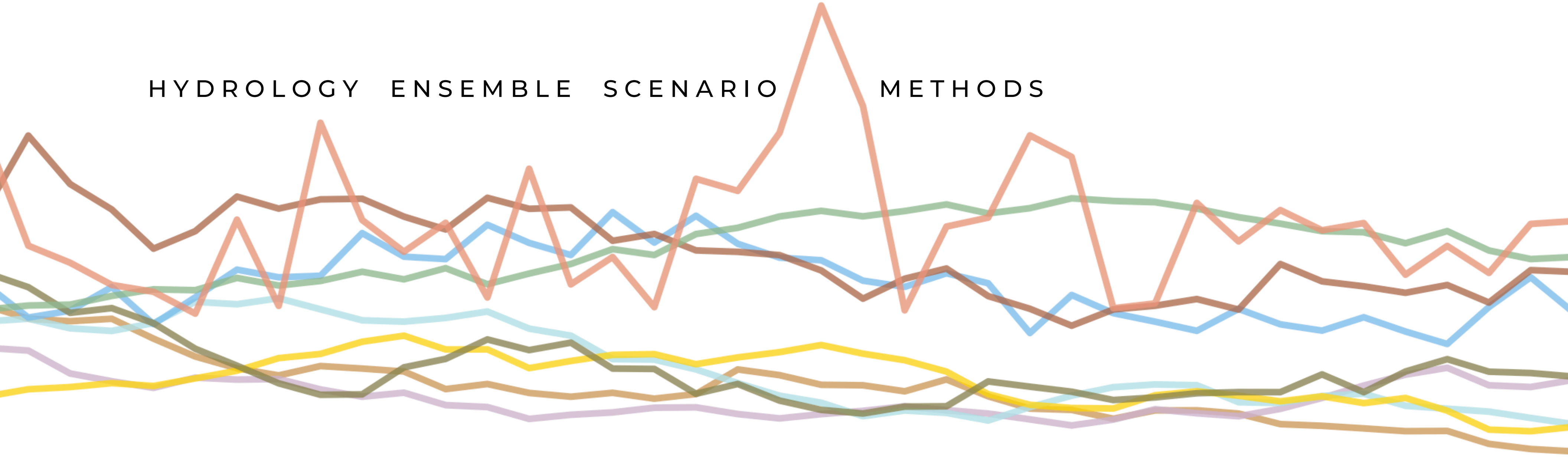
HYDROLOGY ENSEMBLE SCENARIO METHODS



INTRODUCTION

Virga Labs developed a technical methodology to generate the Colorado River Simulation System (CRSS) hydrology ensemble scenarios[1] presented in RiverViz. This document outlines the following segmentation analysis steps used to evaluate different hydrology ensembles:

1. Data Collection and Aggregation
2. Modeling
3. Post Modeling Analysis and Hydrology Ensemble Selection

1) DATA COLLECTION & AGGREGATION

A. Collect Source Hydrology Trace Data
B. Apply Calculations and Summary Statistics
C. Format Cohesive Data Set

A. Collect Source Hydrology Trace Data

This analysis incorporated Lees Ferry flow from the following historic records, statistical models and physically based hydrology models that are available as hydrology ensembles for use in CRSS:

- Observed Resampled:
  - Uses the Index Sequential Method (ISM) on the observed Lee Ferry flow records from 1906 to Present.
  - Implies that future trends will be similar to those seen in the past.
  - Developed and maintained by the Bureau of Reclamation.
  - “ObsRes” is shortened in CRSS to represent Observed Resampled, and is titled “Full Hydrology (1906 - Present)” in RiverVIz.
- General Circulation Models:
  - Uses an ensemble of future general circulation models projections.
  - Implies that future trends will be impacted by climate change.

[1] A hydrology ensemble is defined as a set of hydrology traces that are grouped because they were generated using a particular method. Hydrology ensembles can be interpreted as, and are referred to in the RiverViz tool as, “scenarios”, e.g. “Gradual Aridity Increase”.

    - Adapted for use in CRSS by the Bureau of Reclamation
    - “GCM” is the acronym in CRSS for General Circulation Model, and is titled “General Climate Model (CMIP-3)” in RiverViz.
- Hot Drought RCP 45 100:
    - Uses the Representative Concentration Pathway (RCP) 4.5 warming factor (stabilization scenario) and a 10% streamflow reduction coefficient per degree temperature rise.
    - Implies that future trends will be impacted by increasing emissions and resulting increasing temperatures.
    - Developed by Brad Udall, Senior Water and Climate Scientist at Colorado State University
    - “Hot Drought RCP 45 100” is the descriptor used by Virga Labs, and is titled “Hot Drought (RCP 4.5)” in RiverViz.
- Hot Drought RCP 85 100:
    - Uses the Representative Concentration Pathway (RCP) 8.5 warming factor (business as usual) and a 10% streamflow reduction coefficient per degree temperature rise.
    - Implies that future trends will be impacted by increasing emissions and resulting increasing temperatures.
    - Developed by Brad Udall, Senior Water and Climate Scientist at Colorado State University
    - “Hot Drought RCP 85 100” is the descriptor used by Virga Labs, and is titled “Hot Drought (RCP 8.5)” in RiverViz.

These data sets were selected to represent a reasonable range of plausible hydrologic futures in the Basin, including traces that generate lower flows than are included in the current official Stress Test hydrology ensemble. For each underlying hydrologic record the variables used are Lee Ferry annual flow in MAF, trace id, and year. In total, there are 443 traces which are broken down as follows:

- Observed Resampled
    - 107 individual traces
    - Projected time frame from 2021 to 2060
- General Climate Model
    - 112 individual traces
    - Projected time frame from 2021 to 2060
- Hot Drought 45 100
    - 112 individual traces
    - Projected time frame from 2018 to 2100
- Hot Drought 85 100
    - 112 individual traces

- Projected time frame from 2018 to 2100

The average flow (MAF) for the traces within each data set is shown in Figure 1. As one can see, the Hot Drought RCP 45 100 and 85 100 traces are lower on average than ObsRes and GCM. Additionally, the flat line behavior of the ObsRes average is due to the ISM used to generate the hydrology traces. The data was tabulated such that we could engineer summary statistics, or features, to apply modeling techniques.

Figure 1: Average annual Lee Ferry flows by data set between 2020 and 2060, the time period considered in this modeling and analysis. This figure illustrates the differences in average trend characteristics between data sets.

B. Apply Calculations and Summary Statistics

A clustering algorithm was applied to summary statistics, or features, of the data to capture characteristics that the model can use to determine similarity between traces based on those characteristics. The characteristics selected were consistent with previous work to establish a framework for robust planning in the Colorado River Basin led by the Kasprzyk Research Group and Nathan Bonham, research affiliate of the Bureau of Reclamation. Each trace was summarized based on these methods so we could compare across traces on the same summary statistics. The following sections detail the calculations for each of the summary methods.

Rolling Window Deficit Time Series

The first method used to calculate summary statistics is the rolling window deficit or surplus. First, we considered a baseline flow, for example 13 MAF. These baselines were selected based on the previous work of Bonham and historical averages. From that baseline flow, we compared each trace's annual Lee Ferry flow to the baseline by subtracting the Lee Ferry flow from the baseline flow. In this case, a positive value represents a deficit and negative values represent a surplus. Then, we considered a time window (in years) and within that time window, we calculated a rolling surplus or deficit. For example, if we used a 5 year time window, for each trace, we add up the surplus or deficit values to get a total 5 year surplus/deficit value. This means that for each year we looked at the following 5 years to get a rolling window surplus or deficit value.

The values used as baseline flows (MAF) are as follows: 9 | 11 | 13 | 15

The values used as time windows (years) are as follows: 1 | 2 | 5 | 8 | 10 | 20

For each trace, we calculated the rolling deficit or surplus for each combination of baseline flow and time window. Then for each trace, we calculated:

- maximum value of the surplus or deficit rolling window values for each year and baseline flow
- minimum, value of of the surplus or deficit rolling window values for each year and baseline flow
- average of the surplus or deficit rolling window values for each year and baseline flow
- range of the surplus or deficit rolling window values for each year and baseline flow
- window change count or number of times the rolling windows changed from surplus to deficit or vice versa for each year and baseline flow

Frequency and Duration Statistics

We then calculated a variety of frequency and duration statistics for each trace by categorizing each year for each trace as wet, dry, or normal compared to the entire trace. There were three methods used to determine the threshold values for what constitutes wet, dry, or normal.

1. Mean and standard deviation: For each trace we calculated the mean and standard deviation of Lee Ferry flow. A year is considered wet if its Lee Ferry flow is one standard deviation above average or more, and dry if its Lee Ferry flow is one standard deviation below average or less. Anything in between is considered normal.

2. Terciles: For each trace, we calculated the value at which 1/3 of the Lee Ferry flows are below that value. This served as the threshold below which is considered dry. Then we calculated the value at which 1/3 of the Lee Ferry flows were above that value which served as the threshold above which is considered wet. Each year was compared against these thresholds and was categorized appropriately as wet or dry.
3. Historical periods to determine when a flow is considered wet or dry: The average Lee Ferry flow from 2000 to 2018 was the threshold below which trace years were considered dry and the average Lee Ferry flow from the pluvial period 1906 to 1931 was the threshold above which trace years were considered wet. Each year for each trace was compared against these thresholds and was categorized as wet or dry appropriately.

Each year in each trace received a wet, non wet, dry, or normal categorization for each of these methodologies. Then for each trace and each method, we calculated the count of:

- consecutive dry years
- consecutive normal years
- consecutive wet years
- consecutive non wet years (i.e., dry OR normal)
- total dry years
- total normal years
- total wet years
- total non wet years

C. Format Cohesive Data Set

To format a data set on which to perform unsupervised2 learning, we combined both sets of calculations into a single data frame where each trace was represented once by its trace ID and had all of the above mentioned summary statistics. The data frame consisted of 145 variables, including trace ID. The variables corresponded to a combination of factors discussed earlier and summarized in Tables 1 and 2.

---

[2] An unsupervised machine learning algorithm is a machine learning model which is trained on unlabeled data in order to solve clustering, similarity, or association problems.

Summary Statistic	Time Window	Baseline Flow
max	1	15
min	2	13
average	5	11
count	8	9
range	10	
	20	

Table 1. Rolling Window Deficit Time Series table with the 5 Summary Statistics x 6 Time Windows x 4 Baseline Flows, or 120 variables total.

Pattern	Category	Method
consecutive	dry	1
total	normal	2
	wet	3
	non wet	

Table 2. Frequency and Duration Statistics table illustrating the 2 Patterns x 4 Categories x 3 Methods, or 24 variables total.

2) MODELING

Two modeling techniques were applied in sequence. First, we applied K-means clustering to identify traces that shared similar characteristics. Once these first clusters were identified, we applied dynamic time warping clustering within each of the clusters. The latter groups the traces based on similar temporal trends. These techniques are detailed below.

Steps:

A. Unsupervised Learning Method: K-Means
B. Applying Dynamic Time Warping to K-Means Clusters
C. Visualization

A. Unsupervised Learning Method: K-Means

First, we applied the unsupervised learning method of K-means clustering. Unsupervised indicates that we had no target hydrology ensembles and the algorithm found natural clusters in the data based on all the summary features previously calculated.

Model Training

Before beginning we chose how many clusters the algorithm should identify. With this algorithm, more clusters leads to less variance within each cluster. However, there is a point at which returns diminish, meaning there is a point at which increasing the number of clusters reduces the variance, but at a decreasing rate. One way to identify this dynamic is with an elbow graph (Figure 2). We plotted the sum of the square error (SSE) against the number of clusters and found that while the SSE always declined with more clusters, returns diminished after 2 clusters (i.e., the SSE decreases at a far less steep rate after 2). For this reason, we selected two clusters for the K-Means model training.

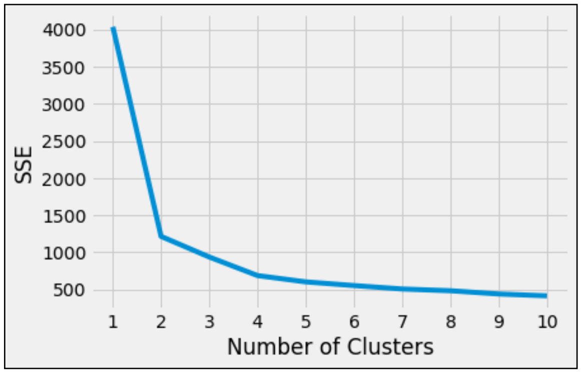


Figure 2: Elbow plot illustrating how the SSE decreases as the number of clusters increases. Note the change in slope at 2 clusters. This indicates that we continue to reduce error as we increase the number of clusters, but at a decreasing rate.

After training the model to cluster the traces into two classes based on the summary features, the model created two classes:

- Class 0:
  - 109 GCM traces
  - 107 ObsRes traces
- Class 1:
  - 112 Udall Hot Drought 45 100
  - 112 Udall Hot Drought 85 100
  - 3 GCM traces

After associating the cluster with each trace we added the centroids of each cluster to evaluate the distance between each trace and the centroid.

Cluster Analysis and Visualization

In order to analyze and visualize the K-Means results, we examined the time series traces that were most closely aligned with the centroid (i.e. nearest to centroid). Figure 3 and 4 each display 10 traces that are closest to their respective centroid based on Euclidean distance, the default distance for this model.

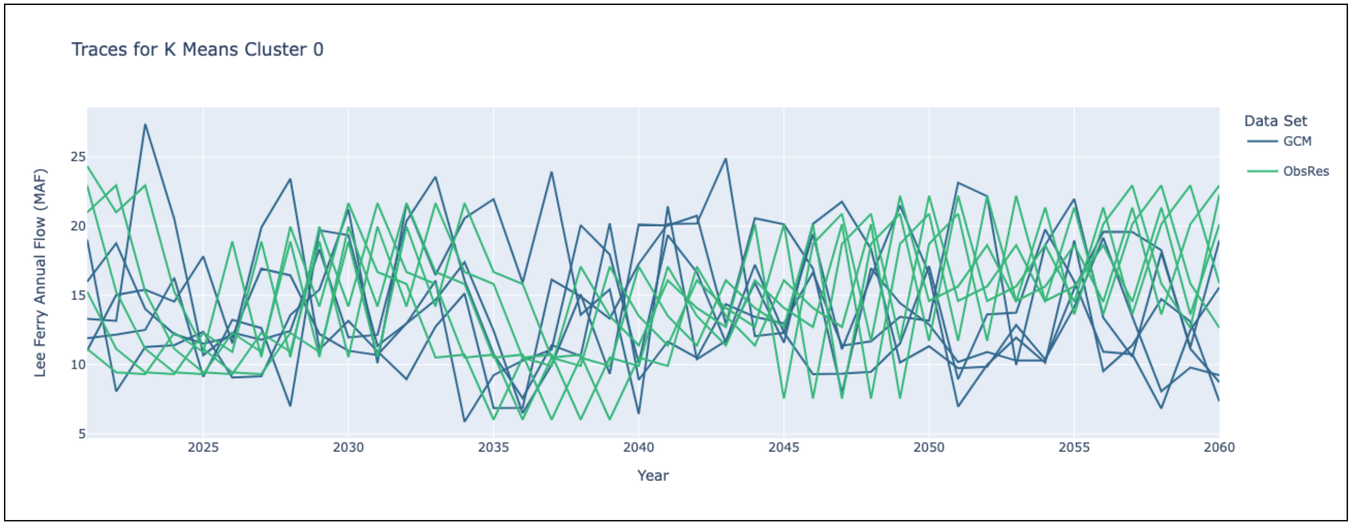


Figure 3: The 10 traces closest to the center of K-Means Class 0 colored by source hydrology ensemble.

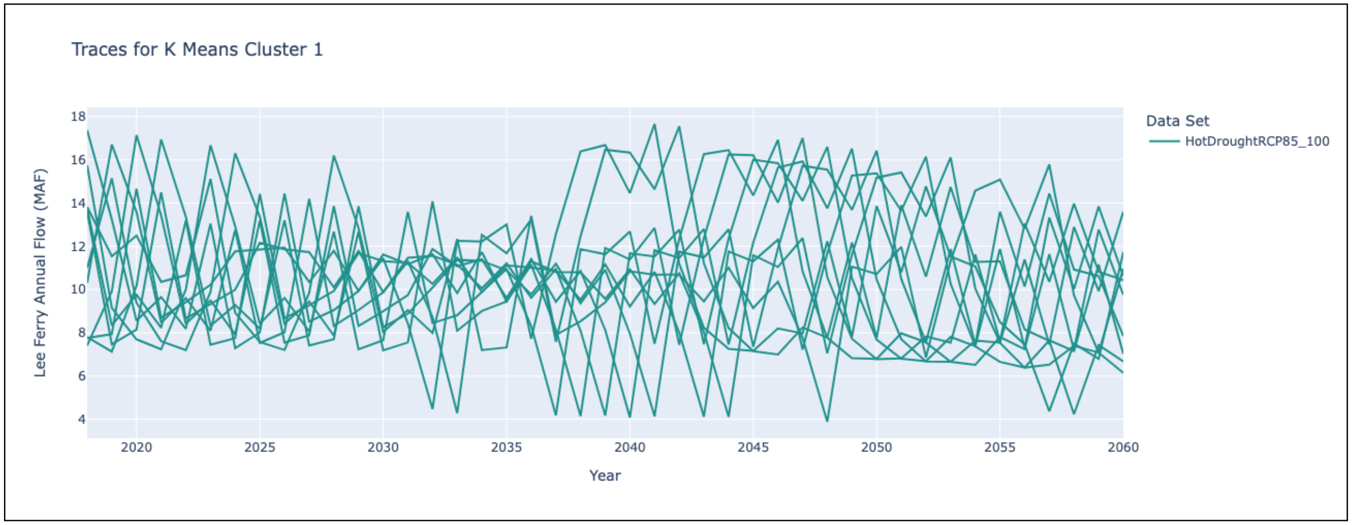


Figure 4: The 10 traces closest to the center of K-Means Class 1 colored by source hydrology ensemble.

We then generated histograms to assess the distance to the centroid for each cluster, or how closely each cluster matched its central tendency. In general, distributions that clump closer to distance 0 (left hand side of the x-axis) indicate that the data within the cluster has less variance (Figure 5 and Figure 6). As illustrated by the histograms, there are very few traces extremely close to the centroid, but most traces are tightly clumped and the data is highly modal. In addition, the scale for both clusters is similar.

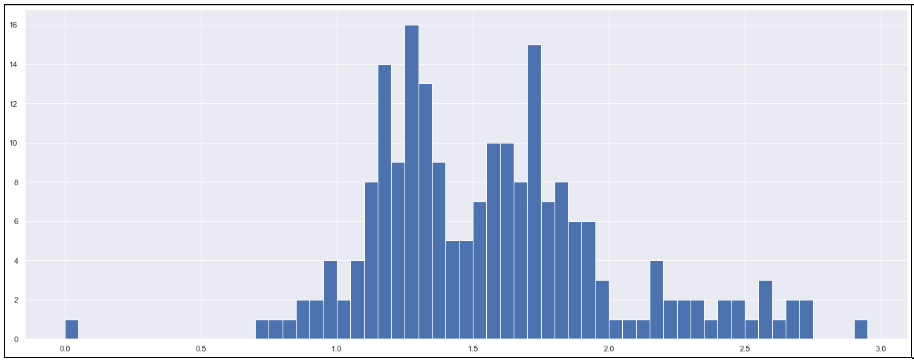


Figure 5: Histogram of distance to the centroid for Cluster 0.

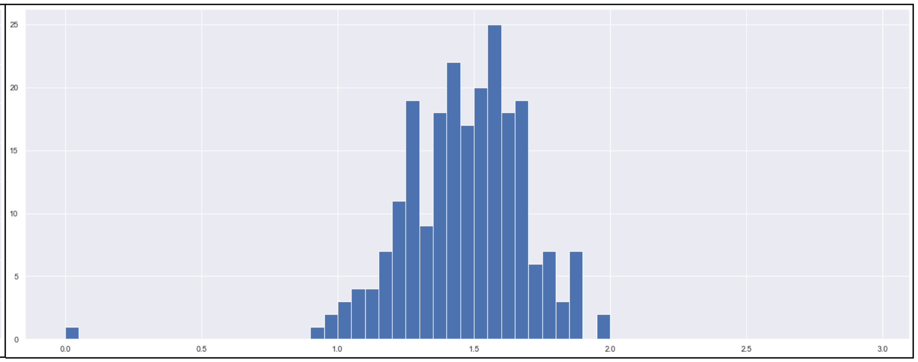


Figure 6: Histogram of distance to the centroid for Cluster 1.

Average Trace Behavior

Next we analyzed the average trace behavior for each cluster. As seen in Figure 7, Class 0 has higher Lee Ferry flows on average and Cluster 1 has much lower Lee Ferry flows on average. If we examine the source hydrology ensembles that comprise each of the clusters, class 0 contains 109 GCM traces and 107 ObsRes traces while class 1 contains 112 Udall Hot Drought 45 100, 112 Udall Hot Drought 85 100, and 3 GCM traces.

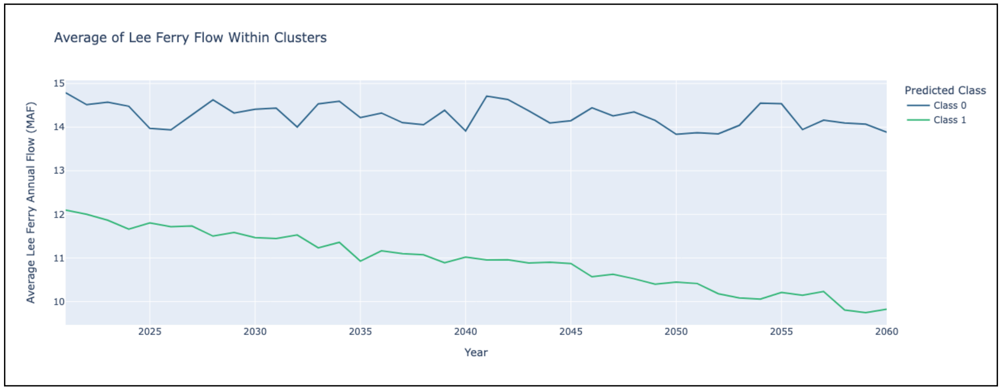


Figure 7: Annual average Lee Ferry flow by predicted class.

Cluster Separation

Next we examine how different the two clusters are to determine if the model did an acceptable job separating the data. To accomplish this, we examined the features and selected the ones that were highly correlated with the clusters but not with each other. Then we plotted them against one another in a scatter plot and colored them by cluster (Figure 8). In the plot, we were looking for distinct groups and we found that in most cases, the points separated well. This indicates that the K-Means clustering did an adequate job separating the data into unique groups.

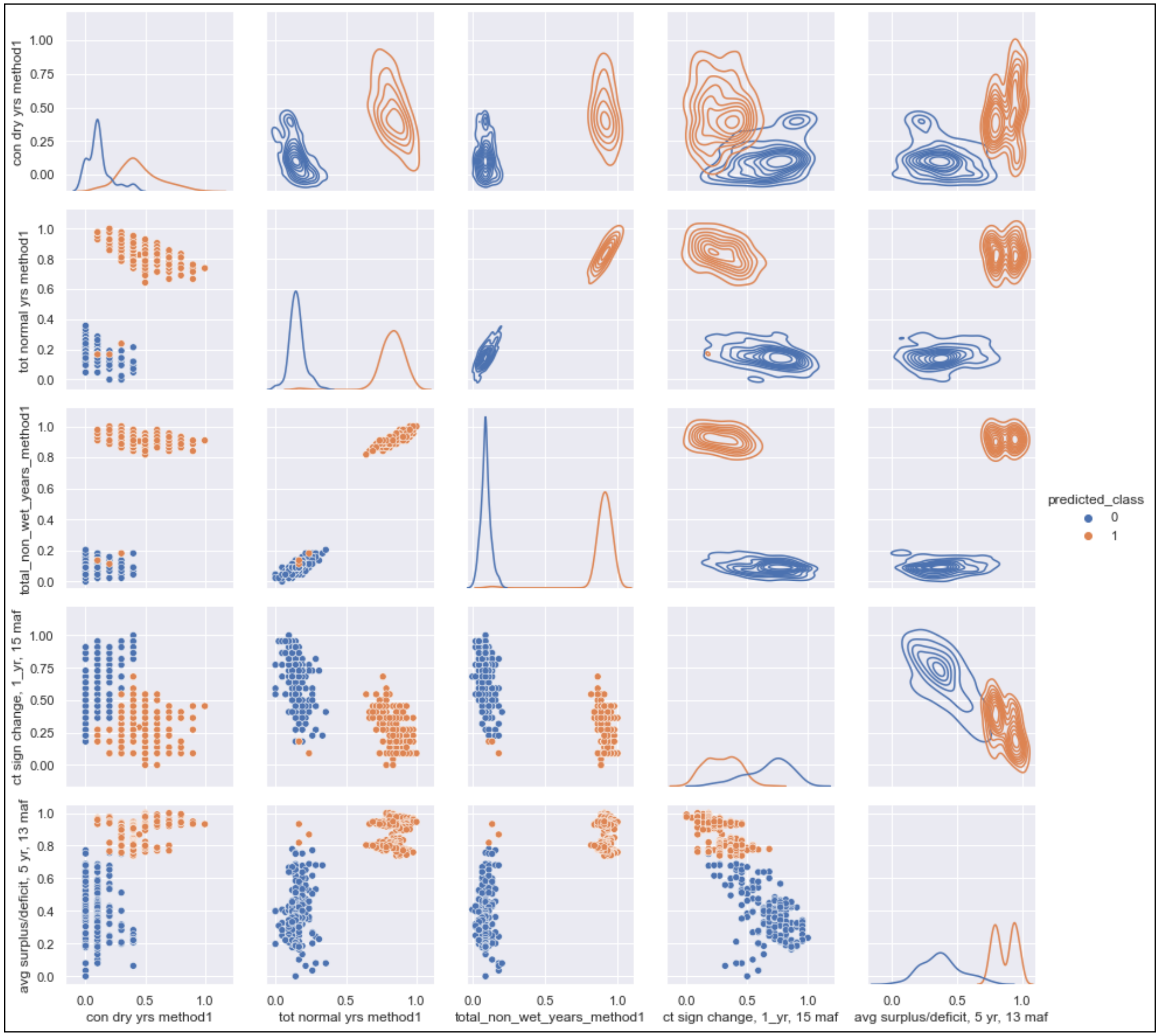


Figure 8: The separation between the two clusters based on four of the input variables. Each plot combines two variables at a time plotted against one another, either as a scatter plot colored by class label in the cells below the diagonal, or as distribution maps in the upper diagonal.

B. Applying Dynamic Time Warping to K-Means Clusters

Now that we have the first pass of clustering done in which we grouped traces by feature or characteristic, we now apply another unsupervised learning algorithm to further sort them by temporal features. The intention with this approach is to create groups of traces that follow similar trends so that when they are averaged or looked at together, they tell a cohesive story. This algorithm is called Dynamic Time Warping (DTW). In order to try to create traces that match the hydrology ensembles and also have similar temporal features, we apply DTW to the results of the K-means clusters.

Model Training

To train this model, we separated the data by K-means cluster. Within each cluster, we ran the unsupervised learning algorithm with a target of 10 Dynamic Time Warping clusters each in order to allow the algorithm to sufficiently separate temporal patterns. Therefore in total, there were 20 clusters into which a trace could fall. In each cluster we added the new center called the barycenter and calculated the Euclidean distance from each trace to the barycenter. At this point, model training was complete.

C. Visualization

We visualized the 20 clusters that resulted from the above modeling by displaying all the traces for each cluster. Please see figures 9 - 18 for Cluster 0 and figures 19 - 28 for Cluster 1.

All DTW Trace Clusters for K-Means Cluster 0

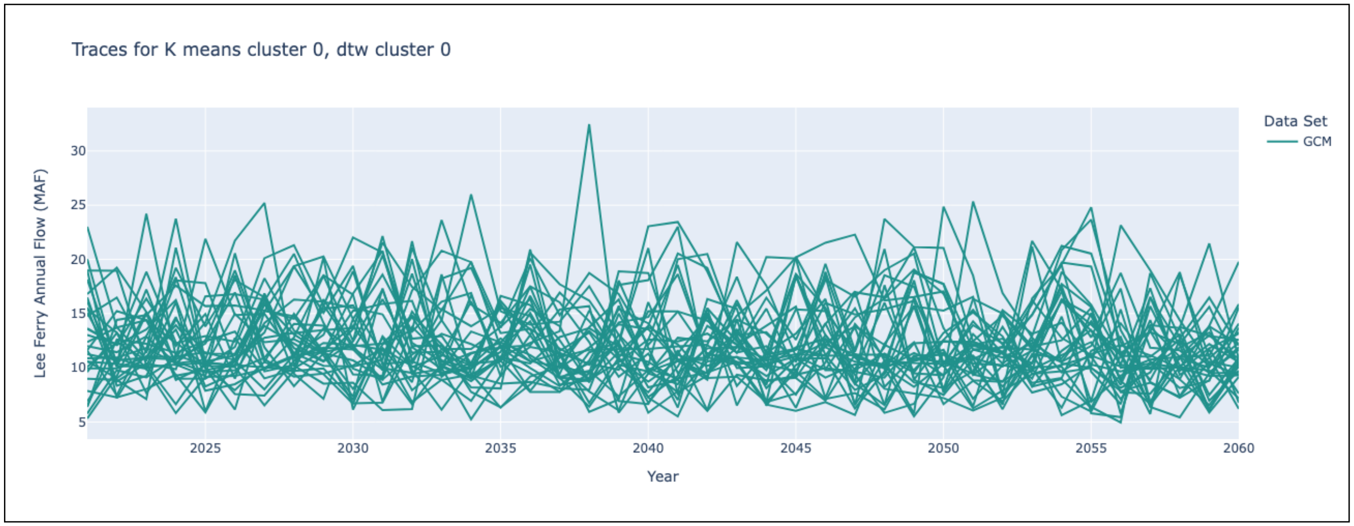


Figure 9: All traces associated with the k means group 0 and dynamic time warping group 0. This represents a cluster of traces that the algorithms placed with one another based on their rolling window/ frequency and duration characteristics and temporal features.

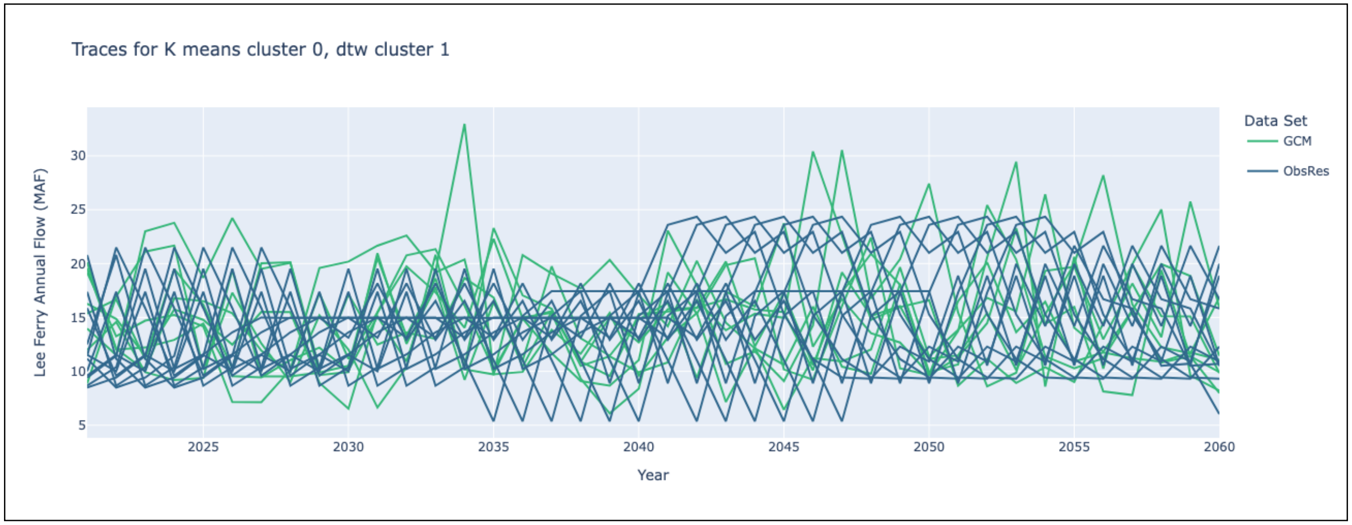


Figure 10: All traces associated with the k means group 0 and dynamic time warping group 1. This represents a cluster of traces that the algorithms placed with one another based on their rolling window/ frequency and duration characteristics and temporal features.

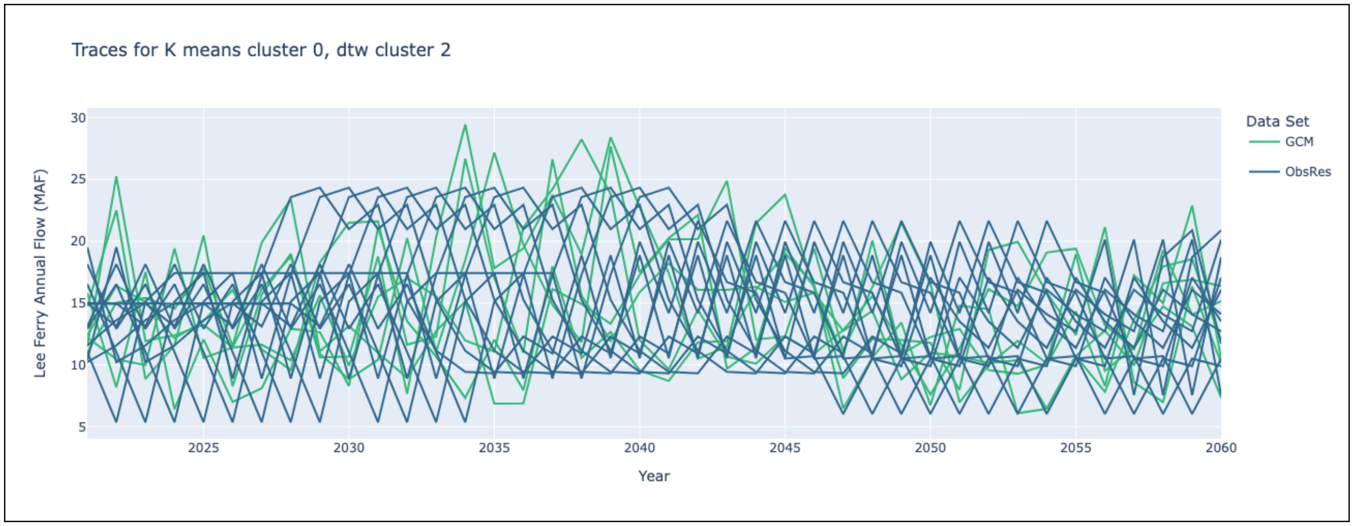


Figure 11: All traces associated with the k means group 0 and dynamic time warping group 2. This represents a cluster of traces that the algorithms placed with one another based on their rolling window/ frequency and duration characteristics and temporal features.

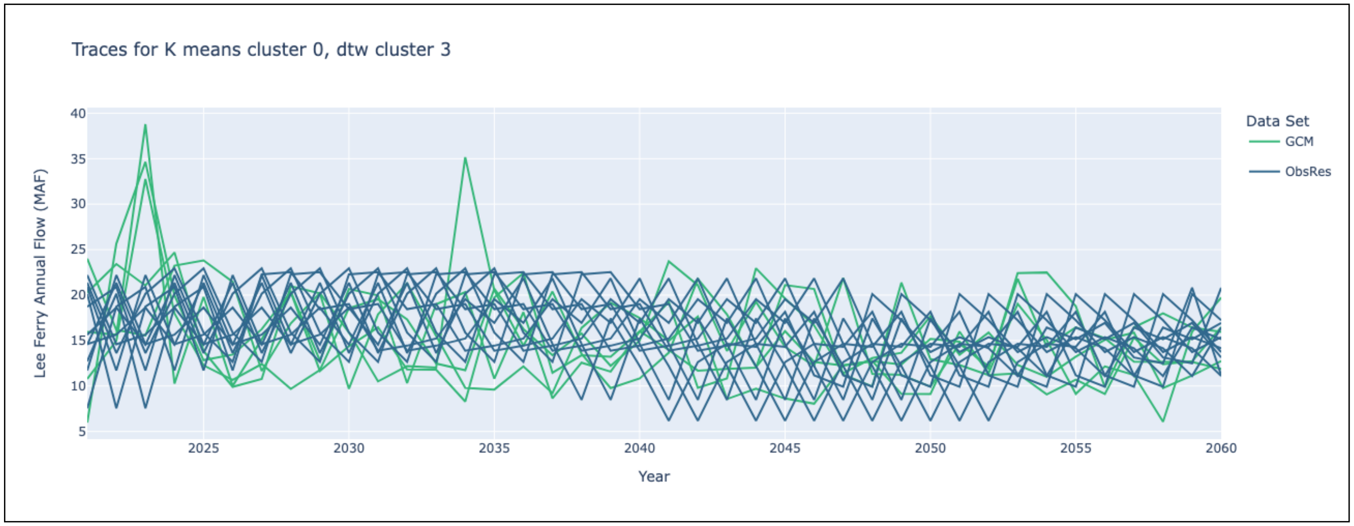


Figure 12: All traces associated with the k means group 0 and dynamic time warping group 3. This represents a cluster of traces that the algorithms placed with one another based on their rolling window/ frequency and duration characteristics and temporal features.

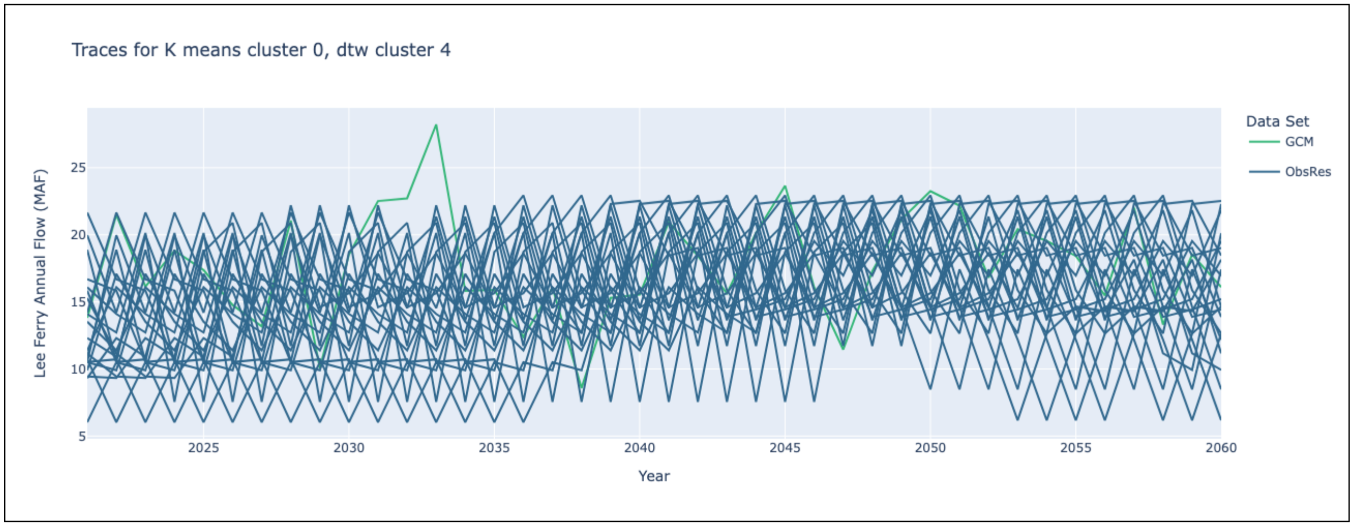


Figure 13: All traces associated with the k means group 0 and dynamic time warping group 4. This represents a cluster of traces that the algorithms placed with one another based on their rolling window/ frequency and duration characteristics and temporal features.

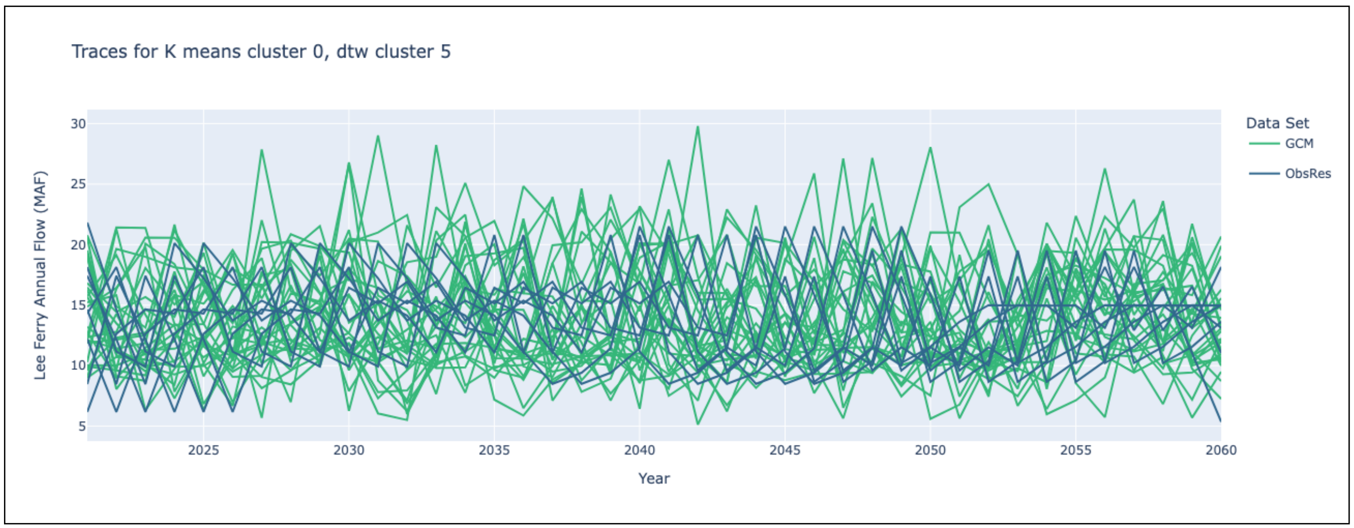


Figure 14: All traces associated with the k means group 0 and dynamic time warping group 5. This represents a cluster of traces that the algorithms placed with one another based on their rolling window/ frequency and duration characteristics and temporal features.

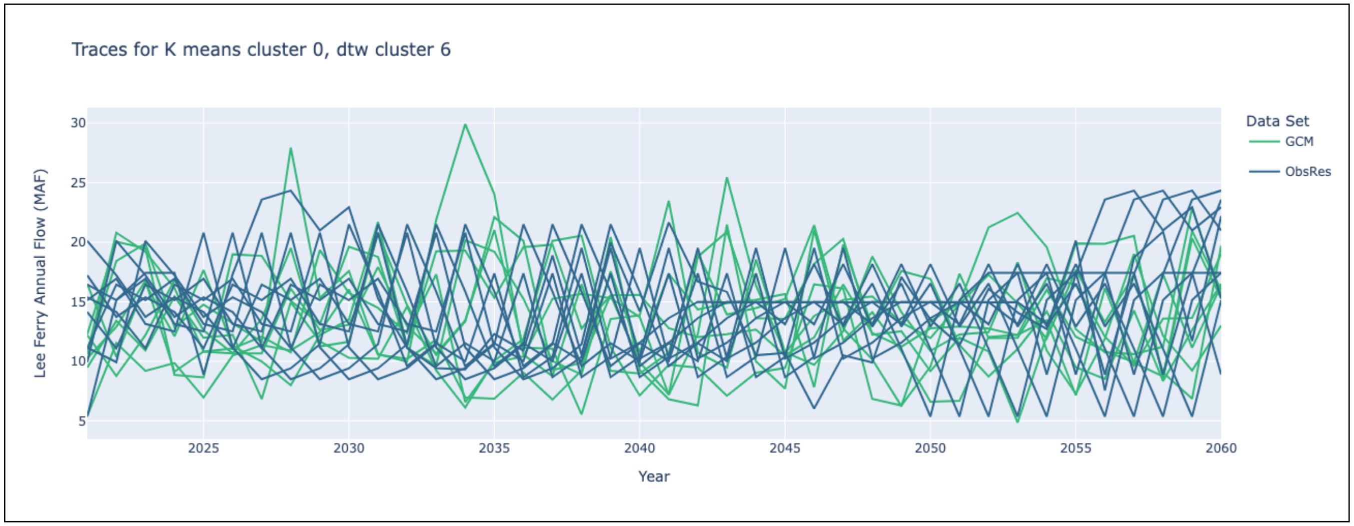


Figure 15: All traces associated with the k means group 0 and dynamic time warping group 6. This represents a cluster of traces that the algorithms placed with one another based on their rolling window/ frequency and duration characteristics and temporal features.

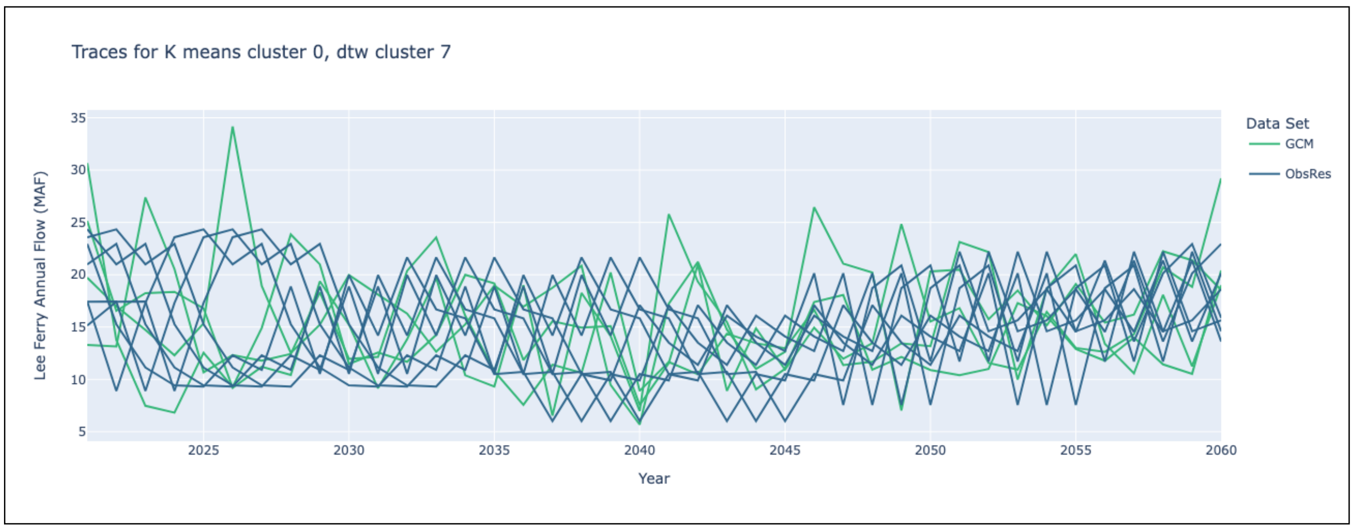


Figure 16: All traces associated with the k means group 0 and dynamic time warping group 7. This represents a cluster of traces that the algorithms placed with one another based on their rolling window/ frequency and duration characteristics and temporal features.

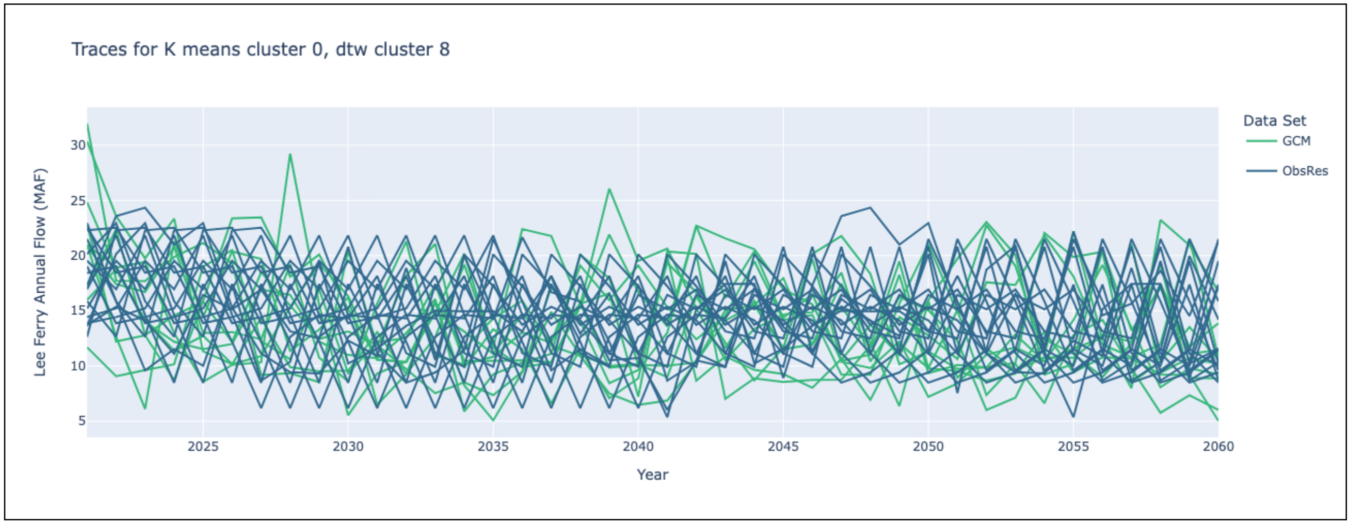


Figure 17: All traces associated with the k means group 0 and dynamic time warping group 8. This represents a cluster of traces that the algorithms placed with one another based on their rolling window/ frequency and duration characteristics and temporal features.

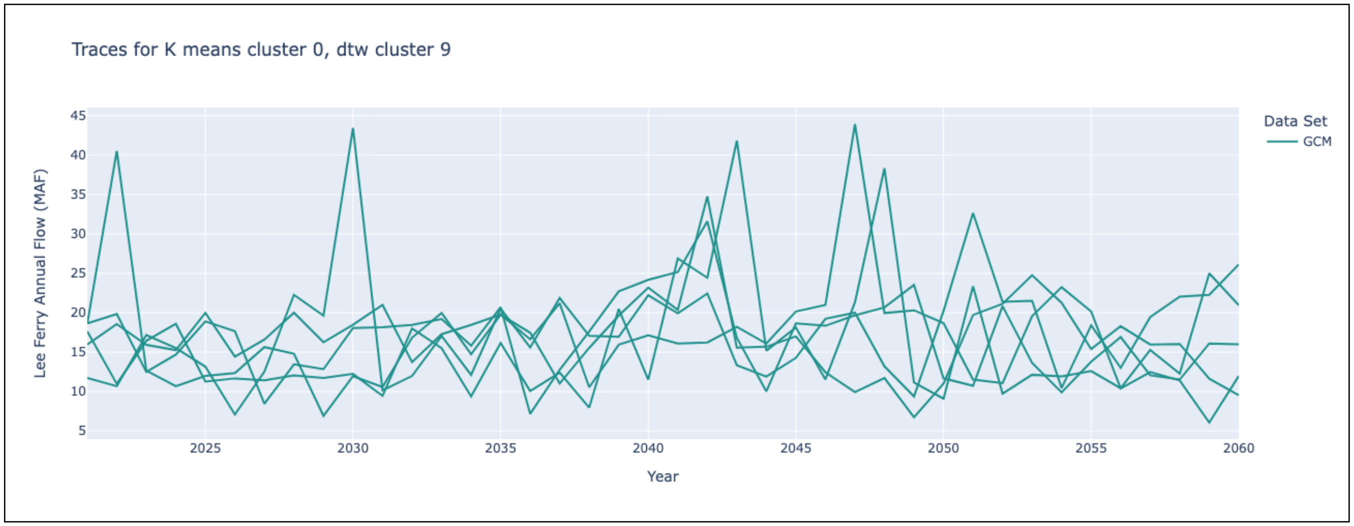


Figure 18: All traces associated with the k means group 0 and dynamic time warping group 9. This represents a cluster of traces that the algorithms placed with one another based on their rolling window/ frequency and duration characteristics and temporal features.

All DTW Trace Clusters for K-Means Cluster 1

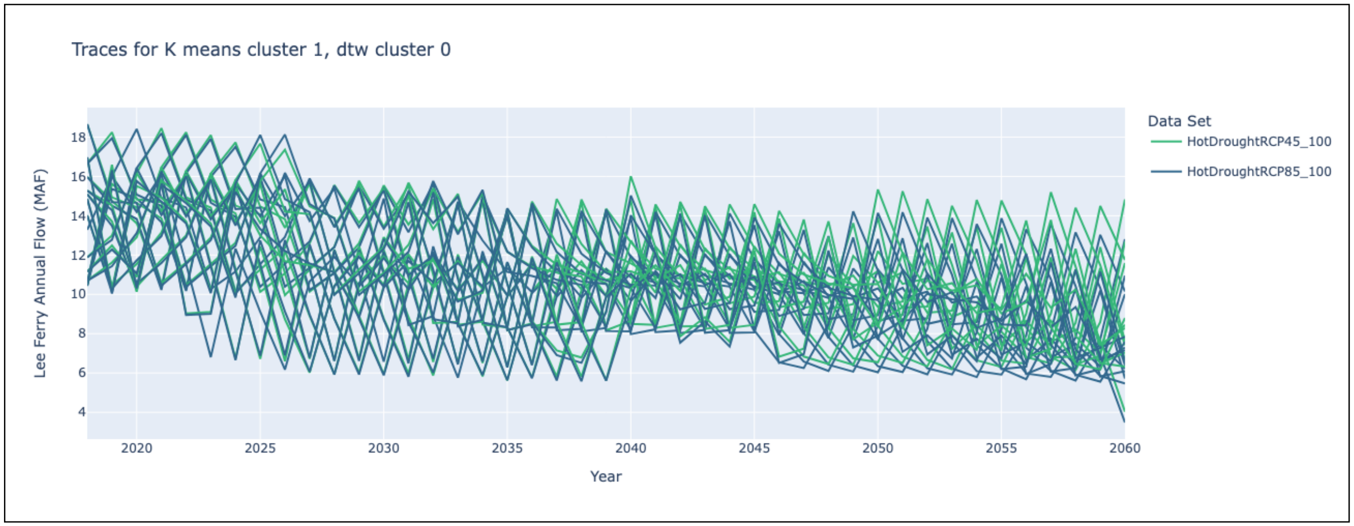


Figure 19: All traces associated with the k means group 1 and dynamic time warping group 0. This represents a cluster of traces that the algorithms placed with one another based on their rolling window/ frequency and duration characteristics and temporal features.

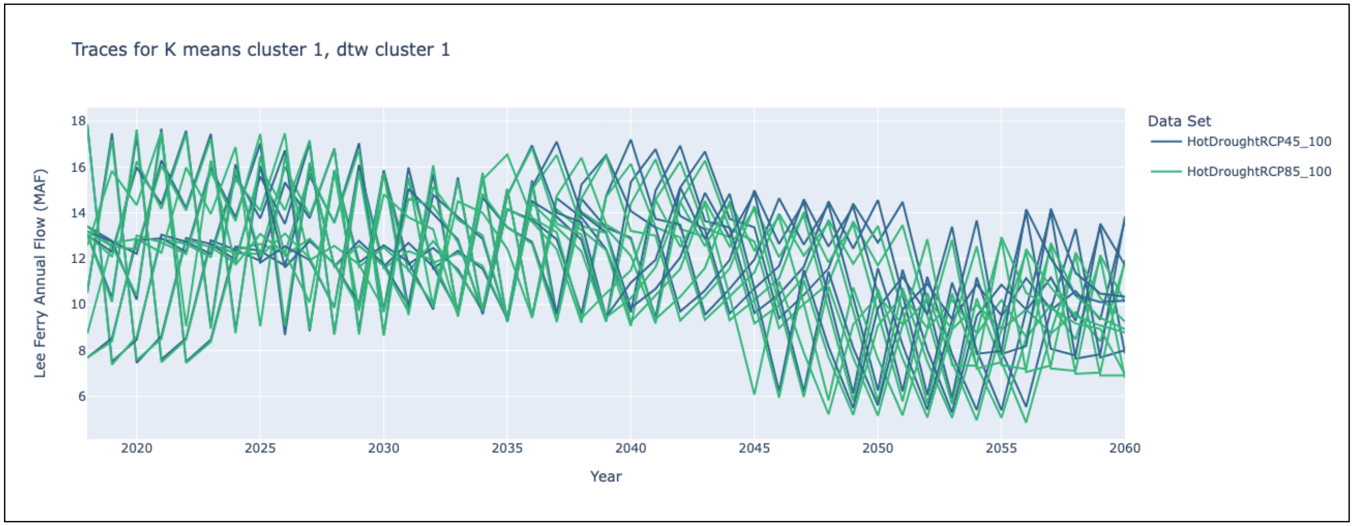


Figure 20: All traces associated with the k means group 1 and dynamic time warping group 1. This represents a cluster of traces that the algorithms placed with one another based on their rolling window/ frequency and duration characteristics and temporal features.

VIRGA LABS

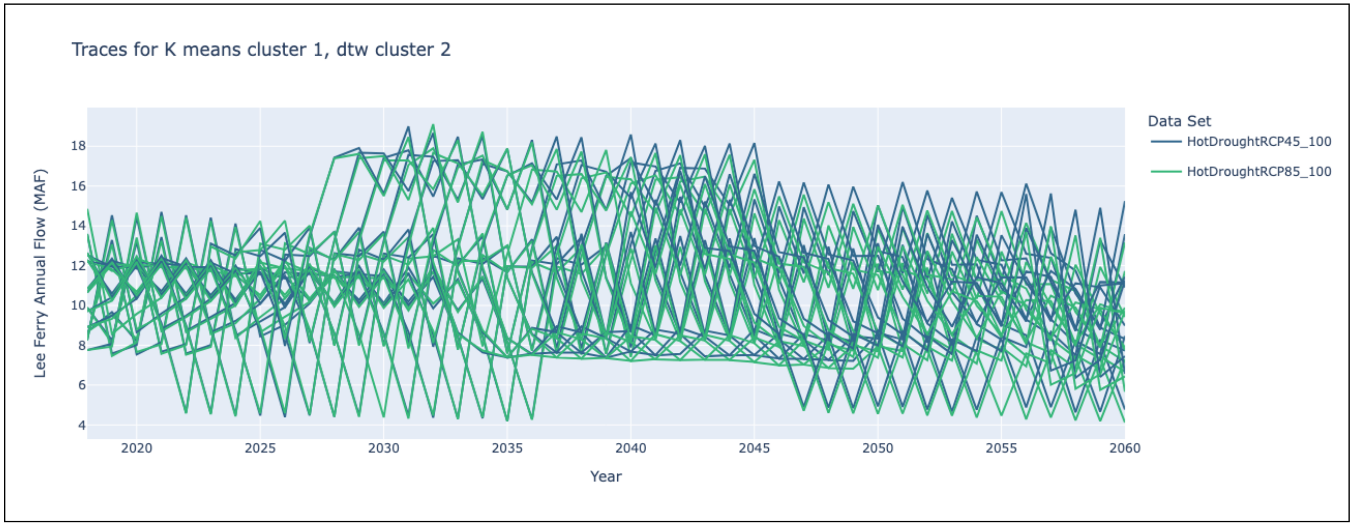


Figure 21: All traces associated with the k means group 1 and dynamic time warping group 2. This represents a cluster of traces that the algorithms placed with one another based on their rolling window/ frequency and duration characteristics and temporal features.

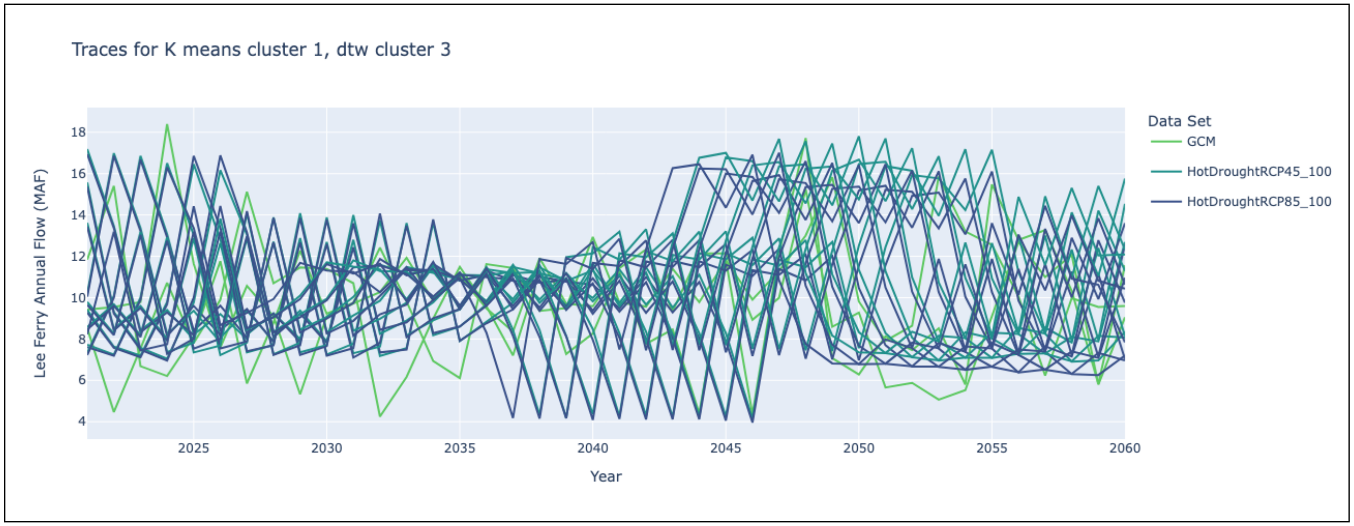


Figure 22: All traces associated with the k means group 1 and dynamic time warping group 3. This represents a cluster of traces that the algorithms placed with one another based on their rolling window/ frequency and duration characteristics and temporal features.

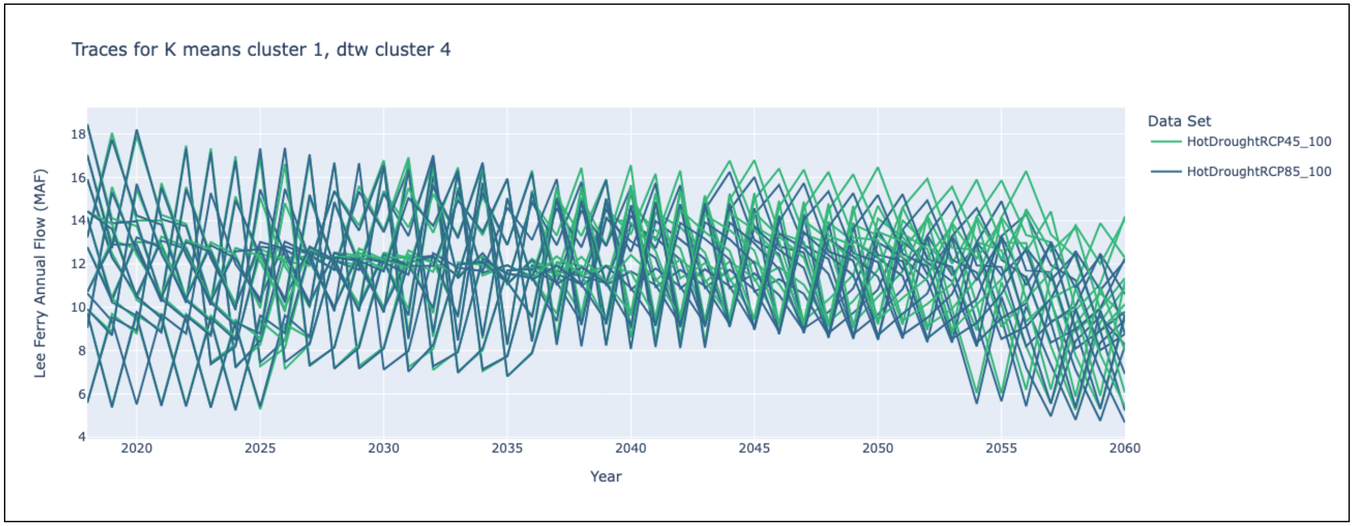


Figure 23: All traces associated with the k means group 1 and dynamic time warping group 4. This represents a cluster of traces that the algorithms placed with one another based on their rolling window/ frequency and duration characteristics and temporal features

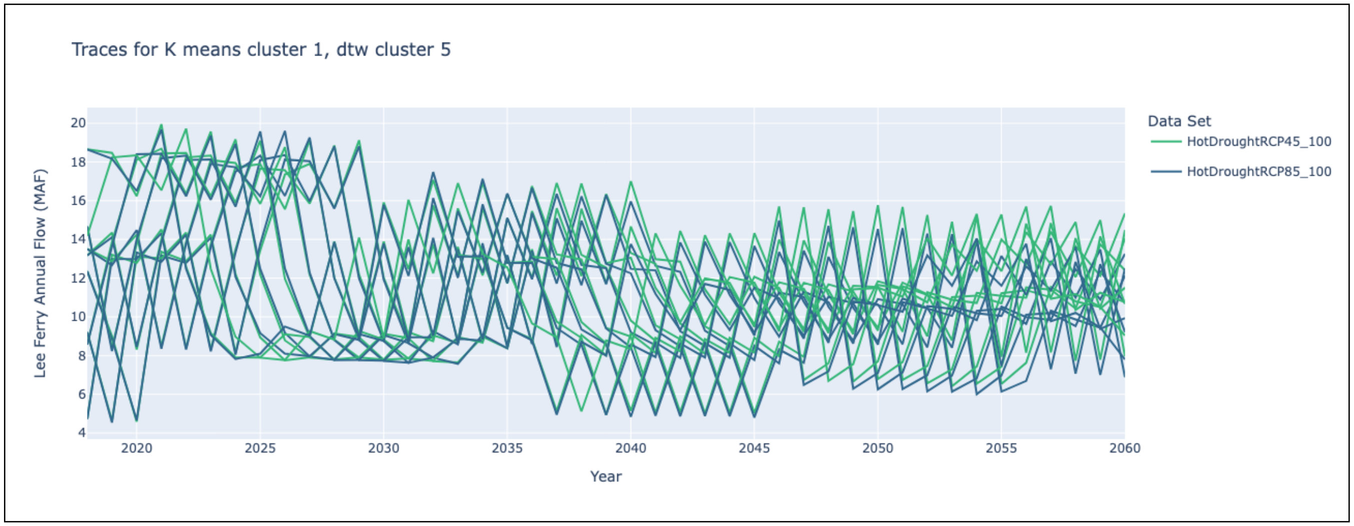


Figure 24: All traces associated with the k means group 1 and dynamic time warping group 5. This represents a cluster of traces that the algorithms placed with one another based on their rolling window/ frequency and duration characteristics and temporal features.

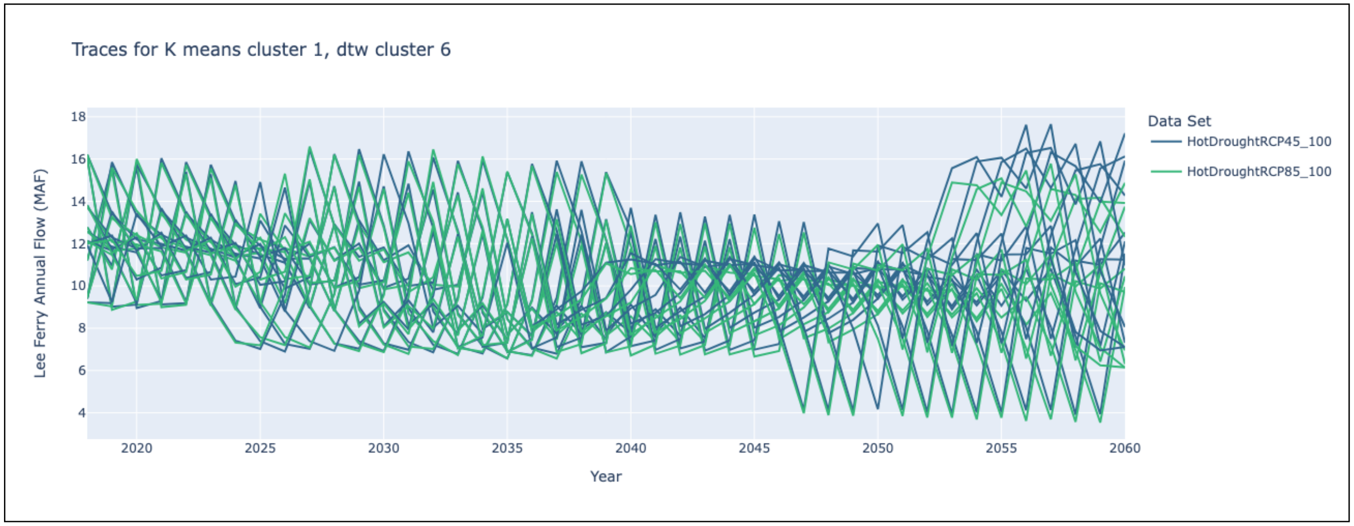


Figure 25: All traces associated with the k means group 1 and dynamic time warping group 6. This represents a cluster of traces that the algorithms placed with one another based on their rolling window/ frequency and duration characteristics and temporal features.

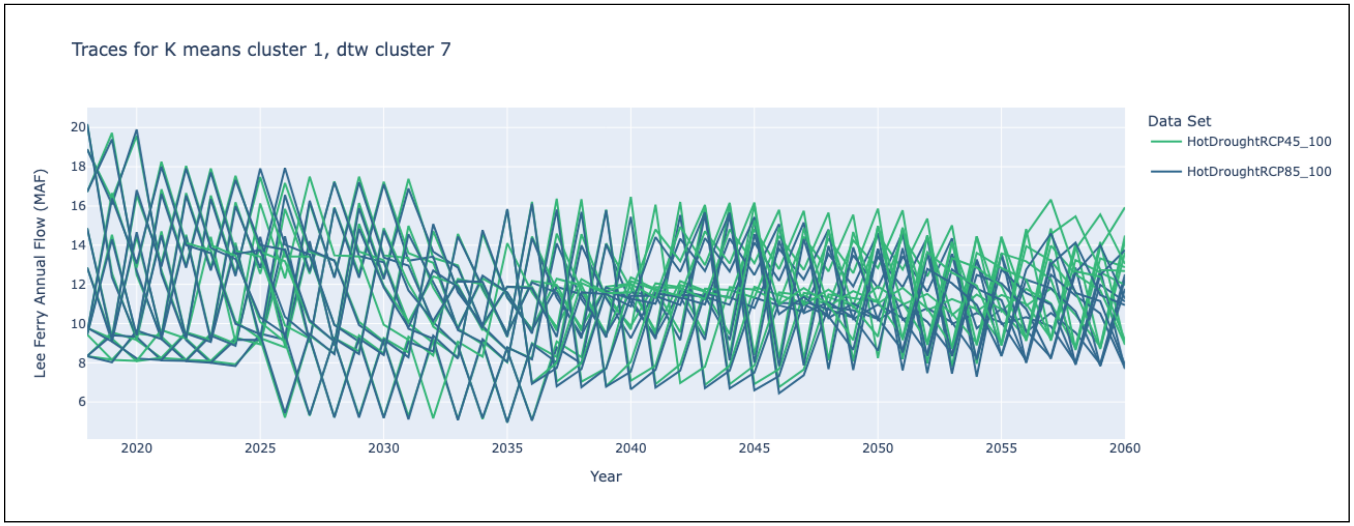


Figure 26: All traces associated with the k means group 1 and dynamic time warping group 7. This represents a cluster of traces that the algorithms placed with one another based on their rolling window/ frequency and duration characteristics and temporal features.

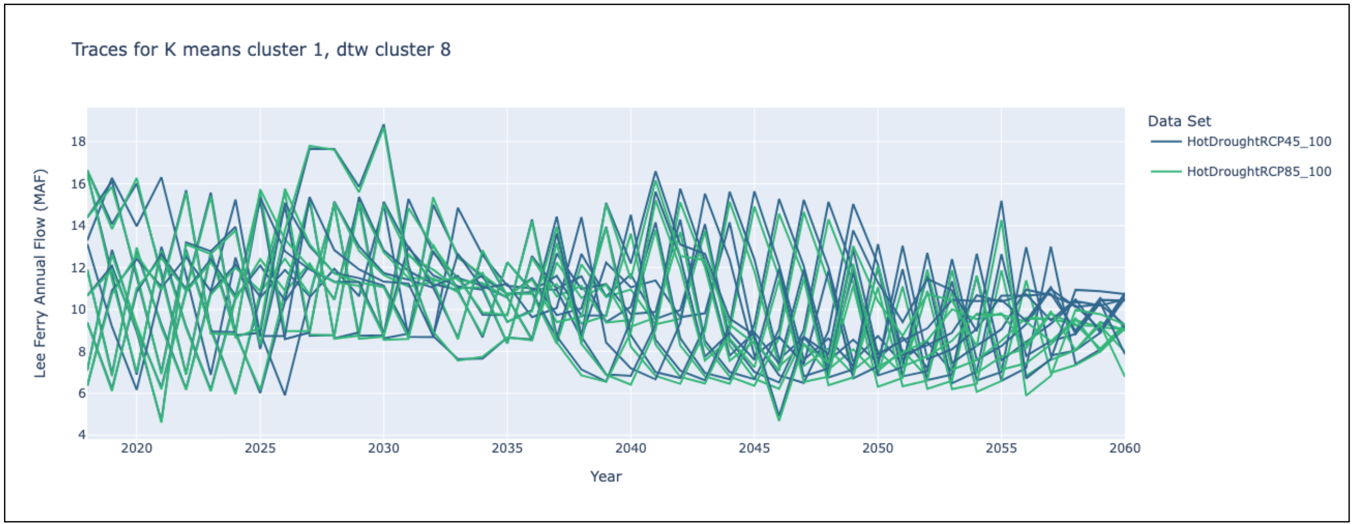


Figure 27: All traces associated with the k means group 1 and dynamic time warping group 8. This represents a cluster of traces that the algorithms placed with one another based on their rolling window/ frequency and duration characteristics and temporal features.

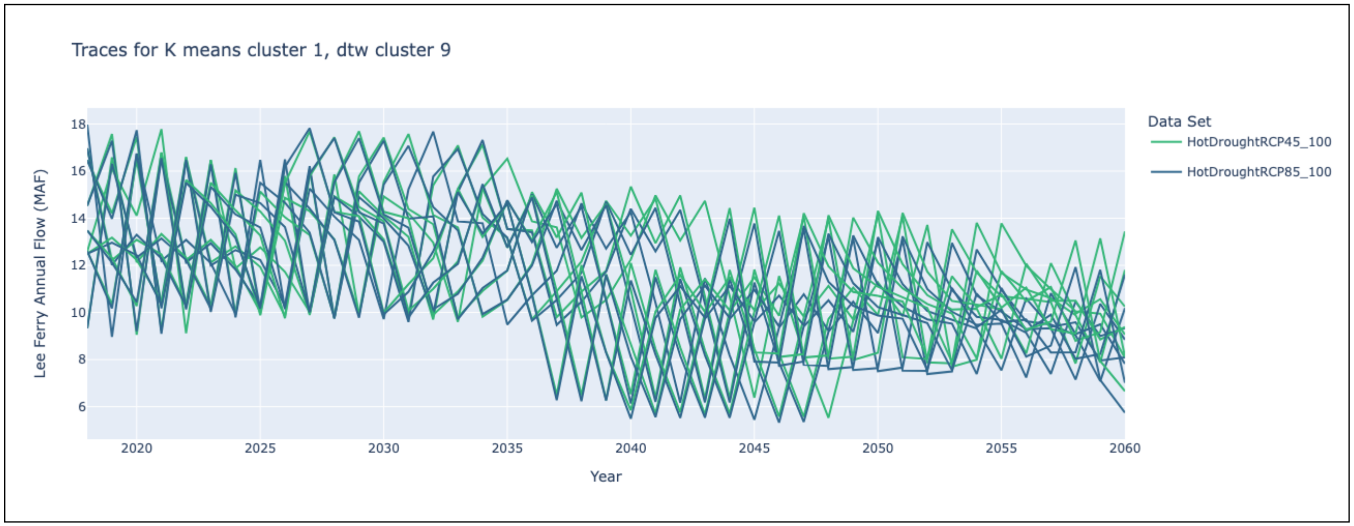


Figure 28: All traces associated with the k means group 1 and dynamic time warping group 9. This represents a cluster of traces that the algorithms placed with one another based on their rolling window/ frequency and duration characteristics and temporal features.

3)POST MODELING ANALYSIS & SELECTION

To make hydrology ensemble scenario selections within each cluster we averaged the traces by year and plotted them all on the same axis. Figure 28 includes all 20 average Lee Ferry flow cluster averages and demonstrates the similarity between many of the clusters on average. We looked at both the average behavior of the clusters as well as the variance of the individual traces in the clusters to narrow down the clusters we selected to use as hydrology ensemble scenarios.

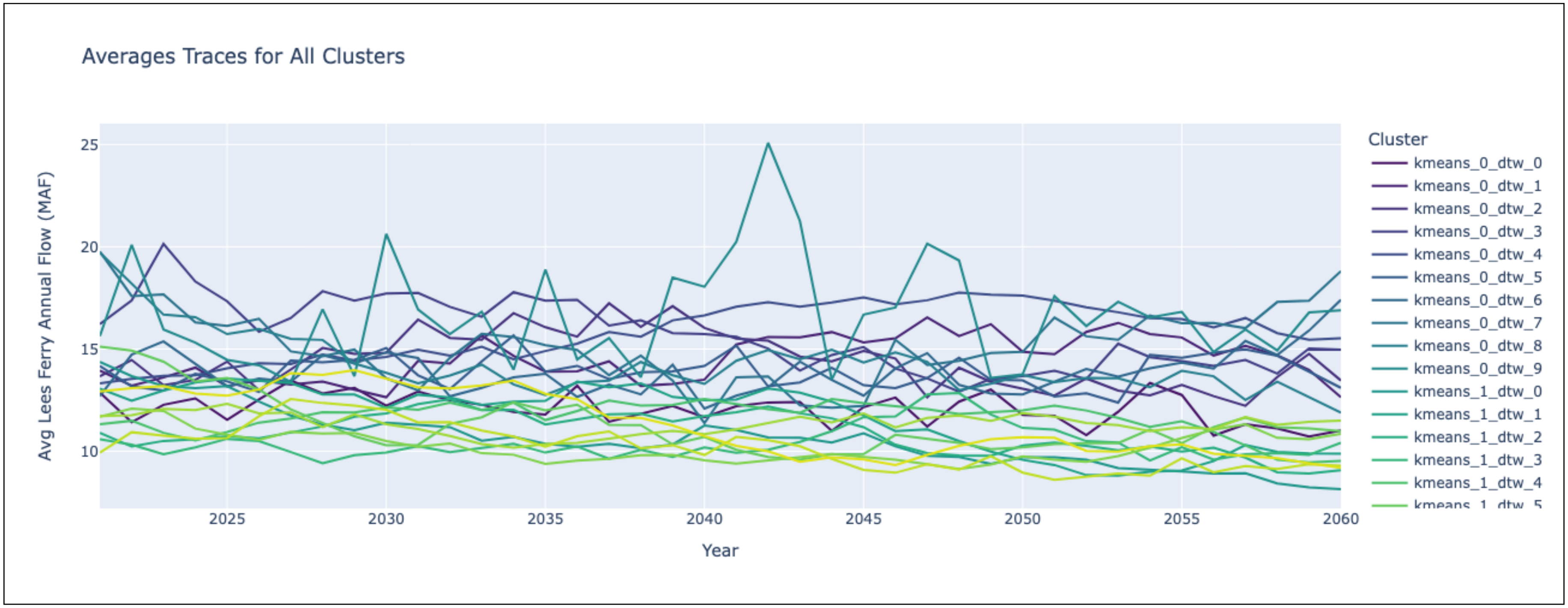


Figure 28: Average annual Lee Ferry flow by cluster for all 20 clusters returned from both algorithms. Each line represents the average behavior of traces that were grouped with one another based on their rolling window/frequency and duration characteristics and temporal features.

After reviewing these features and looking for clusters that exhibited distinct and interesting patterns, such as large spikes in average flow, different variances among traces, and periods of time that showed flows that were different from other average traces. We selected the clusters highlighted in Figure 29 then developed descriptions, statistics, and short names for each of the 9 clusters. These details can be found in the *Scenario 101* document.

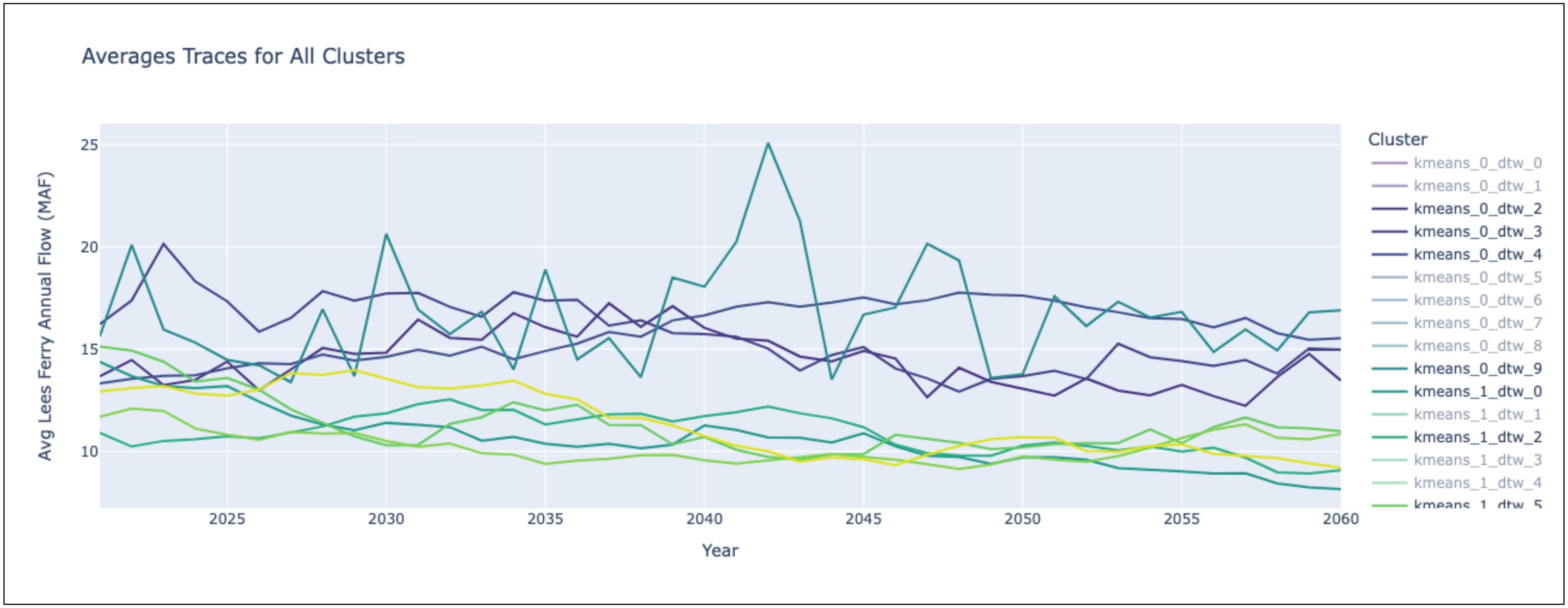


Figure 29: Average annual Lee Ferry flow by cluster for only the 9 clusters that were selected for use. Each line represents the average behavior of traces that were grouped with one another based on their rolling window/frequency and duration characteristics and temporal features.