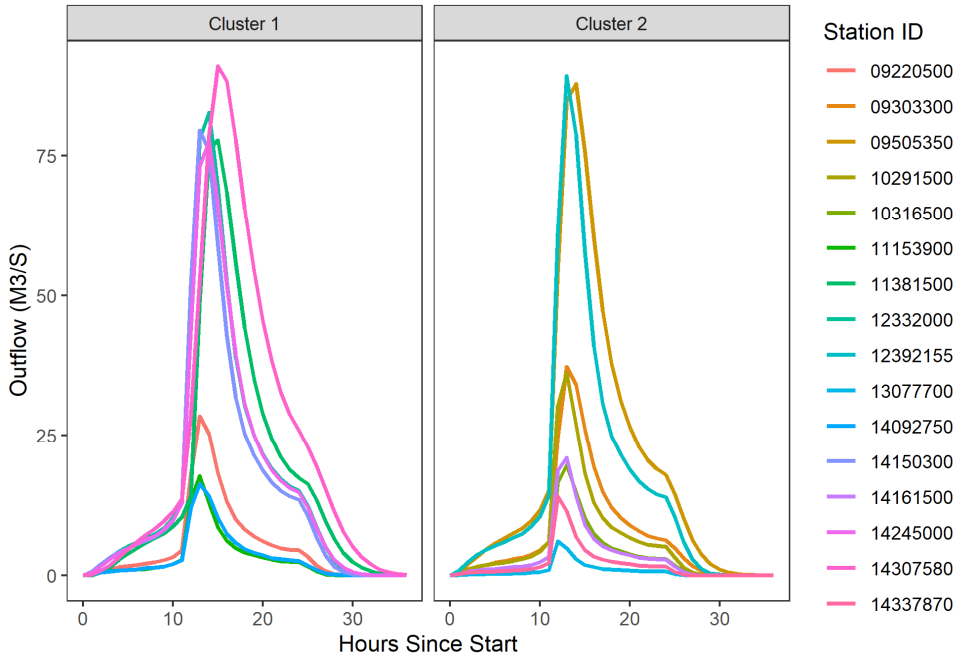

Figure 3.11: Hydrographs for randomly selected basins in Clusters 1 and 2 labelled by Station ID.

Comparing the recession rates from the center basin model tests versus those of the randomly selected basins, some of the same trends are evident. (Q) and (-dQ/dt) are still highly correlated, with an average correlation coefficient across models of 0.88 (Figure 3.12). The relationship between recession rates and discharge in Cluster 1 was much more linear compared to Cluster 2, resulting in a better fitting linear model as can be seen visually (Figure 3.6) in addition to higher R^2 values of 0.86 (Cluster 1) vs 0.049 (Cluster 2) (Table 3.1). However, for the random basin test, the relationship between (Q) and (-dQ/dt) does not appear as linear in Cluster 1 or Cluster 2 (Figure 3.13) as it does in the Cluster 1 Center model. The average R^2 values are 0.55 (Cluster 1) and 0.28 (Cluster 2), suggesting that the Cluster 2 basins from the random basin test are better fit by a linear model than that of the Cluster 2 Center basin test. Also evident in this test are the relative fits of linear models between Cluster 1 and Cluster 2. A linear model better fits the data for Cluster 1 compared to Cluster 2 as shown visually (Figure 3.13) and by the comparison of R^2 values (Table 3.2). However, the KS test performed on R^2 values indicated that these values are not statistically significant ($p > 0.05$).

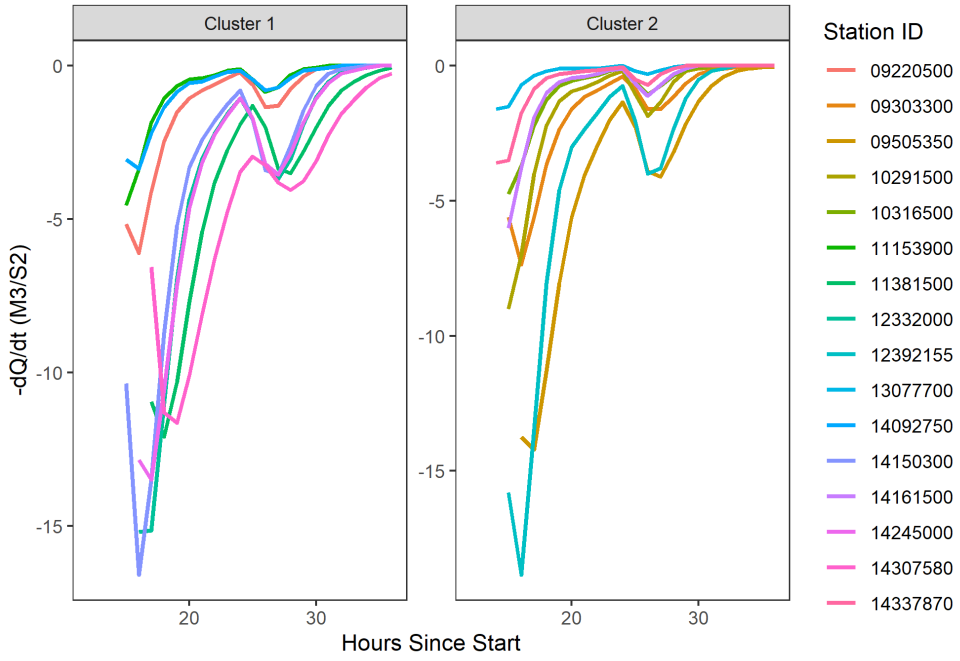

Figure 3.12: Recession rate (-dQ/dt) plotted against time for randomly selected basins in Clusters 1 and 2 labelled by Station ID.

Table 3.2: Calculated summary statistics for Clusters 1 and 2 Random Basin Model Test

Cluster	Station ID	Time to Peak (TP) (Hr)	Time to Baseflow from Peak (TB) (Hr)	Regression Intercept (a)	Regression Slope (b)	R^2
Cluster 1	09220500	13	18	-0.21	0.46	0.40
	11153900	13	16	-0.30	0.32	0.07
	11381500	15	21	-0.18	0.65	0.85
	12332000	14	19	0.01	0.50	0.66
	14092750	13	17	-0.33	0.44	0.29
	14150300	13	19	0.04	0.46	0.55
	14245000	14	19	0.01	0.50	0.67
	14307580	15	21	-0.02	0.54	0.93
	Average	13.75	18.75	-0.12	0.48	0.55
Cluster 2	09303300	13	19	-0.18	0.51	0.58
	09505350	14	21	-0.09	0.60	0.83
	10291500	13	17	-0.14	0.40	0.23
	10316500	13	16	-0.26	0.32	0.07
	12392155	13	18	0.09	0.43	0.39
	13077700	12	15	-0.58	0.54	0.09
	14161500	13	15	-0.28	0.30	-0.02
	14337870	12	15	-0.49	0.54	0.05
	Average	12.875	17	-0.24	0.45	0.28

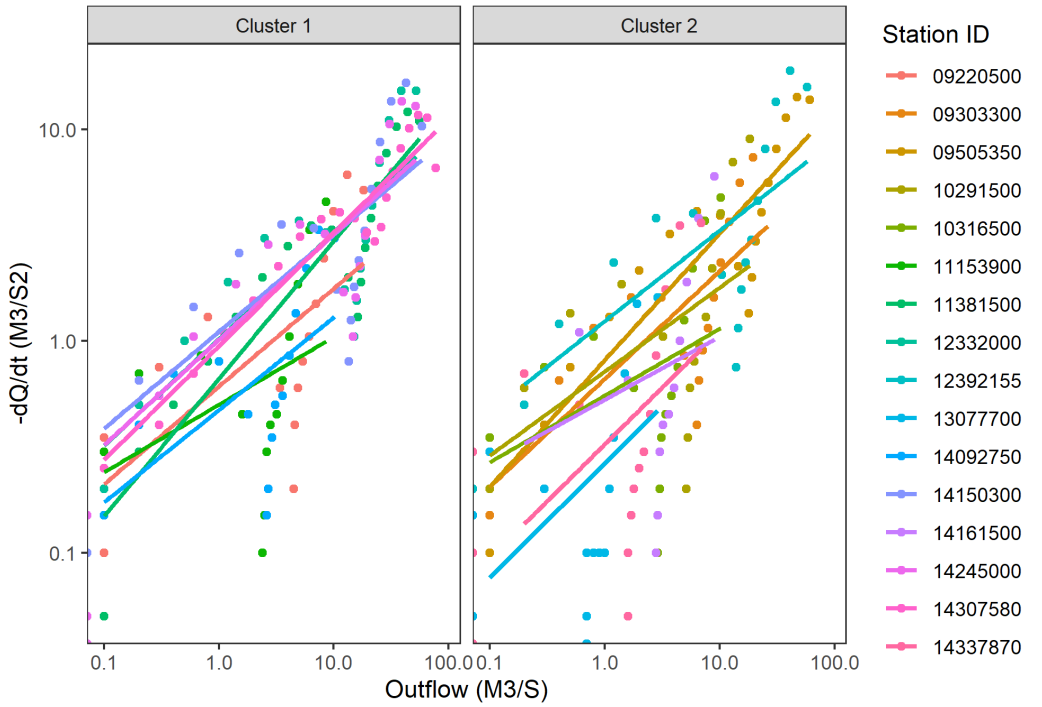

Figure 3.13: Linear regression of the Outflow against Recession rate (-dQ/dt) for the baseline scenario and subsequent model runs labelled by Station ID.

A KS Test was also performed for the summary statistics calculated for the randomly selected basins for the statistics: TP, TB, b, and R^2. The resulting p values are 0.62, 0.62, 0.66, and 0.28 for each of the 4 summary statistics respectively. With an α of 0.05, we cannot reject any of the null hypotheses that the statistics are describing different distributions. The results of this test suggest that the basins sampled for this random basin test are not different between Clusters 1 and 2.

Comparing the R^2 values from modeling of the randomly selected basins vs the center basins suggests that the random basins in Cluster 2 are better explained by the linear model than for the Cluster 2 Center, as indicated by the R^2 values (0.28 - Cluster 2 average, vs 0.049 – Cluster 2 Center average). However, even though the fit is better on average for Cluster 2, the linear models for all the randomly selected basins are heteroscedastic. Figure 3.14 compares the fits of residuals for two model runs as part of the random basin test: Cluster 1 Station ID

11153900 and Cluster 2 Station ID 13077700. Figure 3.14 A shows the model run that is the least heteroscedastic, while Figure 3.14 B shows the model run that is the most heteroscedastic. These differ from those of the center basin model tests in that there is no major discernable difference between the most and least heteroscedastic linear models.

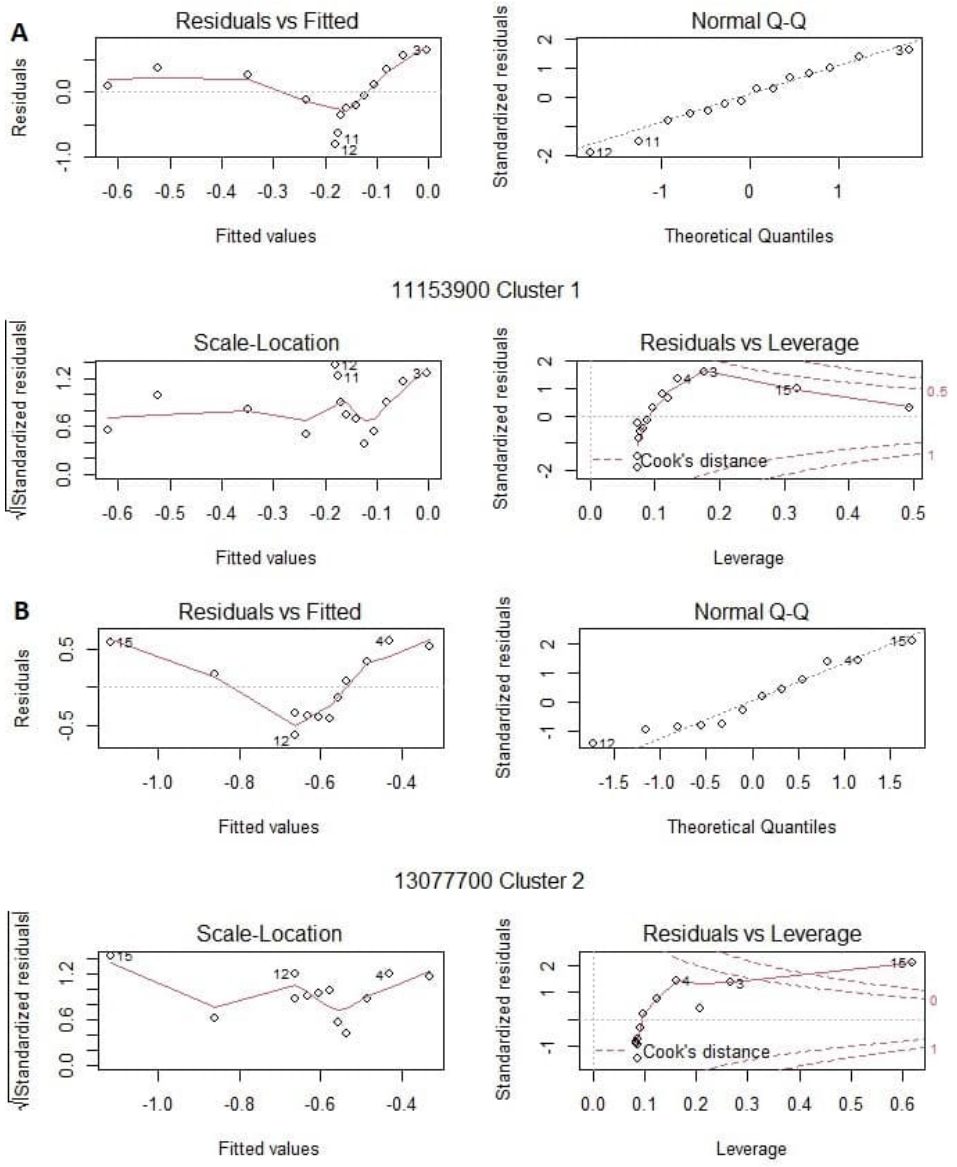

Figure 3.14: A) Plots assessing distribution of residuals for Station ID 11153900 Cluster 1 B) Plots assessing distribution of residuals for Station ID 13077700 Cluster 2.

CHAPTER 4

DISCUSSION ON RESULTS OF K MEANS CLUSTERING ANALYSIS AND HYDROLOGIC MODELING IMPLICATIONS

4.1 Number of Clusters in K-means Analysis

By requiring the number of clusters to be input as a parameter, the K-means algorithm will always produce a result, specifically data partitioned into the clusters. However, this does not necessarily mean the data fits well in those clusters nor the clusters are very different from one another. To assess the performance of the K-means clustering algorithm, it is necessary to validate the results in a separate framework from the clustering method itself. Previous studies commonly apply statistical tests to their distributions (Elsen & Tingley, 2015; Sawicz 2011, 2014), but not specific validation modeling efforts. Elsen & Tingley (2015) applied a 1-way ANOVA to assess the statistical integrity of their clusters on a geographical basis in addition to modeling an average hypothetical montane species migration for real world mountain ranges in the context of their distributions. While they use modeling in a predictive capacity, it could have also been applied in a validation effort to see if species exhibit migration patterns that support their classifications. However, Aytac (2020) argued cluster validation methods cannot be applied due to the inherent inexactness of unsupervised machine learning methods. Their purpose was to transfer results from gauged catchments to ungauged catchments, and by doing so, they note that hydrological modeling will be necessary for new catchments along with calibration and validation as those basins will have unique hydrometeorological conditions. However, they did not use any statistical or other validation techniques to assess the results of their clustered HRUs. Validation efforts can determine if the clusters hold up to testing conditions that differ from the

clustering inputs, i.e. does the hydrologic performance of these basins differ? If they hold up to these external tests, I argue that the results of the clustering analysis are robust and meaningful. Validation methods such as the one I implement are needed for unsupervised clustering methods such as K-means clustering to assess the integrity of the clustering results as a whole.

In this study, I found that a K-means clustering algorithm can be used to classify 604 reference basins from the GAGES-II dataset according to their distributions of elevation, slope, and aspect into two distinct clusters. The notion that two clusters was deemed the best number of clusters for this dataset suggests that even though there is some hydrologic difference between the clusters, the topographic differences between clusters is negligible. Figure 2.3 shows wss does not level out even with a larger number of clusters and the relationship is roughly linear. In addition, the selection criteria for two clusters resulted in several clusters with a relatively high wss around ~12000 (Figure 2.3). Even though discernable hydrograph characteristics are evident for the Cluster 1 Center and Cluster 2 Center models, these differences present could also be attributed to selecting and modeling any two basins in Clusters 1 and 2. Furthermore, this result of the K-means clustering possibly indicates that clustering techniques are perhaps best not applied to continuous variables such as topography. Both Elsen & Tingley (2015) and I discretize our topographic datasets before being used in our respective clustering algorithms. This dichotomizing of continuous variables reduces the amount of information in the dataset and thereby the statistical power to detect relationships between a variable and outcome are reduced (Altman, 2006). While the practice of discretizing continuous data should not be considered a methodological pitfall, K-means clustering is possibly best not applied to a non-categorical dataset. K-means works very well for HRU creation, as noted by Aytac (2020) due to the categorical nature of HRU inputs such as soil types, land cover types, but not as well for

continuous variables such as topography alone. K-means partitions data into classes according to the number of clusters *n* supplied as a parameter, however if the input data is continuous by nature, even discretizing it will not yield the results that would be expected from categorical data. The clustering of continuous variables such as topography is possible but is likely not best applied in such a manner.

K-means clustering methods are widely adopted and chosen for their simplicity and ease of implementation. However, one of the biggest drawbacks of K-means clustering is that the number of clusters must be supplied by the user as a parameter. While there are methods to suggest the optimal number of clusters to be used such as the Elbow method and Silhouette methods, these methods still result in a best estimate for number of clusters that the K-means algorithm will then compute. The Hartigan-Wong algorithm used for this K-means analysis does not allow for automatic determination of clusters, however several studies (Ray & Turi, 1999; Ming & Chiang, 2010) have proposed automated methods to determine the number of clusters within the K-means algorithm itself. This hard-coded number of clusters leads to another issue with K-means by essentially handicapping the algorithm. The Hartigan-Wong algorithm seeks to minimize the total intra-cluster variation by defining it as the sum of squared Euclidean distances between each basin and the cluster centroid. So, by working to minimize the Euclidean distances with a fixed number of clusters, the search can be prone to confusing local minima for the true minima (Pelleg & Moore, 2000). The algorithm could likely better optimize the results by dynamically altering the number of clusters (Pelleg & Moore, 2000), but developing and implementing a unique solution in the context of the Hartigan-Wong algorithm is beyond the scope of this project.

4.2 Hydrological Implications from K-means Results

The clustered basins shown in Figure 3.1 show that there is a large size difference between Cluster 1 Center and Cluster 2 Center (1080.38 km^2vs 32.18 km^2). The hydrograph characteristics for the baseline runs (Figure 3.9), show that given the same amount of precipitation, same CN, channel cross sectional geometries, Impervious percentage, and only a different lag time, Cluster 1 Center has a peak discharge that is an order of magnitude larger than that for Cluster 1 Center. It has been noted that the lag time is calculated based on individual reach geometries which is contingent upon the size of the underlying drainage basin. Lag time is ultimately a function of the size of the drainage basin; therefore, water moves through Cluster 2 Center much faster than Cluster 1 Center as noted by the TP and TB values (Table 3.1).

Figure 2.1 shows the wide range of basin sizes that are part of the GAGES II reference basins dataset used for clustering. Currently the GAGES II dataset has no rules regarding the size of basins within the dataset. This difference in basin size could possibly explain why the K-means analysis returned center basins that are so different in size. However, a visual interpretation of Figure 3.1 suggests the basins are distributed approximately equally across both Clusters 1 and 2 according to their size. I.e., Cluster 1 does not appear to have a disproportionate number of larger basins compared to Cluster 2 or vice versa. Basin size is important to consider because the number of pixels in each discretized elevation, aspect, and slope bin for each basin DEM was counted and normalized, so larger basins could have a disproportionate number of pixels within a certain range compared to another based on its size. Another method could be to further constrict the dataset by only selecting basins smaller or larger than a size threshold as done by Hammond (2018).

The center basins model tests established that the topographical differences between each center basin have a greater impact on the modeled hydrograph than modeling decisions do. This could be simply a characteristic of selecting any two basins in our dataset, that differing topographies yield different hydrograph signatures, or it could be a fundamental finding of the results of the clustering effort. Future work outside of the scope of this project will be needed to test if this is unique to the center basins.

4.3 Recession Rates and Fit of Linear Regression Models

It is common in catchment and surface water hydrology to attempt to explain watershed characteristics using streamflow recession analysis (Kirchner, 2009; Teuling, 2010). Specifically, these analyses make use of a power law relationship that relates discharge (Q) and (-dQ/dt) during recession events (EQ 2.6). When performing power law recession analysis that originates from this relationship (EQ 2.6), there are some noted methodological pitfalls that must be considered. Notably, Dralle (2015) noted the practice of log transforming data to generate a more linear relationship for which to fit a linear regression creates a risk of bias in the fit, which is a fundamental limitation with applying this methodology. The linear regression procedure equally weights the linear deviations from the best fit line, and after transforming the results back into linear space, these deviations diverge exponentially, biasing the fit of the model towards smaller values (Dralle, 2015). Moreover, by using a log-log scale, fluctuations are greater towards 0 than away from 0 (Kirchner, 2009). These potential biases can be seen in the regression line fits (Figure 3.8 Cluster 2) where the regression is fit to the smaller values and not the larger ones. Also noted is the horizontal scatter for the smaller values of (-dQ/dt) (Figures 3.7 and 3.13) which might be attributed to the time interval on which (-dQ/dt) is calculated (Rupp and Selker,

2006). Kirchner (2009) creates recession plots from binned averages of over 8000 individual data points from multiple years of real-world hourly observational data whereas I use strictly experimental data consisting of 1 data point per each hour of the model run, for 36 in total. This can be interpreted to suggest that these recession plots are best created from a large observational dataset as opposed to experimental/theoretical data. With only 36 data points, potential data biases if present would be amplified due to smaller amount of data points, which could explain the poor fitting of the regression line, especially in the Cluster 2 Center model. These biases points cannot be removed as they may represent random fluctuation around a greater average recession trend (Kirchner, 2009) and bias the dataset further by only removing deviations in one direction and not the other. The poor fits of the linear regression could be explained by the scarcity of data points and no removal of outlier values. In addition to the above noted issues, the relationship between discharge and recession rate is not always necessarily linear, even for modeled data. This illustrates that it is generally important to assess the fits and distributions of data before applying this type of recession analysis. These outline various reasons which might explain why a linear regression model poorly fits the data, especially for Cluster 2 Center and Cluster 1 and 2 of the Random Basin Test.

The selection criteria when selecting data points to use to create recession plots is crucial to how well a linear regression will fit. Currently, there is no universally agreed upon procedure for performing the power law recession analysis and therefore no set criteria that dictates how to select observations from the recession limb of the hydrograph (Dralle, 2017). None the less, common selection criterions generally include 1) identifying and isolating periods of recession using the hydrograph and optionally climatic datasets and 2) using these isolated periods to parameterize the power law model (Dralle, 2017). While Kirchner (2009) and Stoelzle (2012)

both applied rigorous rule-based selection criterions that follow the general form outline above, but modified to their respective studies, I considered all data points on the falling limb of the hydrograph and did not identify reasons to omit data points for any model run in either the Center Basin models nor the random basins test. There are use cases for complex and simple selection criterions with benefits and drawbacks for each. I argue a more complex criteria can be better applied to experiments using real-world observations with longer time series scales where the removal data points attributed to measurement error or extreme outliers will have less effect on the integrity of the dataset. That is not the case for these model experiments as I use a very short time series of idealized experimental data with simplified assumptions. Over short time scales with smaller datasets, unless the data is attributed to measurement error, removing outliers might be removing data that is part of some larger trend and therefore biasing regressions fit to short term data due to the lack of data points. My simplified selection approach was applied uniformly across models such that the resulting b parameter from the linear regression should be method dependent but catchment independent, meaning the same relative changes should evident across clusters and across catchments (Dralle, 2017).

Applying the linear regression assumes the residual errors to be homoscedastic, meaning variance remains constant for all modeled values implying a linear relationship between variables. In our case, the residual errors are heteroscedastic, which indicates (-dQ/dt) is not a linear function of (Q) under these idealized hydrologic conditions. Furthermore, the average R^2 values also suggest that a linear model is not best fit to the data, specifically for Cluster 2 of both tests. The recession analysis technique following that of Kirchner (2009) is a widely used approach in surface water/catchment hydrology with noted issues that must be considered. This type of recession analysis requires the use of a linear regression for which to extract regression

slope (b) and regression intercept (a) to parameterize the power law equation (EQ 2.6). However, linear regressions require normally distributed residuals (homoscedasticity) and if this is not the case, then a linear regression is not appropriately applied. These results show that the residuals are not always normally distributed, even for modeled data which is purposefully idealized. The recession analysis of Kirchner (2009) is very effective at discerning hydrologic behavior of catchments according to recession parameters if all criteria for linear regressions are met. This illustrates that such a technique requires close attention to be paid to the distribution of data before deciding to apply it. For data that is not best fit by a linear regression, other recession analysis techniques should be alternatively considered.

The results of the KS test are also interesting in that for the center basin test, TP, TB, and R^2 are shown to be statistically significant, whereas b is not. The fact that b is not statistically significant indicates that difference in regression slopes is not attributed to being representative of different distributions. For these model runs, the shape of the hydrographs and associated timing of TP and TB perform better at quantifying the differences between Cluster 1 Center and Cluster 2 Center than the results from the linear regression. For the random basin models, the null hypothesis was rejected for all four summary statistics, indicating no statistical difference between the randomly sampled basins. While b was not determined to be statistically significant for these specific basin comparisons, the application of statistical significance and recession analysis is still highly useful for discerning which hydrograph characteristics could differ between basins, particularly when using real-world data.

4.4 Conclusion

Using a K-means clustering approach to assess topographic similarity between 604 catchments in the western United States resulted in a dataset with 240 basins in Cluster 1, and 364 basins in Cluster 2. Of these clustered basins, it was shown that the basins that are closest to the center of each cluster (Cluster 1 Center and Cluster 2 Center) have hydrograph characteristics that are attributed to topographic differences between the basins and not solely due to modeling decisions. The summary statistics: TP, TB, and R^2 are shown to be statistically significant from the KS test suggesting a difference between the two basins according to those measures. Though there is a statistical difference between basins closest to the cluster centers, when randomly selecting basins from each cluster and modeling their hydrograph characteristics, the same differences are not observed. All four summary statistics used to quantify the differences in between these model runs were not determined to be statistically significant ($p > 0.05$), indicating the basins sampled are likely from the same distribution. This also illustrates that the clustering methods were not successful in partitioning basins into clusters based on similar topographic characteristics. Small hydrological differences are noticed when modeling the center basins, but these differences could be the same for any two basins. No hydrological differences are discernable for modeling randomly selected basins from each cluster.

The goal of this project was to partition catchments based on topographic differences using K-means clustering methods, and this was achieved. But a hydrological modeling validation effort indicated that there is no meaningful difference between these partitioned catchments. Topography is a continuous dataset with too much variability to be discretized and classified in a meaningful way using unsupervised machine learning methods. K-means is best applied to categorical datasets and has utility in hydrology such as creation of HRUs. The

clustering analyses of Sawicz (2011, 2014) and Aytac (2020) were based on large topographic, hydroclimatic, and/or hydrologic datasets which used complex data refinement criterions along with clustering models developed with advanced decision trees based on expert judgement. Elsen & Tingley (2015) used a topographic dataset alone similar to mine and applied an ANOVA to assess geographical classification integrity but did not otherwise validate their classifications. While statistical tests can assess the statistical integrity of the clusters, an outside validation method such as I have demonstrated needs to be applied to assess the meaningfulness of classifications in a separate context from the classification inputs. The use of K-means clustering as applied through the scope of this project sheds light on the broader application of unsupervised clustering methods in research. While unsupervised learning methods, specifically K-means are simple to set up and do not often require extensive computational resources, we should be cautious of the unvalidated classification outputs. Statistical tests are useful to determine statistical differences between classifications, but classifications can still be statistically different but not necessarily be meaningful. Validation methods such as modeling classification performance to against outside forcing parameters are needed to ensure the results of the clustering effort are meaningful and discernable. Unsupervised clustering methods are very popular research tools and have demonstrated utility in the field of hydrology notably for categorical data such as creation of HRUs, however, it is crucial to understand how these methods work and their respective pitfalls. Furthermore, clustering outputs should not be trusted until validation methods can be applied. Specifically for K-means clustering, I argue it is good practice to validate the clustering outcome by testing these distributions using a separate index as demonstrated by constructing hydrological models of the clustered basins. By performing

unsupervised clustering analyses in this manner, it allows the user to discern whether the clustering results are meaningful or not.

CHAPTER 5

FUTURE WORK

This work identified that a K-means clustering algorithm can be applied to topographic indices to yield discrete clusters. The initial selection criteria consisted of only selecting reference basins part of the GAGES-II dataset that are in the western United States, which resulted in 604 catchments to be used for analysis. While this number of catchments sufficed, it would be interesting to see what the distributions of elevation, slope, and aspect look like for additional catchments. One area of work could consist of analyzing references basins for the entire GAGES-II dataset for the United States. Furthermore, it was identified that basin area could possibly have been a factor to consider deeper as noted by the size difference of the basins closest to the cluster centers. Further work should select basins above or below a size threshold to see if the clustering results would be different by only considering basins of roughly the same area. Also, only 8 basins were selected at random to be modeled from each cluster due to time constraints with a GUI-based model. Future work could consist of constructing hydrological models for all 604 basins in the clustered dataset to compare the resulting hydrograph characteristics.

The specific variable combination of elevation, slope, and aspect were shown to not yield meaningful topographic-based classifications of the basins within the GAGES-II dataset. Perhaps meaningful classifications of topography would be yielded if other variables were additionally considered in the K-Means algorithm. This study shows that catchment topography is too variable to be simplified down to only distributions of elevation, slope, and aspect for each catchment. Further classification studies could also include additional topographic variables including contributing upstream area for streams, basin size, drainage area, topographic wetness

index, or convexity. Including more topographic variables could likely recapture some of the natural topographic variability that was not included in this clustering dataset. In addition, perhaps including other variables needs to be considered such as geologic variables or soil types/soil thicknesses as well.

An additional possible area for further study would be to see if the modeling trends identified for the basins closest to the centers of each cluster are indicative of the findings of the K-means analysis or if they would be the same for selecting any two basins in the dataset. The KS tests established that the statistics being used to describe the difference between Cluster 1 Center and Cluster 2 Center are statistically significant. However, it is unknown if this difference in basins is attributed to the K-means performance or if it is simply an observation of selecting any two basins. To test this, the same approach could be applied to randomly select one basin in each cluster and compare the hydrographs and summary statistics.

Although I have noted the issue with K-means of having to prespecify the number of clusters for the algorithm, studies have been proposed modifications to K-means algorithms where the number of clusters can be varied within the algorithm (Ray & Turi, 1999, Ming & Chiang, 2010). These proposed methods have yet to be applied in a hydrological context and could yield fascinating results if the number of clusters is variable. Perhaps the basins will better fit in their respective cluster while also being partitioned into more than 2 clusters.

REFERENCES

Addor, N. A Ranking of Hydrological Signatures Based on Their Predictability in Space. 8792–8812 (2018) doi:10.1029/2018WR022606.

Alboukadel Kassambara and Fabian Mundt (2020). factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7. https://CRAN.R-project.org/package=factoextra

Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ. 2006 May 6;332(7549):1080. doi: 10.1136/bmj.332.7549.1080. PMID: 16675816; PMCID: PMC1458573.

Aytaç, E. Unsupervised learning approach in defining the similarity of catchments: Hydrological response unit based k-means clustering, a demonstration on Western Black Sea Region of Turkey. *Int. Soil Water Conserv. Res.* **8**, 321–331 (2020).

Bao, Z., Zhang, J., Liu, J., Wang, G., Yan, X., Wang, X. and Zhang, L. (2012), Sensitivity of hydrological variables to climate change in the Haihe River basin, China. Hydrol. Process., 26: 2294-2306. https://doi.org/10.1002/hyp.8348

Bholowalia, Purnima, and Arvind Kumar. "EBK-means: A clustering technique based on elbow method and k-means in WSN." International Journal of Computer Applications 105.9 (2014).

Boussinesq, J.: Essai sur la théorie des eaux courantes, vol. 2, Im- primerie nationale, Paris, France, 1877

Carroll, R. W. H., Deems, J. S., Niswonger, R., Schumer, R. & Williams, K. H. The Importance of Interflow to Groundwater Recharge in a Snowmelt-Dominated Headwater Basin. *Geophys. Res. Lett.* **46**, 5899–5908 (2019).

Corduas, M. Clustering streamflow time series for regional classification. *J. Hydrol.* **407**, 73–80 (2011).

Devito, K. *et al.* INVITED COMMENTARY A framework for broad-scale classification of hydrologic response units on the Boreal Plain : is topography the last thing to consider ? **1714**, 1705–1714 (2005).

Dralle, D., Karst, N. & Thompson, S. E. a , b careful : The challenge of scale invariance for comparative analyses in power law models of the streamflow recession. 9285–9293 (2015) doi:10.1002/2015GL066007.Received.

Dralle, D. N., Karst, N. J., Charalampous, K., Veenstra, A. & Thompson, S. E. Event-scale power law recession analysis : quantifying methodological uncertainty. 65–81 (2017) doi:10.5194/hess-21-65-2017.

Elsen, P. R. & Tingley, M. W. Global mountain topography and the fate of montane species under climate change. *Nat. Clim. Chang.* **5**, 772–776 (2015).

Falcone, J. A. (2011). GAGES-II: Geospatial attributes of gages for evaluating streamflow (Digit. Spat. Data set). Reston, VA: U.S. Geological Survey

Hartigan, J. A., and M. A. Wong. "Algorithm AS 136: A K-Means Clustering Algorithm." *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, 1979, pp. 100–08. *JSTOR*, https://doi.org/10.2307/2346830. Accessed 23 Nov. 2022.

Hammond, J. C., Saavedra, F. A. & Kampf, S. K. How Does Snow Persistence Relate to Annual Streamflow in Mountain Watersheds of the Western U.S. With Wet Maritime and Dry Continental Climates? *Water Resour. Res.* **54**, 2605–2623 (2018).

Hollister, J.W. (2021). elevatr: Access Elevation Data from Various APIs. R package version 0.4.1. https://CRAN.R-project.org/package=elevatr/

J. A. Cunge (1969) On The Subject Of A Flood Propagation Computation Method (Musklngum Method), Journal of Hydraulic Research, 7:2, 205-230, DOI: 10.1080/00221686909500264

Jehn, F. U. *et al.* Using hydrological and climatic catchment clusters to explore drivers of catchment behavior. 1081–1100 (2020).

J. MacQueen (1967). Some methods for classification and analysis of multivariate observations. Proc. Fifth Berkeley Symp. on Math. Statist. and Prob., Vol. 1 (Univ. of Calif. Press, 1967), 281—297

Kirchner, J. W. Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward. *Water Resour. Res.* **45**, 1–34 (2009).

Klos, P. Z., Link, T. E., and Abatzoglou, J. T. (2014), Extent of the rain-snow transition zone in the western U.S. under historic and projected climate, Geophys. Res. Lett., 41, 4560– 4568, doi:10.1002/2014GL060500.

Knowles, N., M. D. Dettinger, and D. R. Cayan (2006), Trends in snowfall versus rainfall in the western United States, J. Clim., 19(18), 4545– 4559.

Köplin, N., Schädler, B., Viviroli, D., and Weingartner, R.: Relating climate change signals and physiographic catchment properties to clustered hydrological response types, Hydrol. Earth Syst. Sci., 16, 2267–2283, https://doi.org/10.5194/hess-16-2267-2012, 2012.

Kroll, C. N., Croteau, K. E. & Vogel, R. M. Hypothesis tests for hydrologic alteration. J. Hydrol. 530, 117–126 (2015).

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K.(2021). cluster: Cluster Analysis Basics

and Extensions. R package version 2.1.2.

Ming, M. & Chiang, T. Intelligent Choice of the Number of Clusters in K-Means Clustering: An Experimental Study with Different Cluster Spreads. *J. Classif.* **190**, 173–190 (2010).

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

Pelleg, Dan, and Andrew W. Moore. "X-means: Extending k-means with efficient estimation of the number of clusters." *Icml.* Vol. 1. 2000.

Prancevic, J. P. & Kirchner, J. W. Topographic Controls on the Extension and Retraction of Flowing Streams. *Geophys. Res. Lett.* **46**, 2084–2092 (2019).

Ray, S. & Turi, R. H. Determination of number of clusters in k-means clustering and application in colour image segmentation. *Proc. 4th Int. Conf. Adv. pattern Recognit. Digit. Tech.* 137–143 (1999).

Rupp D. and Selker J. , "Information, artifacts, and noise in dQ/dt-Q recession analysis", Adv. Water Resour., Vol. 29, No.2, (2006), pp 154-160

SAPUTRA, D. M., SAPUTRA, D. & OSWARI, L. D. Effect of Distance Metrics in Determining K-Value in K-Means Clustering Using Elbow and Silhouette Method. 172, 341–346 (2020).

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A. & Carrillo, G. Catchment classification: Empirical analysis of hydrologic similarity based on catchment function in the eastern USA. *Hydrol. Earth Syst. Sci.* **15**, 2895–2911 (2011).

Sawicz, K. A. *et al.* Characterizing hydrologic change through catchment classification. *Hydrol. Earth Syst. Sci.* **18**, 273–285 (2014).

Stein, L., Clark, M. P., Knoben, W. J. M., Pianosi, F. & Woods, R. A. How Do Climate and Catchment Attributes Influence Flood Generating Processes? A Large-Sample Study for 671 Catchments Across the Contiguous USA. Water Resour. Res. 57, 1–21 (2021).

Stewart, I. T., Cayan, D. R. & Dettinger, M. D. Changes toward earlier streamflow timing across western North America. *J. Clim.* **18**, 1136–1155 (2005).

STOELZLE, MICHAEL, KERSTIN STAHL, and MARKUS WEILER. "As simple as possible? Drought recognition based on streamflow recession." geology 50 (2012): Q95.

Taylor, R. G., Scanlon, B., Döll, P., Rodell, M., van Beek, R., Wada, Y., et al. (2012). Ground water and climate change. Nature Climate Change, 3(4), 322–329. http://doi.org/10.1038/nclimate1744

Teuling, A. J., Lehner, I., Kirchner, J. W. & Seneviratne, S. I. Catchments as simple dynamical systems : Experience from a Swiss prealpine catchment. **46**, 1–15 (2010).

US Army Corps of Engineers, USACE (1998) HEC-1 flood hydrograph package user's manual. Hydrologic Engineering Center, Davis, CA.

USACE (2000) Hydrologic Modeling System HEC-HMS User's Manual. Hydrologic Engineering Center, Davis, CA.

USACE (2002) Hydrologic Modeling System HEC-HMS Applications Guide. Hydrologic Engineering Center, Davis, CA.CHAPTER 2

Viviroli, D., Dürr, H. H., Messerli, B., Meybeck, M. & Weingartner, R. Mountains of the world, water towers for humanity: Typology, mapping, and global significance. *Water Resour. Res.* **43**, 1–13 (2007).

Wagener, T., Sivapalan, M., Troch, P. & Woods, R. Catchment Classification and Hydrologic Similarity. *Geogr. Compass* **1**, 901–931 (2007).