Regional cluster analysis of groundwater quality to determine aquifer sources of wells

Kyle Hardage, PhD
Engineering Geologist
DWR SGMO Modeling Tools & Support

April 4, 2022

CALIFORNIA DEPARTMENT OF WATER RESOURCES
SUSTAINABLE GROUNDWATER MANAGEMENT OFFICE

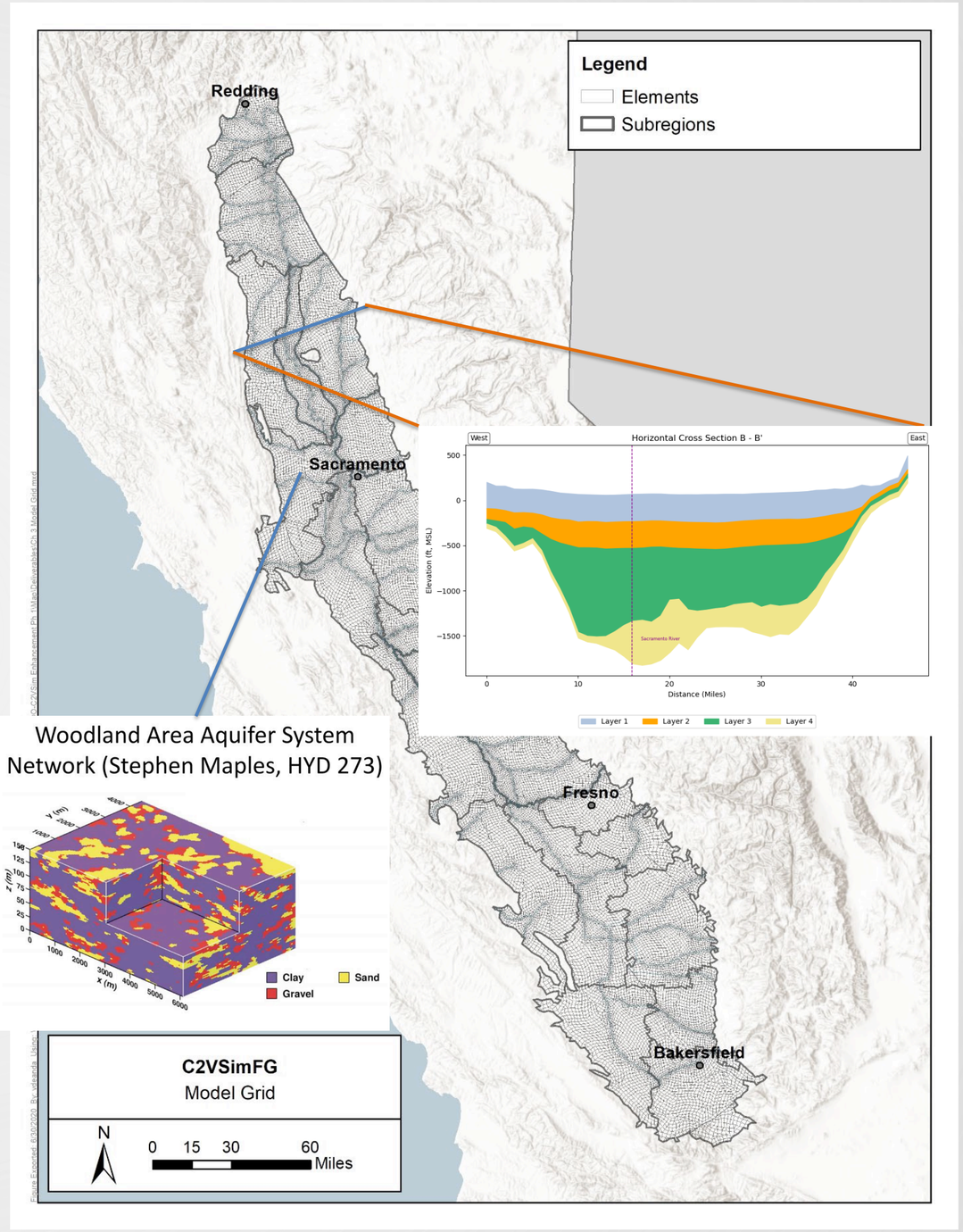
Acknowledgements

- Vivek Bedekar & Matt Tonkin (SSP&A)
- Eric Senter (SGMO/Water Data Library)
- Jason Preece (SGMO)
- Aaron Button (GAMA)

Background

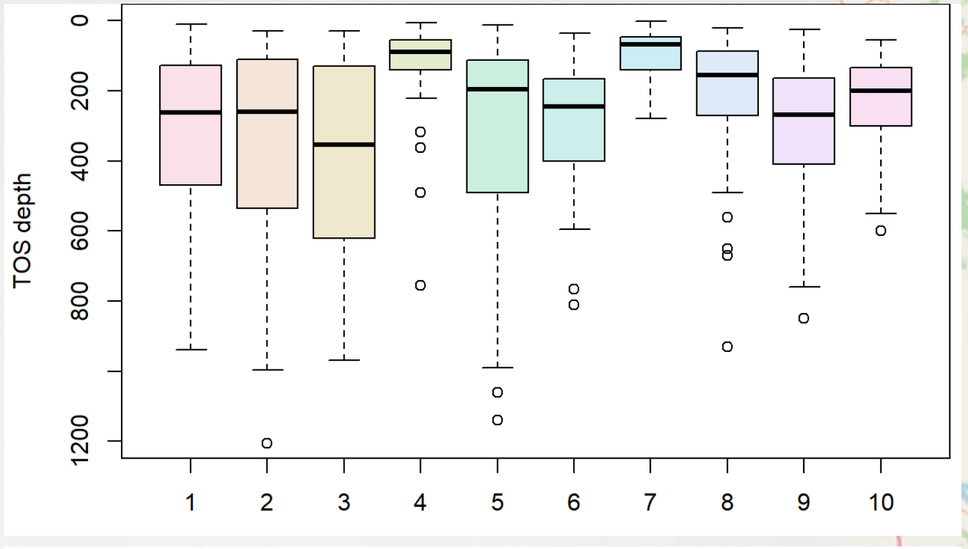
- C2VSim-FG – 4 layers
  - Miss stratigraphy complexity
- Well construction data
  - Screening intervals
- Sacramento vs.

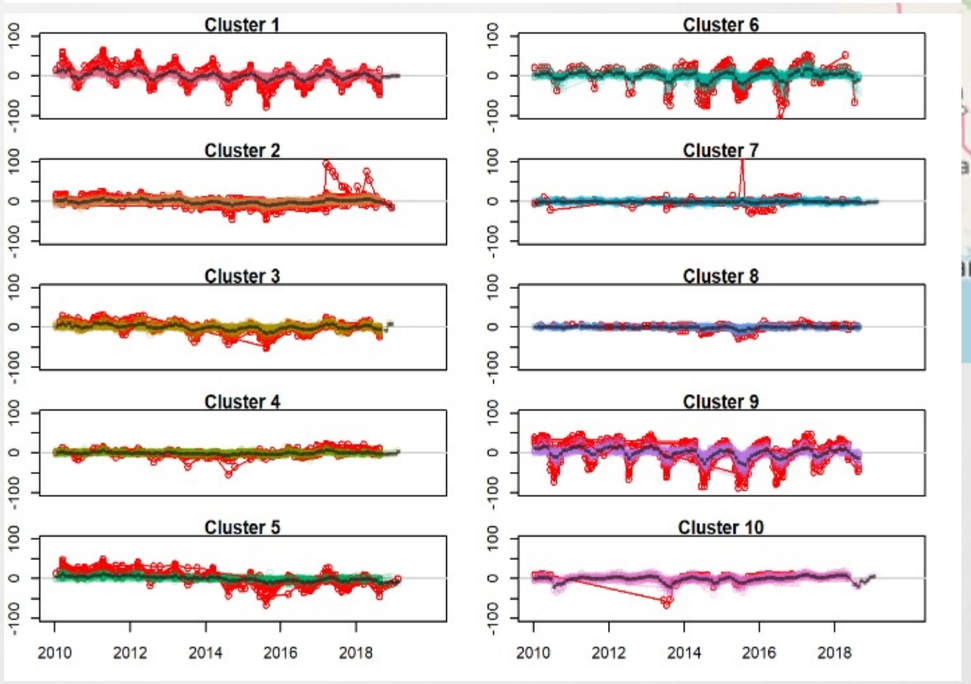
  San Joaquin data



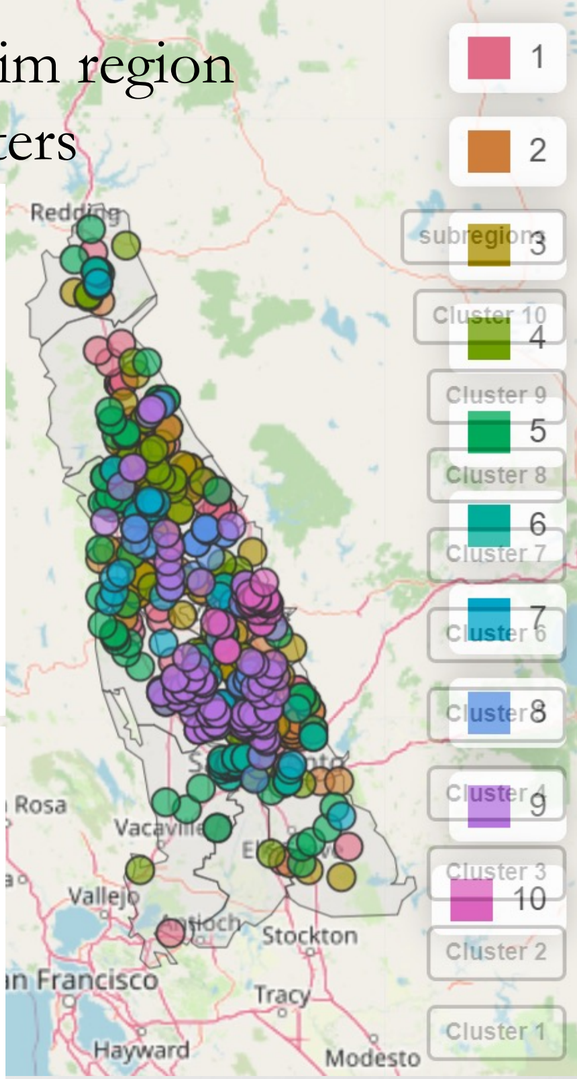
Background

- Previous clustering: type hydrographs
- Advantage
  - time series
- Disadvantage
  - single variable





SVSim region clusters



SSP&A 2020

Method

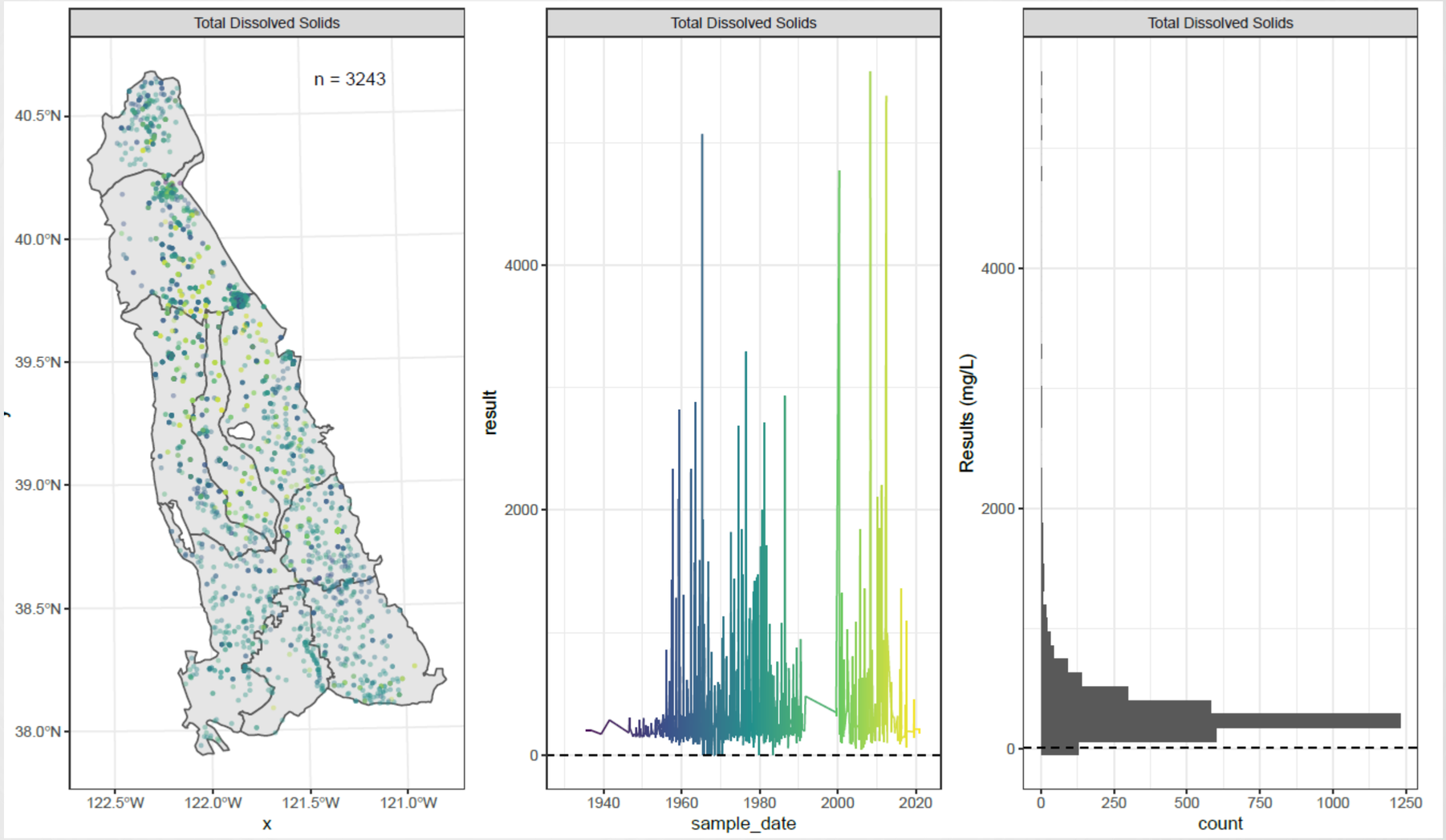
- Water quality multivariate clustering
- ~76 natural chemicals
- Advantage: many variables*
- Disadvantage: no time series
  - Hierarchical, Euclidean distance
  - Linkage: Average, Ward
- Goal: geochemical fingerprints
  - 1) screen interval
  - 2) complete set

AlkCaCO3	Cl ★	Kr	Se	HCO3 ★
Alk ★	Cr	Pb	Ag	CO3
Al	Cr6	Li	Na ★	δD
ACB	Co	Mg ★	SpecCon ★	δ15N
Sb	Cu	Mn	Sr	δ18O
Ar	H-2	Hg	SO4	P ortho
As	DO	CH4	Tl	pH ★
Ba	Dosat	Mo	TDS ★	
Be	F	Ne	TOCH	
AlkB	alpha	Ni	PHC	
B ★	beta	nitrate ★	H-3	
Br	hardness ★	nitrite	W	
Cd	HE-4	K	U	
Ca	HE-3/HE-4	Ra-226	V	
C-14	I	Ra-228	Xe	
CDS	Fe	Rn-222	Zn	

Method

Complete set varies

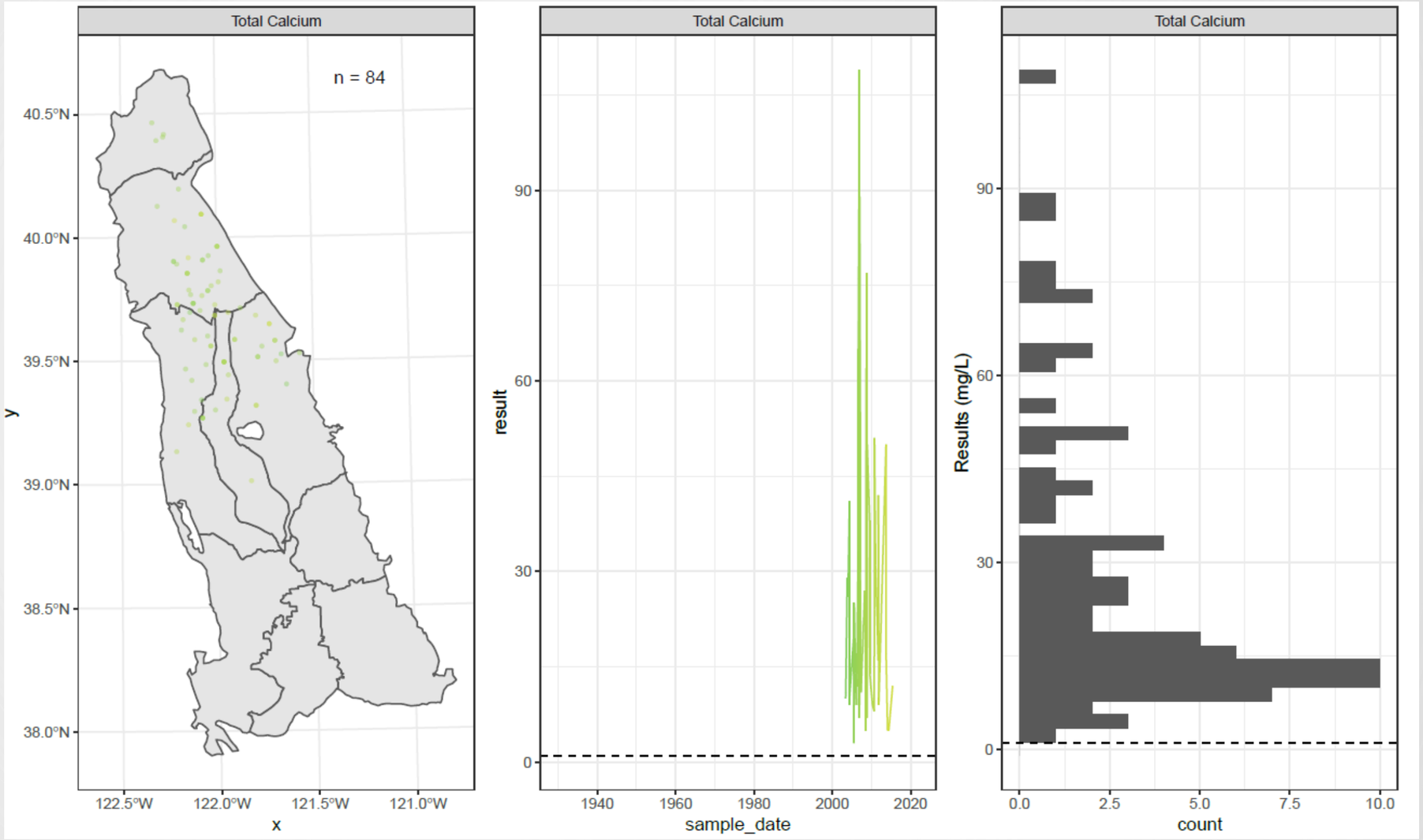
Requires averaging



Method

Complete set varies

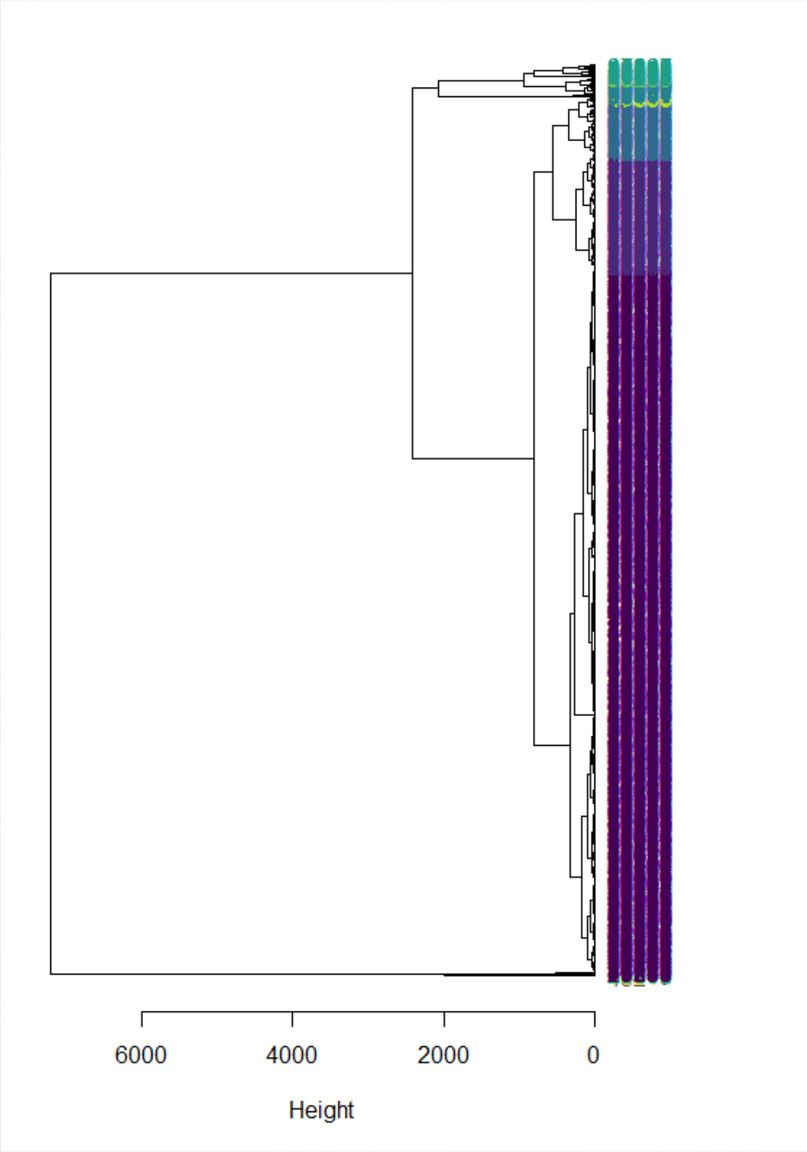
Requires averaging

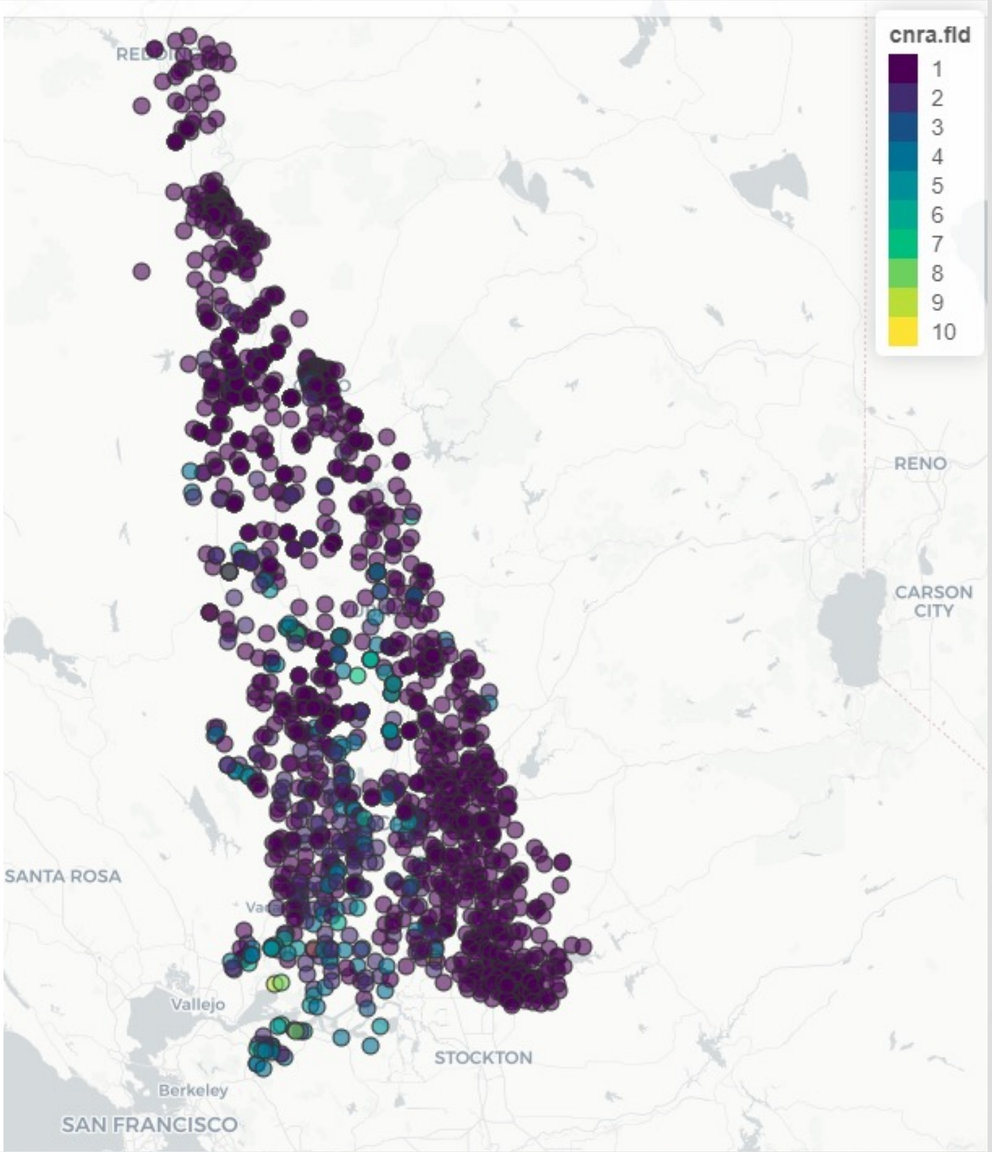


Field data

pH
specific conductance
temperature
N = 1505

Hierarchical agglom.
Euclidean
Average linkage
(cc = 0.93)
Raw data

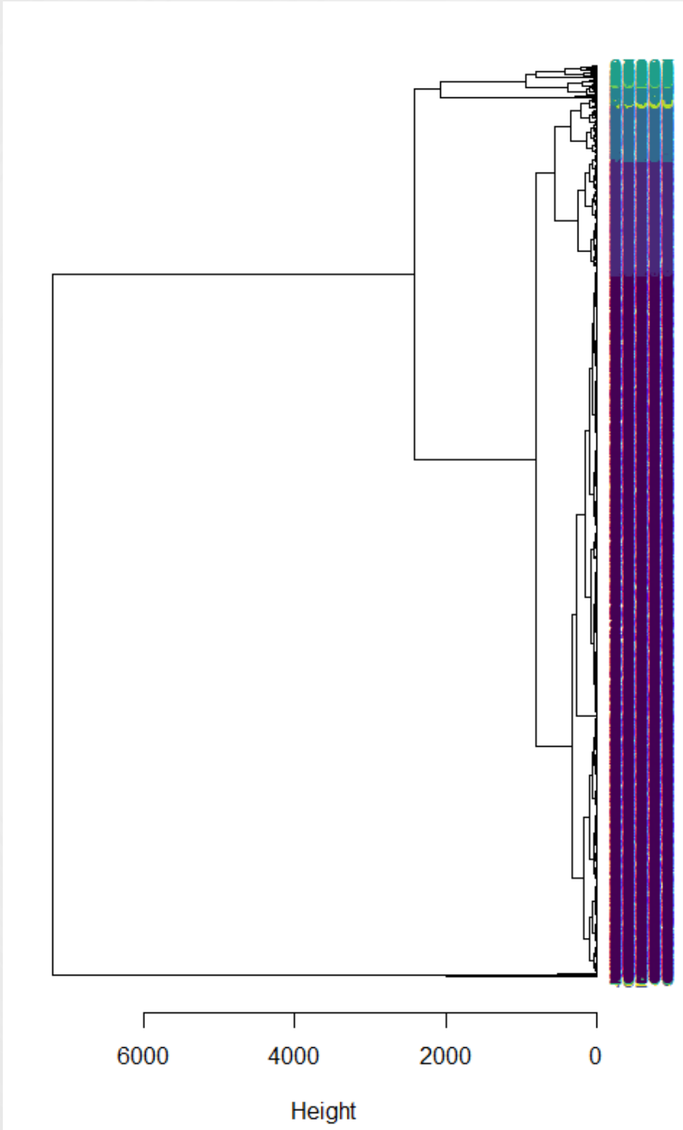


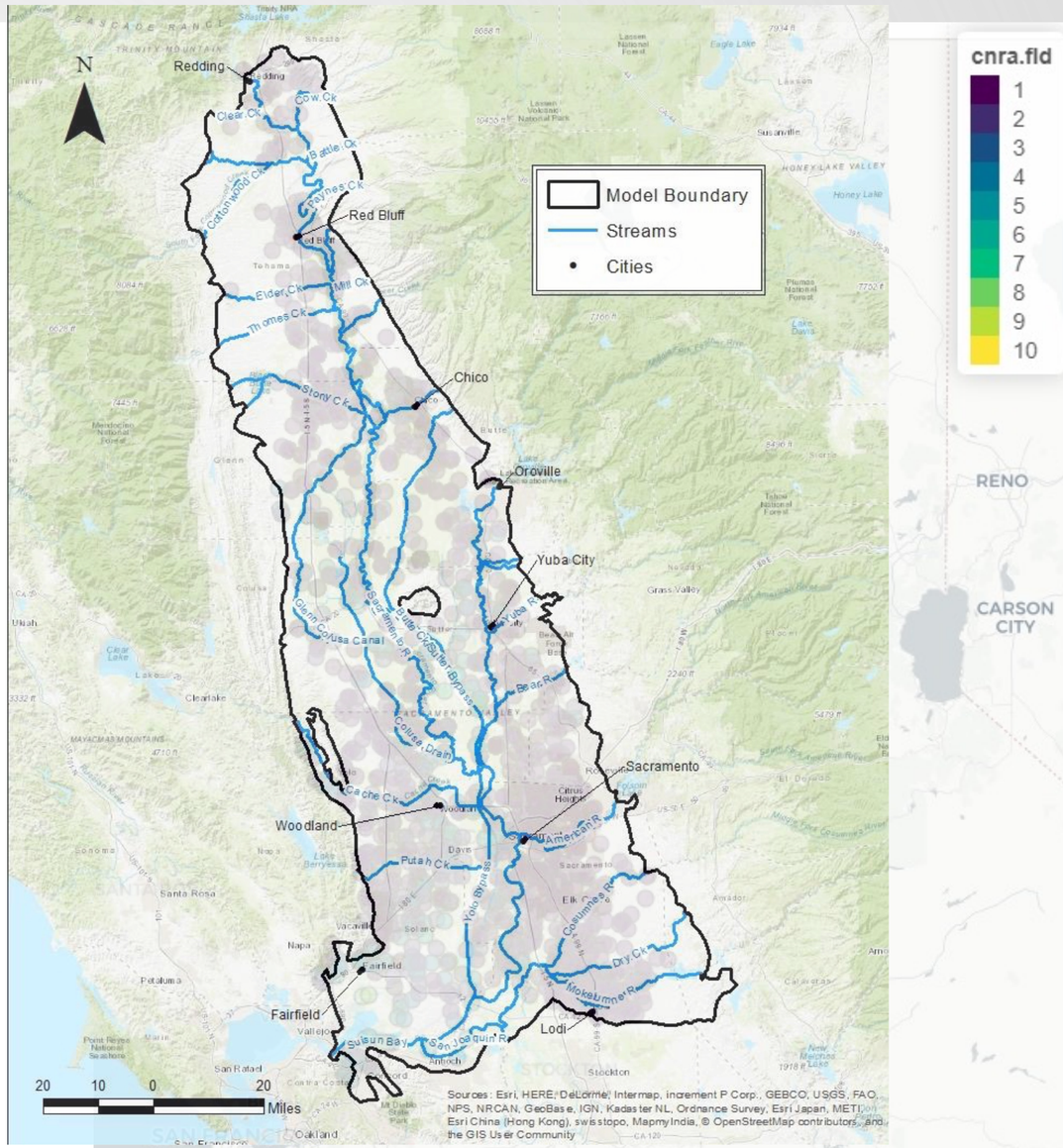


Field data

pH
specific conductance
temperature
N = 1505

Hierarchical agglom.
Euclidean
Average linkage
(cc = 0.93)
Raw data

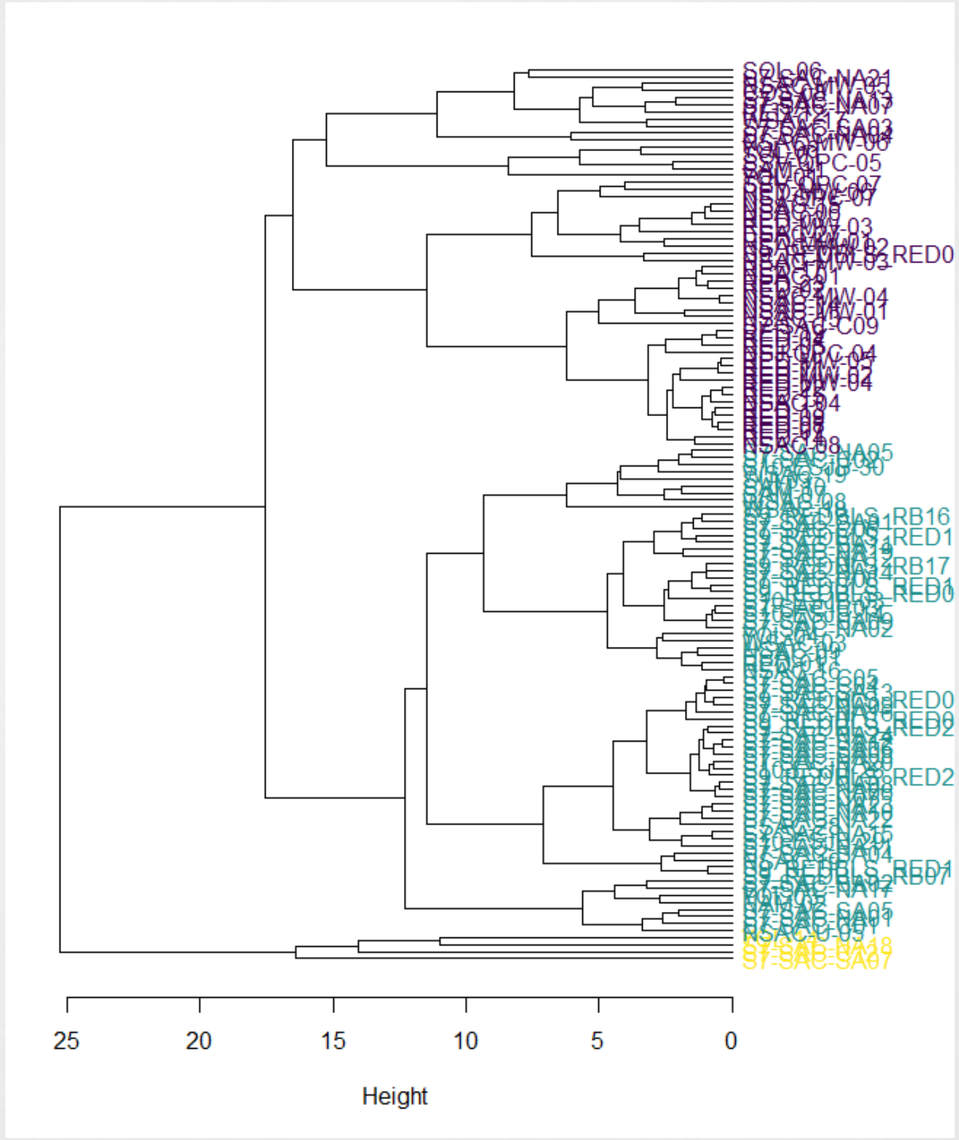


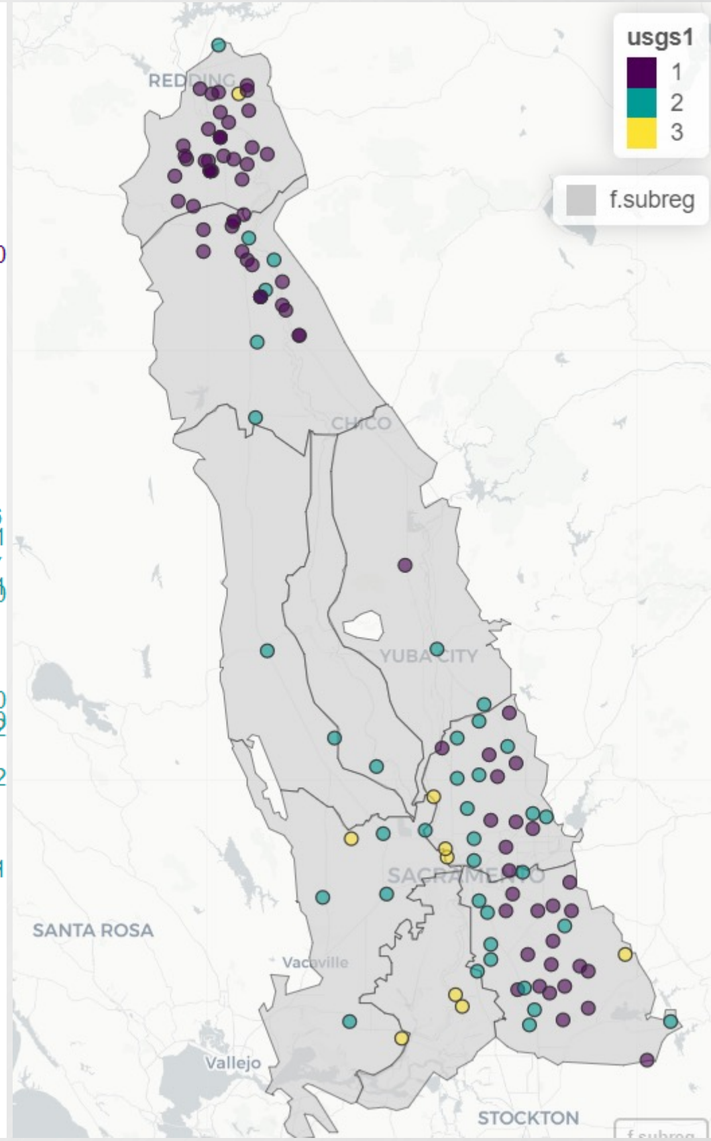


USGS data

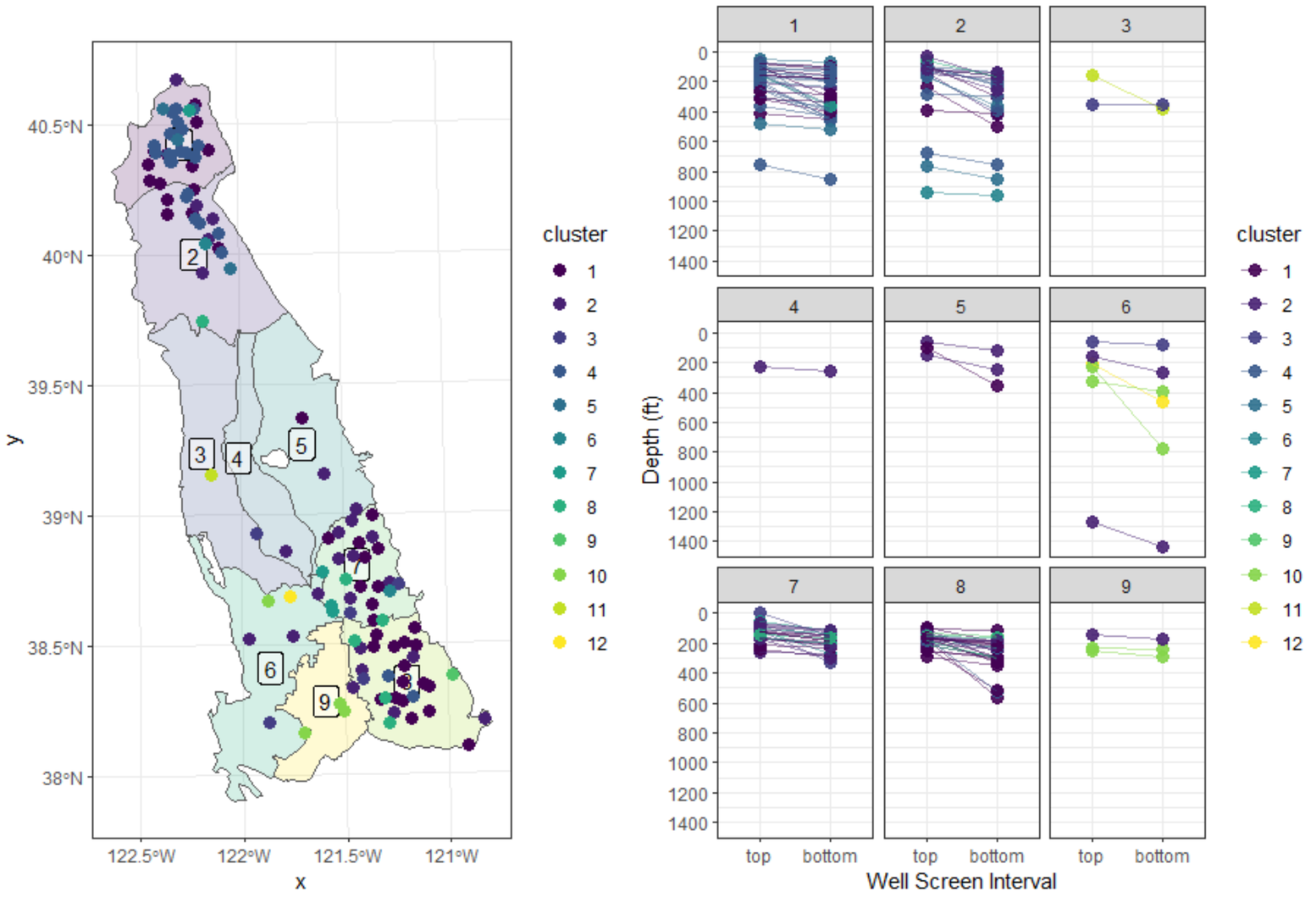
Hierarchical agglom.
Euclidean
Ward linkage
(cc = 0.64)
Z-score normalized

Ag	Cd	Mo	V
Al	Cr	Ni	Fe
As	I	Se	Mn
Be	Li	Sr	

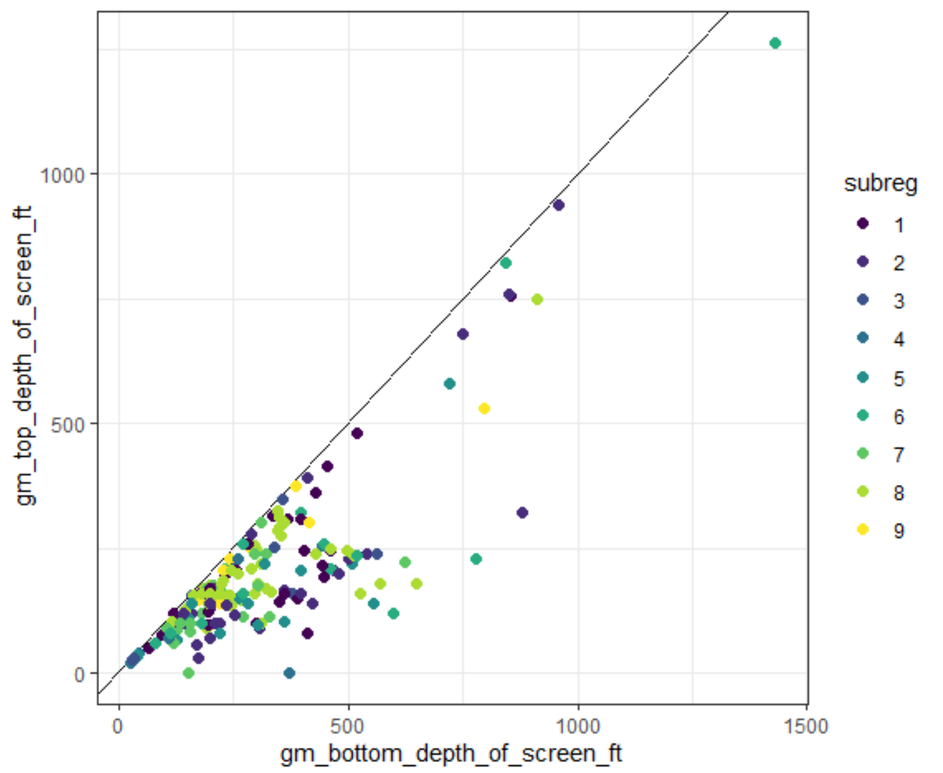




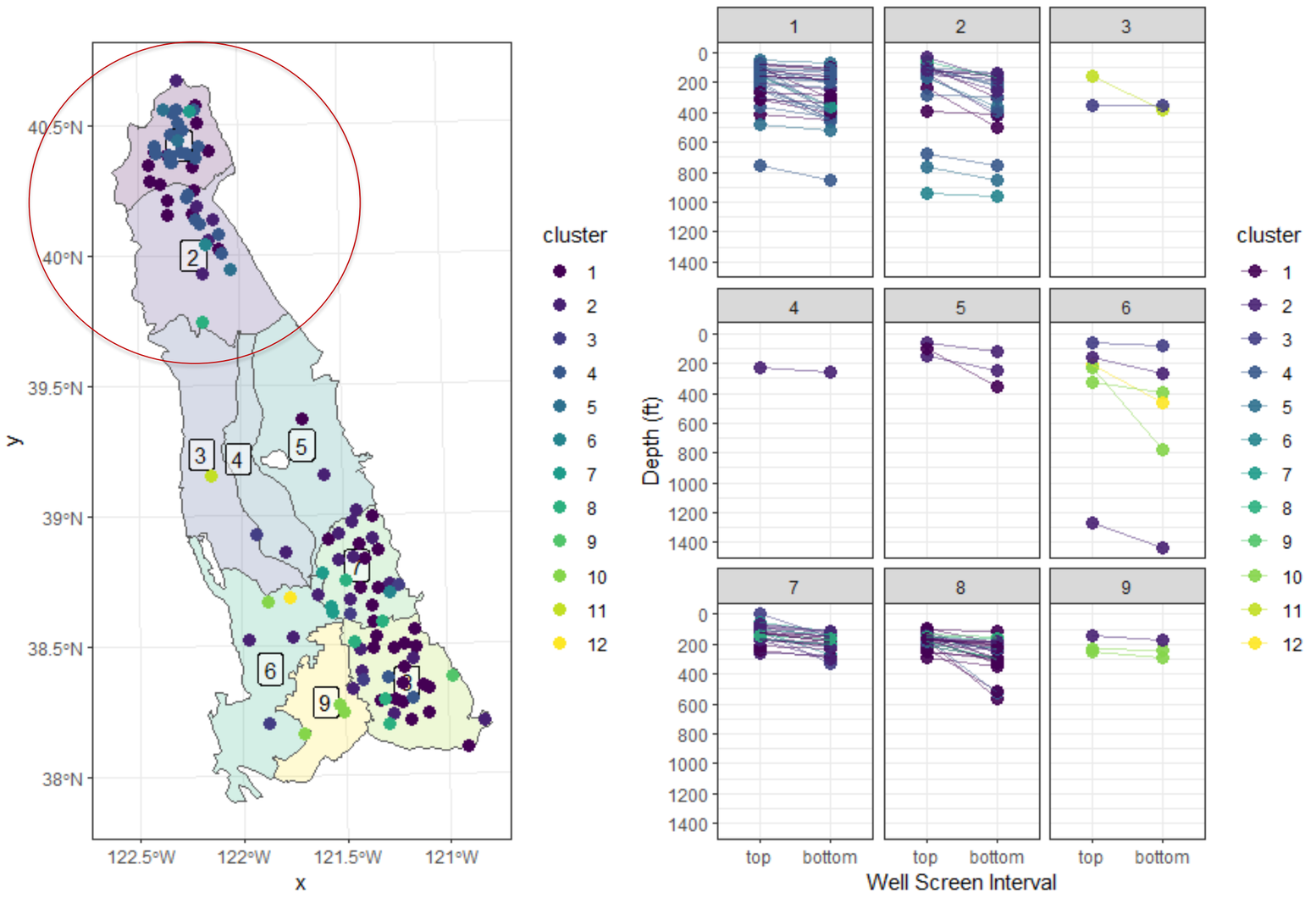
USGS data



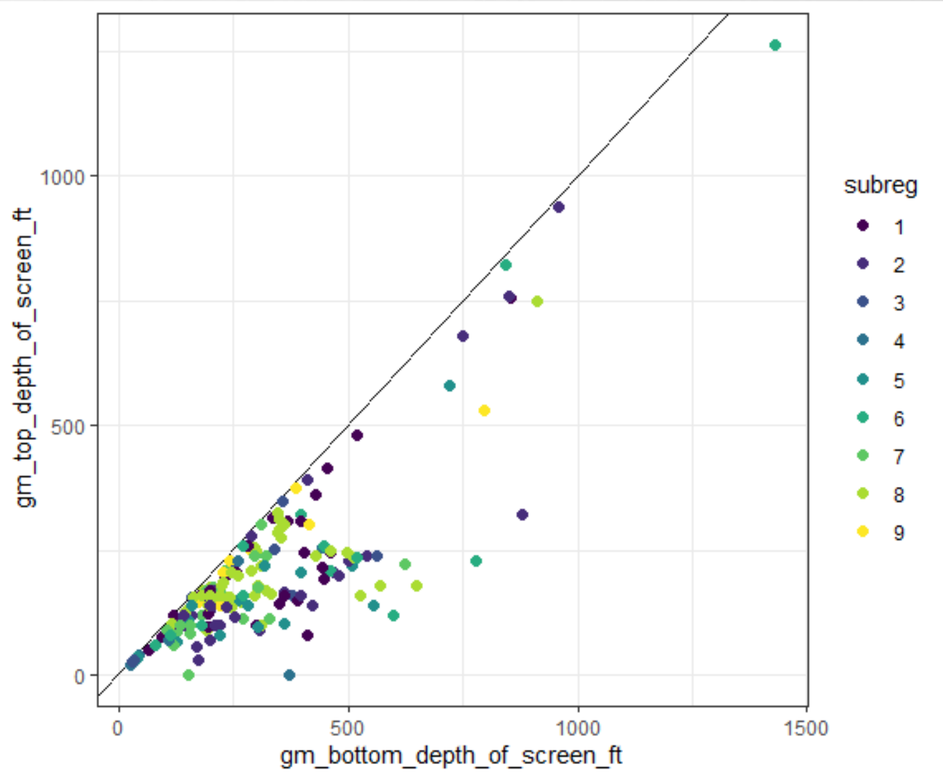
Now with 12 clusters



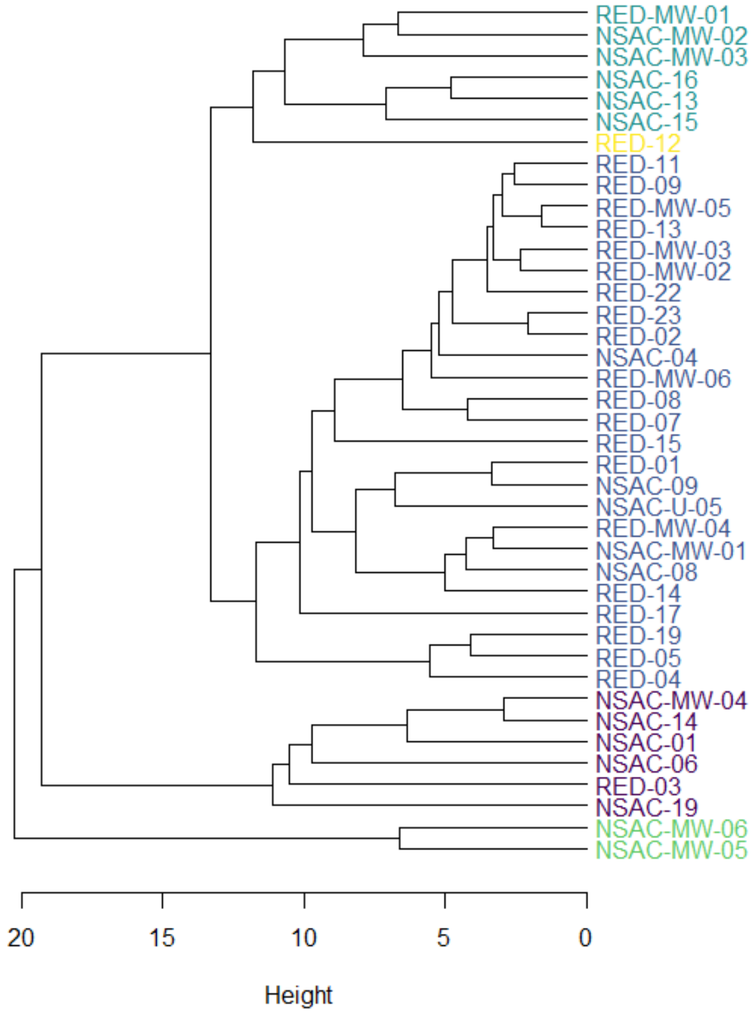
USGS data

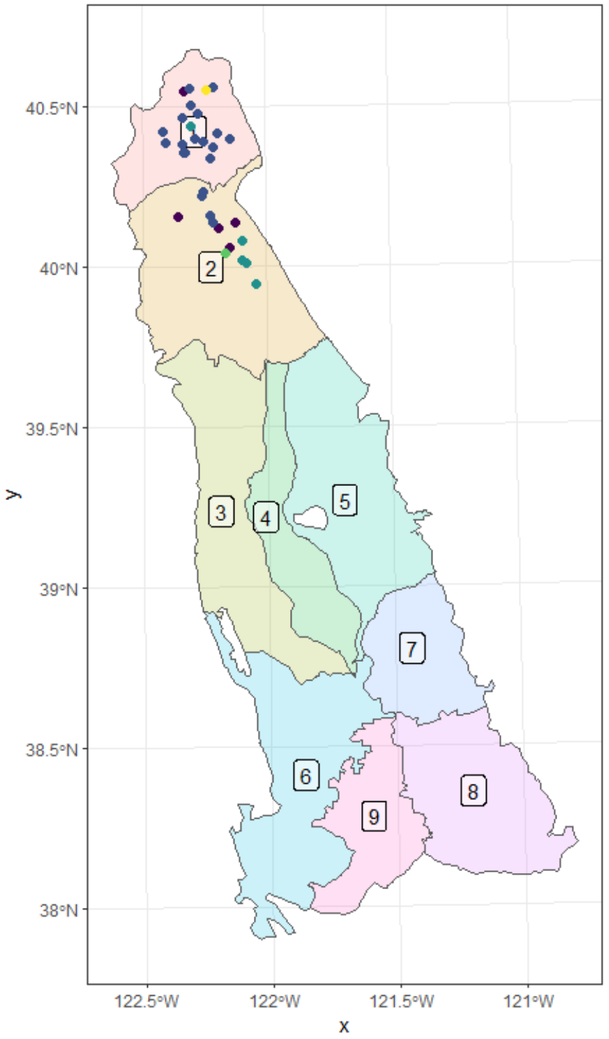


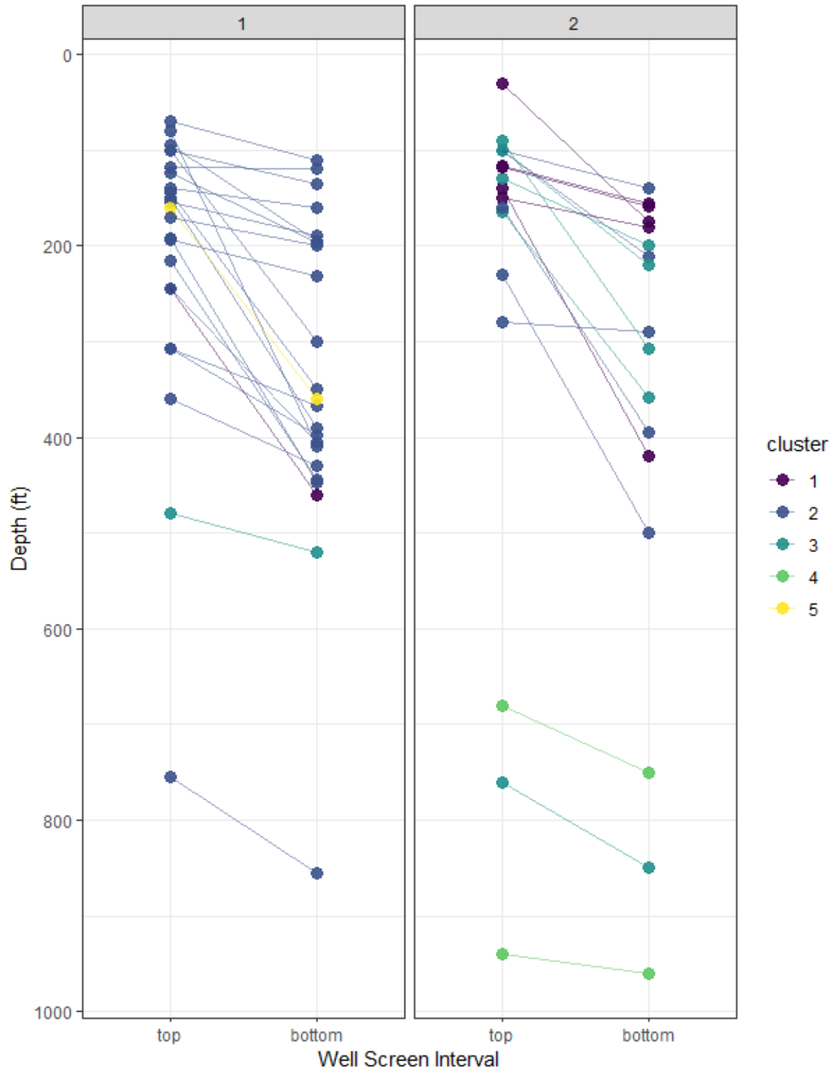
Now with 12 clusters



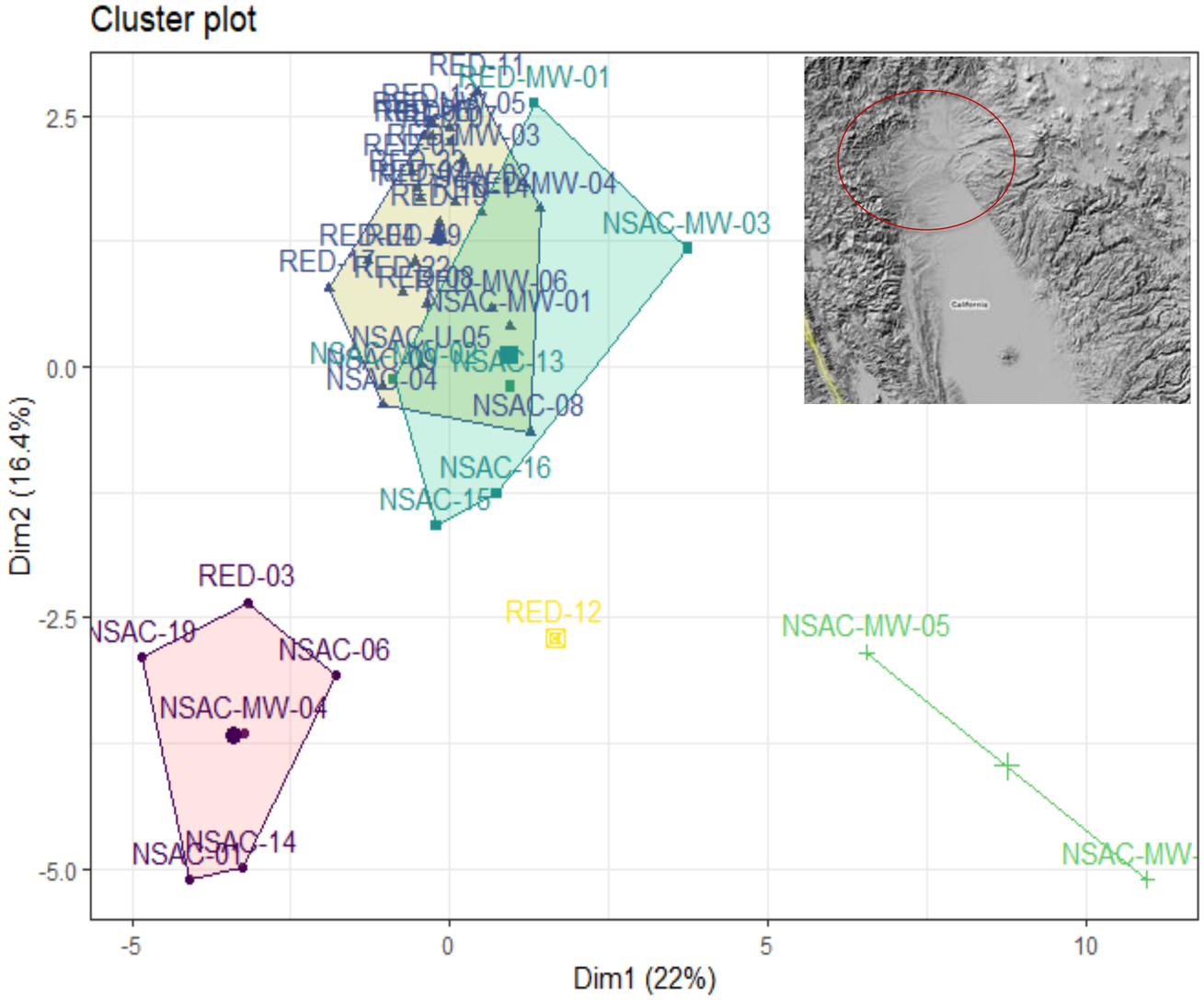
Subregions 1 & 2

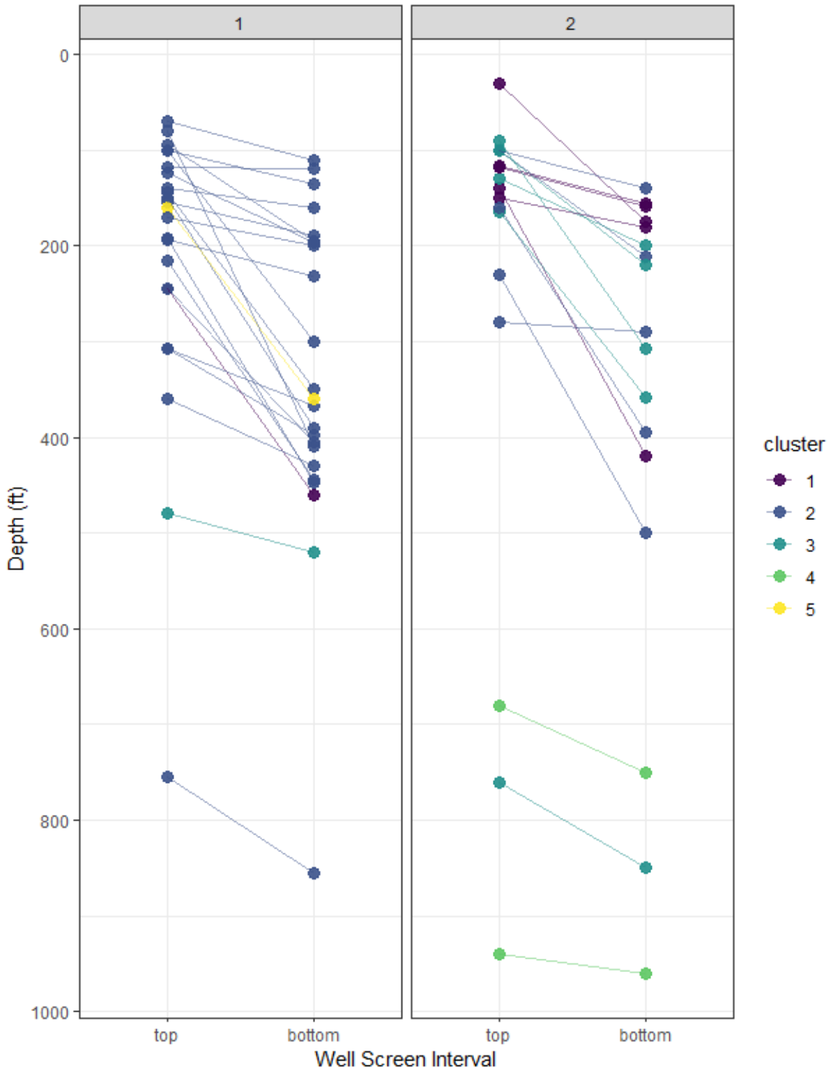




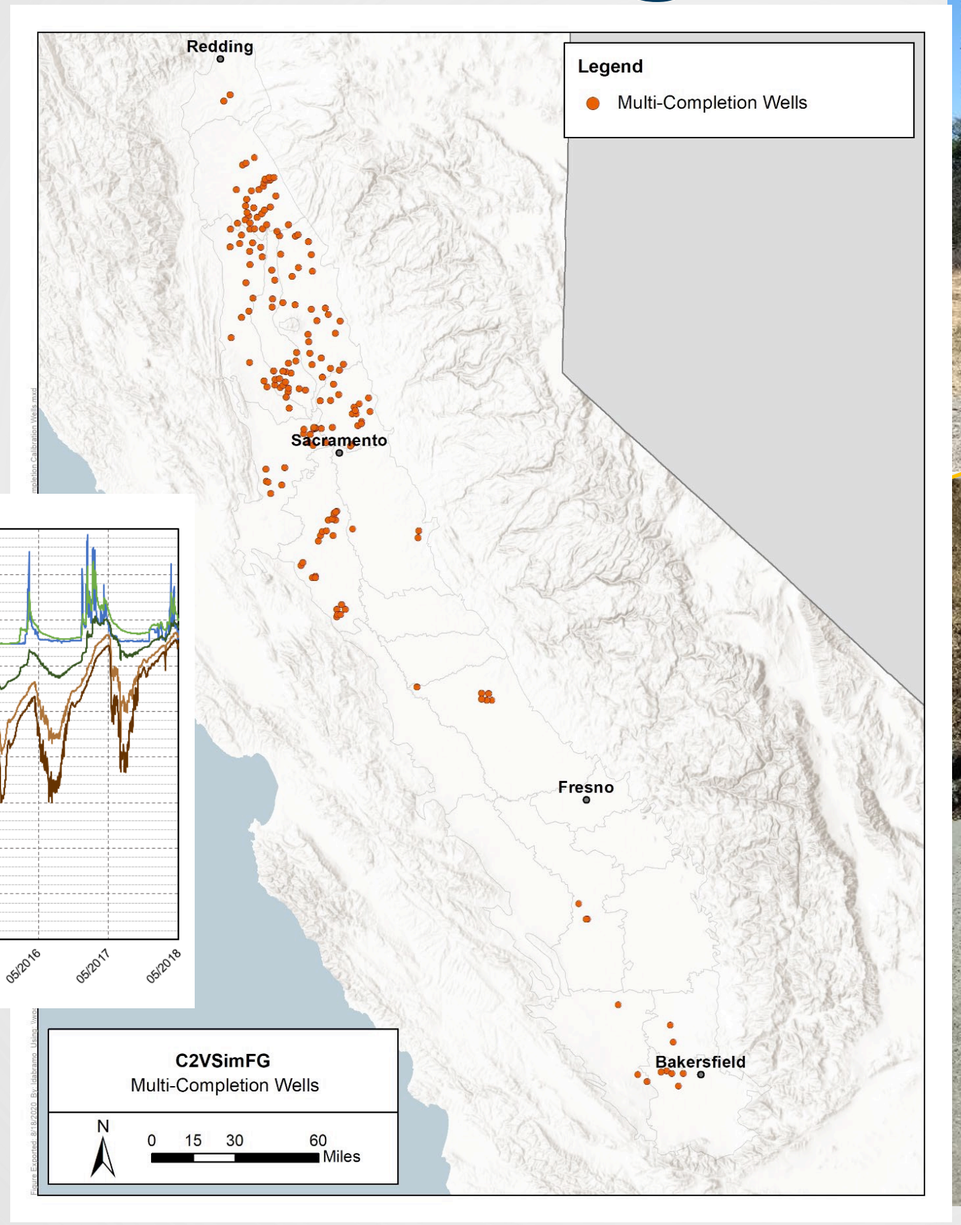


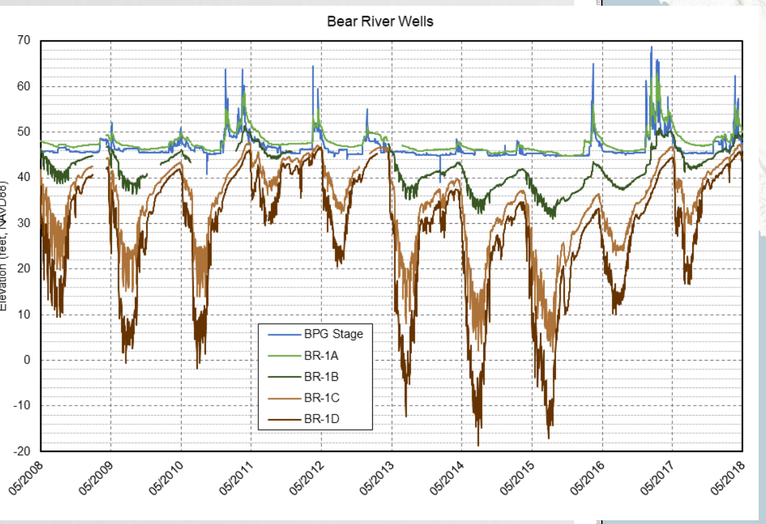
Subregions 1 & 2





Future monitoring





CALIFORNIA DEPARTMENT OF WATER RESOURCES
SUSTAINABLE GROUNDWATER MANAGEMENT OFFICE

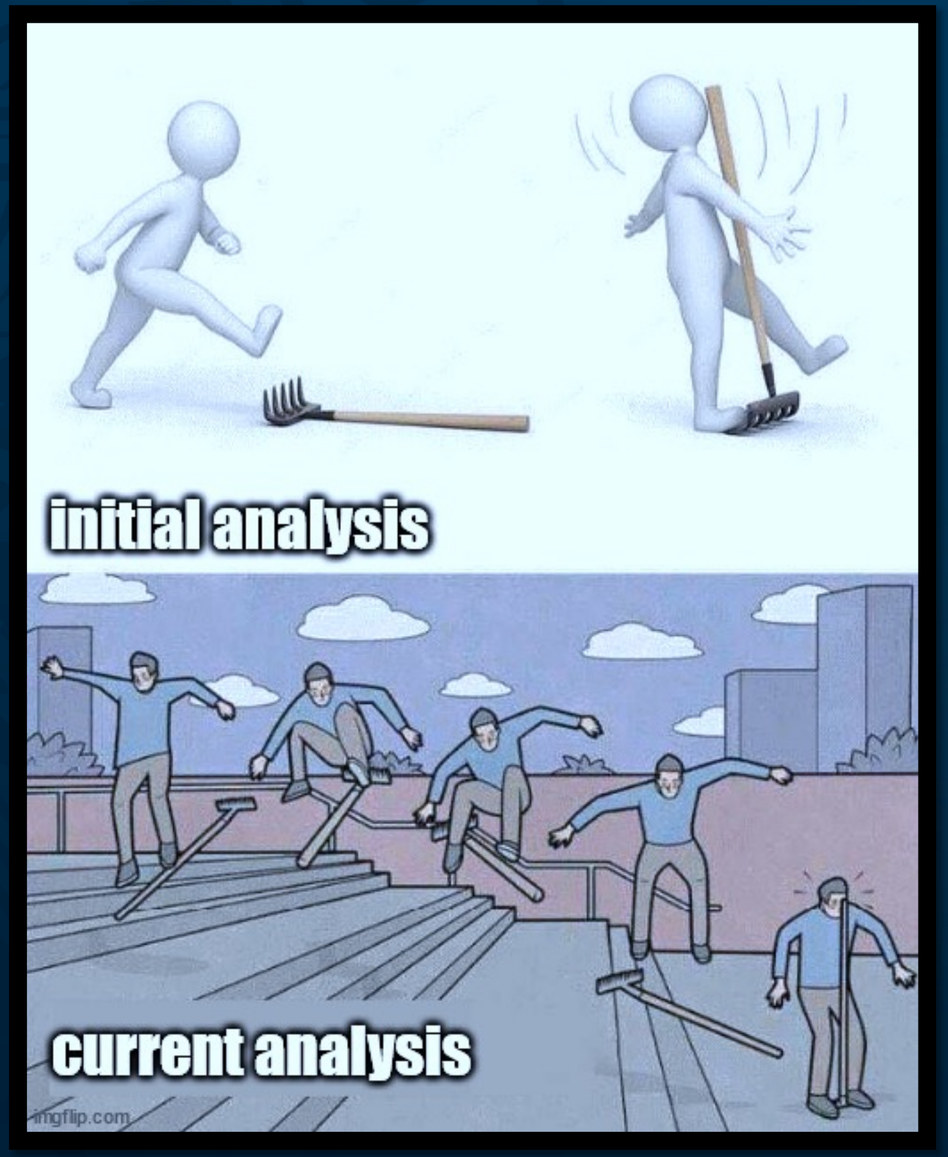
Lessons learned

- GAMA datasets complex
  - > 100 million individual observations, multiple agencies
  - ~350 constituents
    - 77 Naturally occurring vs. not (pesticides, PFAS)
  - Differing analytical methods reported
  - Ongoing-standardization
- Appropriate sets for subregions -> chemical + screen int.
- Complement to stratigraphy, AEM
- Water quality modeling / solute transport (salinity/TDS/SpC)

…and still learning

Kyle Hardage, Ph.D.

Kyle.Hardage@water.ca.gov



CALIFORNIA DEPARTMENT OF WATER RESOURCES
SUSTAINABLE GROUNDWATER MANAGEMENT OFFICE