Appendix B.

Development of Hydrogeographic Fish Assemblages for California

### Nick Santos, Ryan Peek, Rob Lusardi and Sarah Yarnell

### Center for Watershed Sciences, UC Davis

# Introduction

As part of (text from TO), identification of priority ecological endpoints needed. Using the PISCES fish database (ref), we sought to determine xxx. PISCES is and provides…

Focus on flow sensitive species (why they’re important from Ted’s paper, best suited for comparisons to flow regimes and for evaluation of impacts of eflow regimes.

Clustered flow sensitive species by geographic proximity to determine assemblages that are most similar within a defined region, or most distinct based on dissimilarity with other assemblages within region. Chose 5 regions in CA that represent geographically unique regions with HUC4-defined boundaries. Then ran k-means clustering on species ranges at the HUC12 scale within each region to determine species groupings within the region that were most distinct. Expert opinion was used to assess the various clusters and determine which set of clusters was most geographically accurate and ecologically sound (get summarized info from audio recording).

Details on the methods used to determine the hydrogeographic fish assemblages is provided below. The results, including an example of the analysis completed to inform cluster selection, is also provided below. Additional data and plots generated for each regional analysis not presented below is provided in a supplemental html file.

# Methods

## Fish Taxa

Taxa range information was pulled from PISCES (Santos et al, 2014), a comprehensive database of native fish taxa ranges for California that stores range information using HUC 12 subwatersheds as defined within the Watershed Boundary Dataset (WBD)[[1]](#footnote-1). Although clusters for many different subsets of native California fish taxa including anadromous, wide ranging, narrow-ranged, and flow sensitive (as defined in Grantham et al, 2014) could be used to create regional fish assemblages, we decided that that the Flow Sensitive species subset was the most appropriate focus for determining regional fish assemblages that would relate to and be most impacted by natural or altered flow regimes. Thus we completed clustering using the Flow Sensitive species subset (see species list in Appendix B.1).

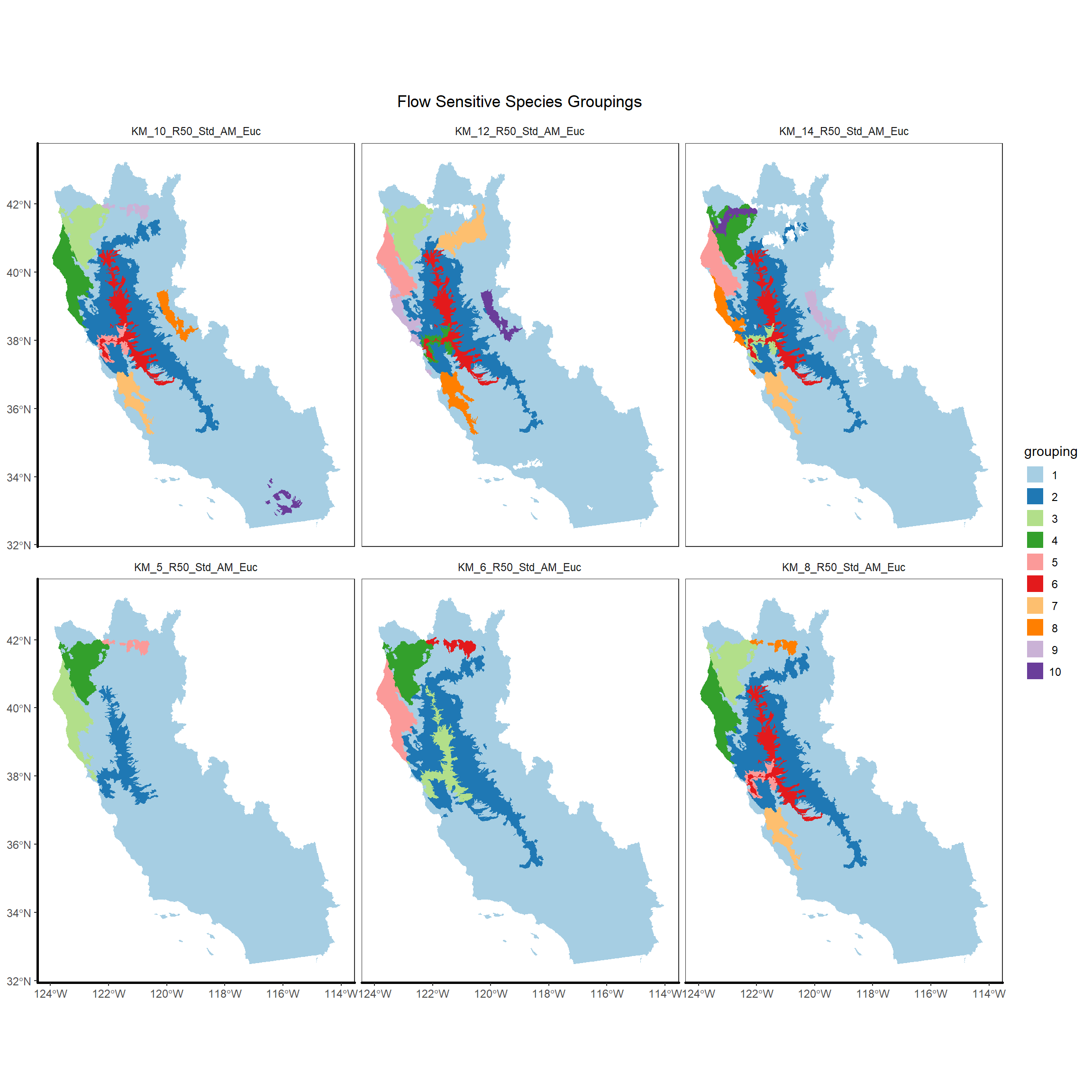


Figure 1 Results of the first Kmeans clustering using flow sensitive species

## Regional Designations

In initial trials of clustering, we found that statewide clustering did not generate useful hydrogeographic assemblages due to the small scale of the HUC12 units, which caused some clusters to cover large portions of the state and include large, disconnected taxa lists. Data output from these early iterations can be found at (link to statewide groupings from earlier methods section below assuming that’s the right link?). We therefore sought to determine appropriate regions within the state for clustering that would be large enough to encompass broad geographical areas, but not so small as to limit the usefulness of clustering. We considered using HUC6 and HUC8 units (ref) or the Moyle Zoogeographic Regions (ref) due to their inherent hydrologic and/or ecological basis. However, we found that the scale of each of these regions was either too narrow (HUC6-8) or too broad (zoogeographic) or simply didn’t account for the inherent multi-scale hydrologic connectivity and thus similarity in many species groupings across the state. We therefore settled on defining 5 regions that were created by first defining hydrologic regions at the HUC4 scale and then combining complete HUC4 units so that watersheds remained connected in general geographic regions. See figure \_\_ for a map of the final regions used. In essence, these regions are similar to zoogeographic regions but xxxx(lumped into fewer regions? Maybe show our region map next to zoogeographic?).

* Central Valley & West Slope Sierra  
  *All HUC4s flowing out of the Central Valley via the San Francisco Bay, including the west slope of the Sierra Nevada, the Coast Range HUCs flowing into the valley, and the Pit River system*
* North Coast  
  *All HUC4s flowing out the Klamath river, as well as all HUC4s from Klamath river past San Francisco Bay to the top of Monterey Bay along the west coast*
* South Coast  
  *All HUC4s from the Monterey Bay to the US-Mexico border along the west coast*
* Great Basin  
  *HUC4s flowing from the eastern Sierra/Modoc Plateau into the Great Basin north of Mono Lake*
* Desert  
  *HUC4s from landlocked desert drainages within California and any great basin HUC4 that contacts California south of Mono Lake (in order to keep contiguous regions), as well as Colorado River HUC4s in the southeast of the state*

In the original clustering, the south coast region was split between the desert and north coast (originally just “coast”) regions. The desert region was not used in clustering due to its sparse coverage of fish. 

## Clustering

As part of this project, species ranges were clustered into hyrdrogeographic assemblages, which are sets of 3-6 spatial groupings of HUC12 level species assemblages within each region (except the desert region). Data were pulled from PISCES, using only a species’ current presence, ignoring historical presence and translocations. This data corresponded to the following data types within PISCES:

* 1: Observed current presence
* 3: Extant range - expert opinion
* 9: Reintroduction after local extirpation

Records in PISCES for a taxon with one of those data types counted the taxon as present in a given HUC12 for the purposes of the clustering. Versions using historical taxa ranges and/or translocated taxa ranges were also tried but rejected as being less useful for modern management.

Initial exploratory clustering was performed using the Kmeans tool in the GeoDa package[[2]](#footnote-2) (Anselin et al, 2006) with final clustering performed using the ArcGIS Grouping Analysis tool in the Spatial Statistics toolbox with a spatial constraint setting of “CONTIGUITY\_EDGES\_CORNERS". This constraint requires groups to be contiguous through at least one shared vertex or HUC12 edge.

The clustering tool Zonation was also considered for this project, but was ultimately not chosen due to prior experience and a lack of future software support.

For each subregion and set of PISCES data types used, clusters were generated for grouping schemes where the maximum number of clusters, K, ranged from two to nine, as illustrated in figure \_\_ below.

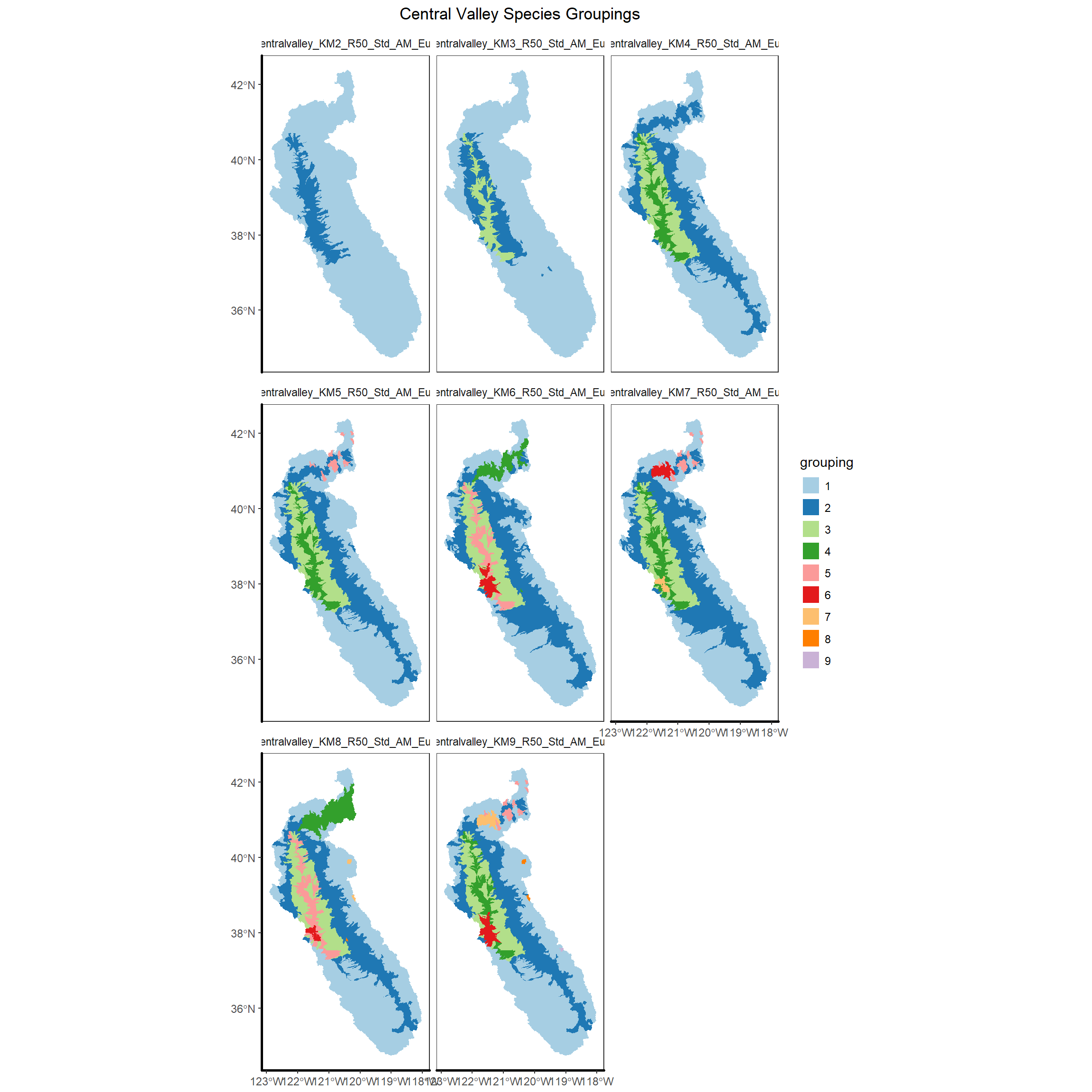


Figure \_\_: An early clustering for the Central Valley and West Slope, showing clusters generated when the maximum number of groups ranges from 2 to 9.

## Evaluation

Once we determined the appropriate regional scale for clustering and the methodology for clustering as desribed above, we generated a series of clusters for each region with k set as 2 to 9. A common challenge with Kmeans-based approaches is determining the appropriate numbers of clusters to use for the data. We considered using selection criteria tied to the stream classifications as a potential way to more objectively select the appropriate number of clusters for each region. These criteria were intended to be hydrologically meaningful by trying to assess the lowest number of clusters associated with

1. the majority of stream classes present in a region were represented as a dominant class in at least one cluster
2. More than half of clusters had a dominant stream class that covered more than 60% of the cluster’s area

However, after repeated clustering iterations, we found that these criteria were still potentially arbitrary and not as objective as we’d hoped. Thus we chose to use expert opinion to determine what number of clusters in each region would be meaningful and useful from a management-perspective. The final number of clusters was therefore determined via expert opinion.

## Earlier Methods

As was mentioned earlier in the methods, the choice of clusters has been a highly iterative process. For documentation of methodology and available data, we include a listing of the clustering attempts and data analysis that were not chosen, but which informed the direction of the project.

* [Flow-class by HUC dominance](https://ucd-cws.github.io/eflow-species/eflow-distance.html)
* [mapping species directly to streams](https://ucd-cws.github.io/eflow-species/eflow-list.html)
* [Groupings of many species groups statewide, using GeoDa, including historical and current assemblage clusters](https://ucd-cws.github.io/eflow-species/map_facets.html)
* Flow Sensitive species by s[ub-regions, using GeoDa](https://ucd-cws.github.io/eflow-species/huc_region_groups_kplots.html)
* New clustering algorithm (ArcGIS), with a minimum number of 2 taxa per clustered HUC
* Comparison of ArcGIS clusters with GeoDa clusters, including translocations, excluding minimum numbers of taxa
* Excluding translocations
* Including all data types (current, translocated, historical)
* [Clustering with a requirement that clusters be contiguous - no translocations or historical](https://ucd-cws.github.io/eflow-species/huc_region_groups_neighbors.html)
* Larger numbers of clusters with contiguous clusters

Code for analyzing the presence data and clusters and generating figures can be found at <http://github.com/ucd-cws/eflow-species>

# Appendix B.1 - Species List for Flow Sensitive Species

**Scientific Name**

**Common Name**

|  |
| --- |

Entosphenus tridentata

Pacific lamprey

|  |
| --- |
|  |

Entosphenus similis

Klamath River lamprey

|  |
| --- |
|  |

Lampetra ayersi

River lamprey

|  |
| --- |
|  |

Lampetra hubbsi

Kern brook lamprey

|  |
| --- |
|  |

Lampetra richardsoni

Western brook lamprey

|  |
| --- |
|  |

Acipenser medirostris

Northern green sturgeon

|  |
| --- |
|  |

Acipenser medirostris

Southern green sturgeon

|  |
| --- |
|  |

Acipenser transmontanus

White sturgeon

|  |
| --- |
|  |

Siphatales bicolor snyderi

Owens tui chub

|  |
| --- |
|  |

Gila coerulea

Blue chub

|  |
| --- |
|  |

Gila orcutti

Arroyo chub

|  |
| --- |
|  |

Lavinia exilicauda exilicauda

Sacramento hitch

|  |
| --- |
|  |

Lavinia exilicauda harengeus

Monterey hitch

|  |
| --- |
|  |

Hesperoleucus symmetricus symmetricus

California roach

|  |
| --- |
|  |

Hesperoleucus symmetricus subspecies

Red Hills roach

|  |
| --- |
|  |

Hesperoleucus symmetricus subditus

Southern coastal roach

|  |
| --- |
|  |

Hesperoleucus symmetricus subspecies

Tomales roach

|  |
| --- |
|  |

Hesperoleucus mitrulus

Northern roach

|  |
| --- |
|  |

Pogonichthys macrolepidotus

Sacramento splittail

|  |
| --- |
|  |

Mylopharodon conocephalus

Hardhead

|  |
| --- |
|  |

Ptychocheilus grandis

Sacramento pikeminnow

|  |
| --- |
|  |

Rhinichthys osculus subspecies

Sacramento speckled dace

|  |
| --- |
|  |

Rhinichthys osculus robustus

Lahontan speckled dace

|  |
| --- |
|  |

Rhinichthys osculus subspecies

Owens speckled dace

|  |
| --- |
|  |

Rhinichthys osculus subspecies

Santa Ana speckled dace

|  |
| --- |
|  |

Pantosteus lahontan

Lahontan mountain sucker

|  |
| --- |
|  |

Catostomus occidentalis occidentalis

Sacramento sucker

|  |
| --- |
|  |

Catostomus occidentalis mnioltiltus

Monterey sucker

|  |
| --- |
|  |

Catostomus microps

Modoc sucker

|  |
| --- |
|  |

Catostomus snyderi

Klamath largescale sucker

|  |
| --- |
|  |

Catostomus luxatus

Lost River sucker

|  |
| --- |
|  |

Catostomus santaanae

Santa Ana sucker

|  |
| --- |
|  |

Chasmistes brevirostris

Shortnose sucker

|  |
| --- |
|  |

Xyrauchen texanus

Razorback sucker

|  |
| --- |
|  |

Spirinchus thaleichthys

Longfin smelt

|  |
| --- |
|  |

Hypomesus pacificus

Delta smelt

|  |
| --- |
|  |

Prosopium williamsoni

Mountain whitefish

|  |
| --- |
|  |

Oncorhynchus tshawytscha

Upper Klamath-Trinity fall Chinook salmon

|  |
| --- |
|  |

Oncorhynchus tshawytscha

Upper Klamath-Trinity spring Chinook salmon

|  |
| --- |
|  |

Oncorhynchus tshawytscha

California Coast fall Chinook salmon

|  |
| --- |
|  |

Oncorhynchus tshawytscha

Central Valley winter Chinook salmon

|  |
| --- |
|  |

Oncorhynchus tshawytscha

Central Valley spring Chinook salmon

|  |
| --- |
|  |

Oncorhynchus tshawytscha

Central Valley late fall Chinook salmon

|  |
| --- |
|  |

Oncorhynchus tshawytscha

Central Valley fall Chinook salmon

|  |
| --- |
|  |

Oncorhynchus kisutch

Central Coast coho salmon

|  |
| --- |
|  |

Oncorhynchus kisutch

Southern Oregon Northern California coast coho salmon

|  |
| --- |
|  |

Oncorhynchus mykiss

Northern California coast winter steelhead

|  |
| --- |
|  |

Oncorhynchus mykiss

Northern California coast summer steelhead

|  |
| --- |
|  |

Oncorhynchus mykiss

Klamath Mountains Province winter steelhead

|  |
| --- |
|  |

Oncorhynchus mykiss

Klamath Mountains Province summer steelhead

|  |
| --- |
|  |

Oncorhynchus mykiss

Central California coast winter steelhead

|  |
| --- |
|  |

Oncorhynchus mykiss

South Central California coast steelhead

|  |
| --- |
|  |

Oncorhynchus mykiss

Southern California steelhead

|  |
| --- |
|  |

Oncorhynchus mykiss gilberti

Kern River rainbow trout

|  |
| --- |
|  |

Oncorhynchus clarki henshawi

Lahontan cutthroat trout

|  |
| --- |
|  |

Cyprinodon macularius

Desert pupfish

|  |
| --- |
|  |

Cyprinodon radiosus

Owens pupfish

|  |
| --- |
|  |

Cottus klamathensis macrops

Bigeye marbled sculpin

|  |
| --- |
|  |

Cottus klamathensis polyporus

Lower Klamath marbled sculpin

|  |
| --- |
|  |

Cottus klamathensis klamathensis

Upper Klamath marbled sculpin

|  |
| --- |
|  |

Cottus gulosus

Riffle sculpin

|  |
| --- |
|  |

Cottus pitensis

Pit sculpin

|  |
| --- |
|  |

Gasterosteus aculeatus microcephalus

Inland threespine stickleback

|  |
| --- |
|  |

Gasterosteus aculeatus williamsoni

Unarmored threespine stickleback

|  |
| --- |
|  |

Archoplites interruptus

Sacramento perch

|  |
| --- |
|  |

Hysterocarpus traskii traskii

Sacramento tule perch

|  |
| --- |
|  |

Eucyclogobius newberryi

Tidewater goby

|  |
| --- |
|  |

Hesperoleucus symmetricus

Kaweah roach

|  |
| --- |
|  |

1. https://nhd.usgs.gov/wbd.html [↑](#footnote-ref-1)
2. Anselin, Luc, Ibnu Syabri and Youngihn Kho (2006). GeoDa: An Introduction to Spatial Data Analysis. Geographical Analysis 38 (1), 5-22. [↑](#footnote-ref-2)