

Regionalization of a California-wide seismic tomography model

UNIVERSITY OF MARYLAND
DEPARTMENT OF
GEOLOGICAL, ENVIRONMENTAL,
AND PLANETARY SCIENCES



Laurent Montesi¹; Arfan Rashid^{2,3}

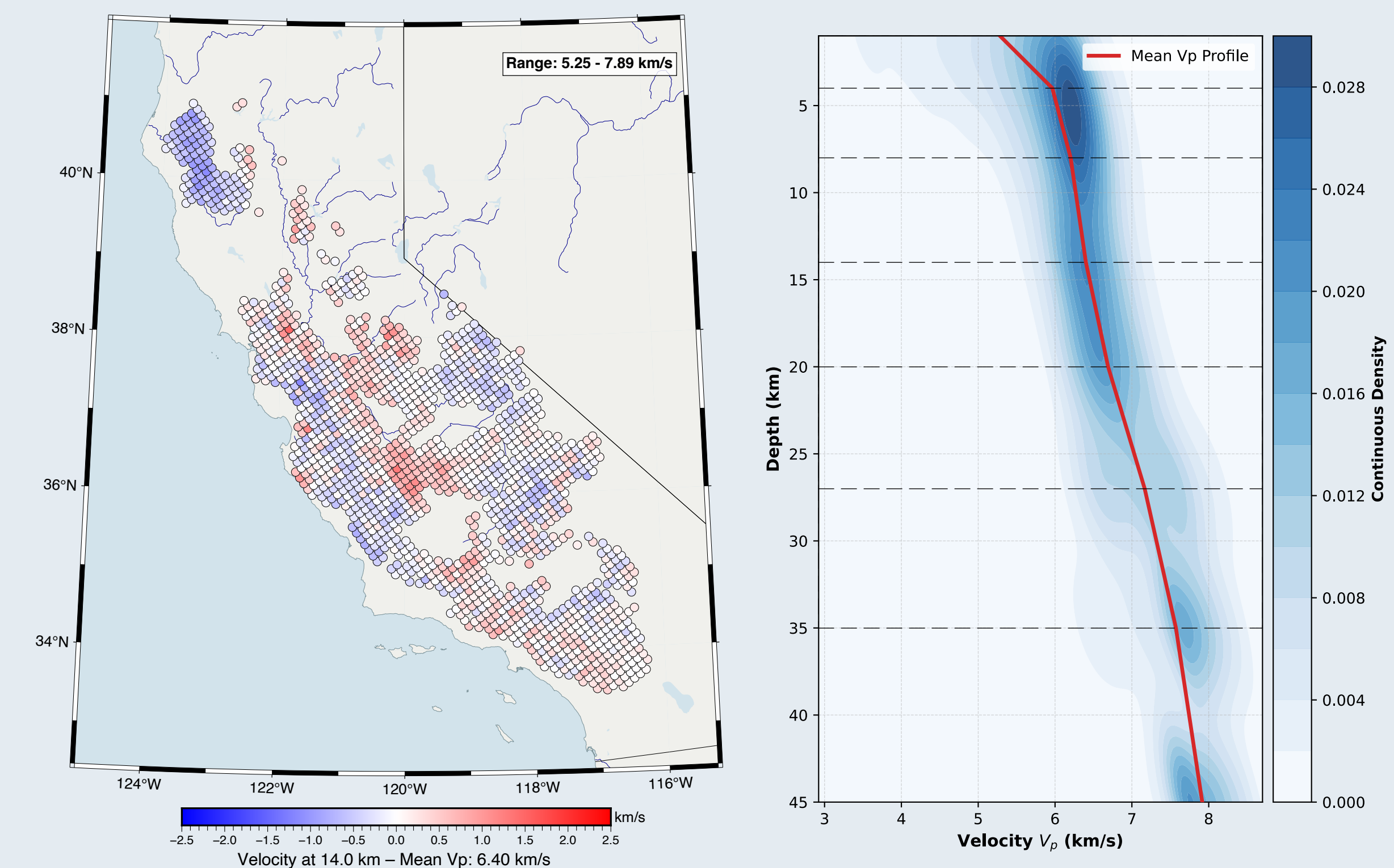
¹University of Maryland (montesi@umd.edu), ²Eleanor Roosevelt High School, Greenbelt, Maryland, ³now at Johns Hopkins University

Overview

The continental crust is composed of lithotectonic blocs with distinct properties. Geological reconstructions help constrain where these blocks are but provide limited information at depth. Seismic tomography, on the other hand gives a smoothed view of heterogeneities and isn't linked with unique rock types. Following the pioneering work of Eymold & Jordan (2019), we define distinct blocks in a regional tomographic model using clustering algorithms and discuss their correspondence with geological divisions.

Seismic topography dataset

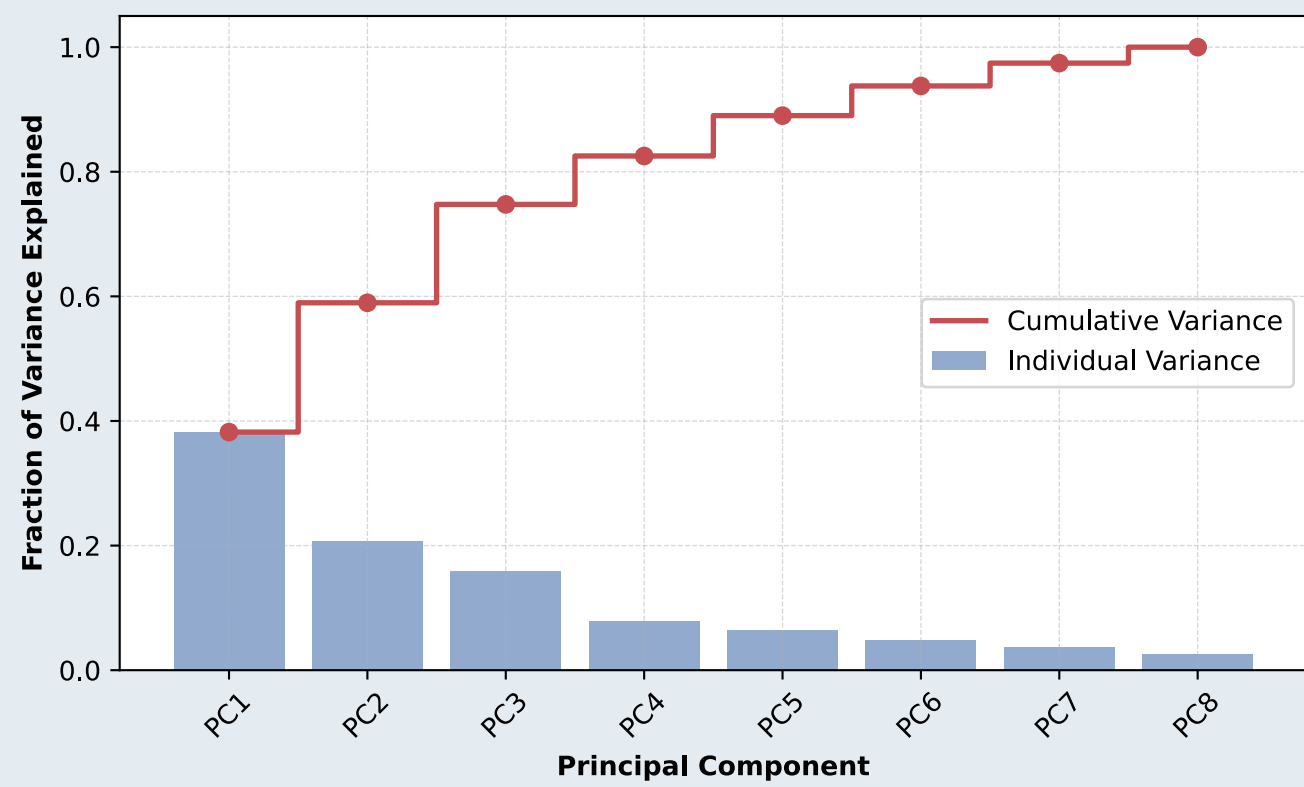
We use the statewide Vp uwlinca model (Lin et al., 2010). It contains Vp and DWS information at 8113 locations and 8 depth horizons (1, 4, 8, 14, 20, 27, 35, and 45 km). Unless specified, we restrict the dataset to the 1476 locations for which DWS exceeds 300 at all depths. We do not interpolate between location of depth slices and this would not introduce new seismological information.



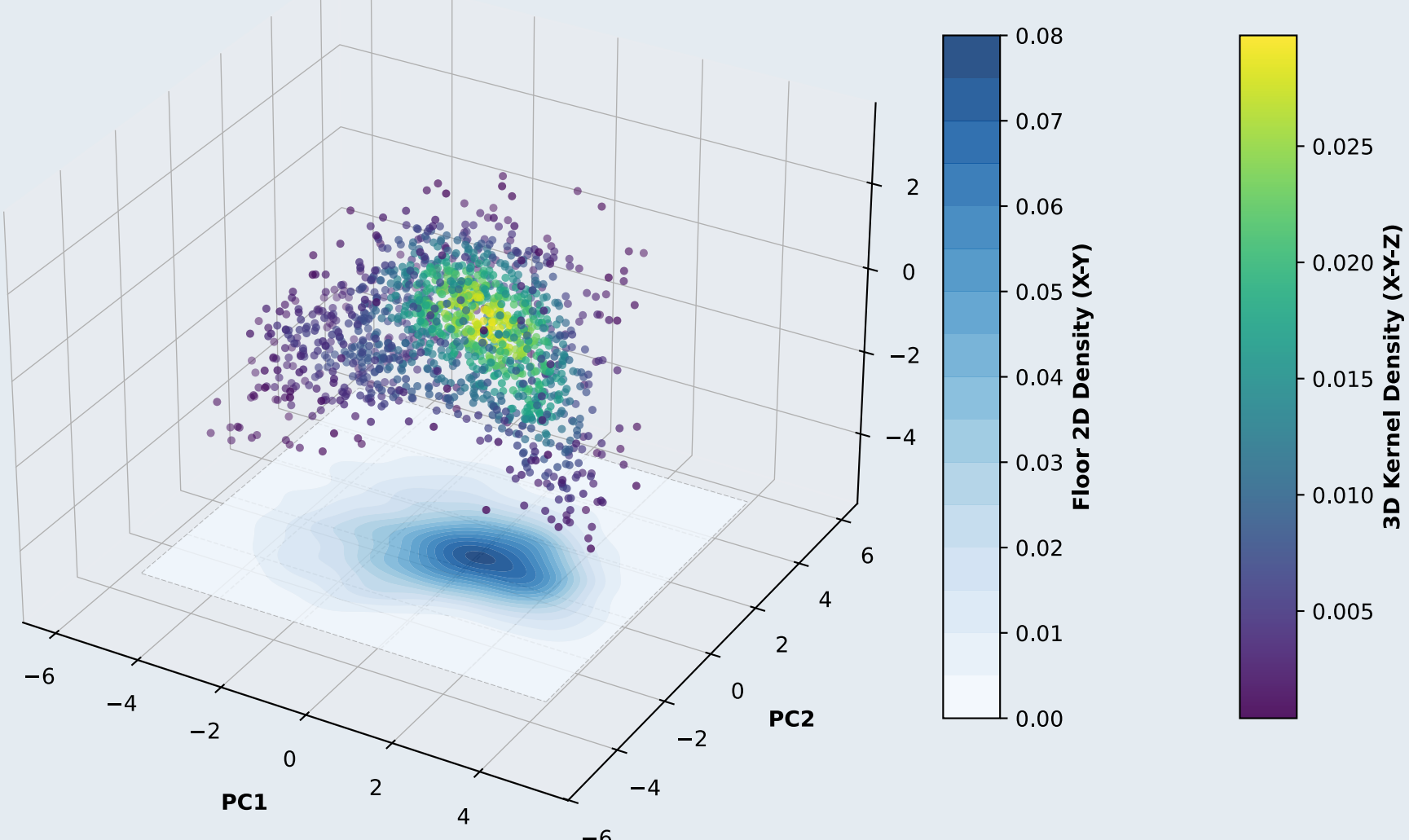
Map of the velocity anomaly at 8 km depth in uwlinca at all retained locations

Heat map of seismic velocity vs. depth in the retained dataset. Velocity is consistent at shallow depth with a diffuse tail to low velocity. No dominant value is present at 27 km depth.

The data is normalized at each depth (Z-score). This involves 1) removing an average Vp value to avoid layers with larger overall velocity from being weighted more heavily and 2) dividing by the standard deviation so that subtle variations in relatively uniform depth slices are considered in the clustering algorithm.



The analysis is conducted on the principal components of the seismic velocity profiles. PCA shows that 5 principal components are needed to capture 90% of the data variance. Unless specified, we ignore the remaining 3 principal components as they may represent noise.

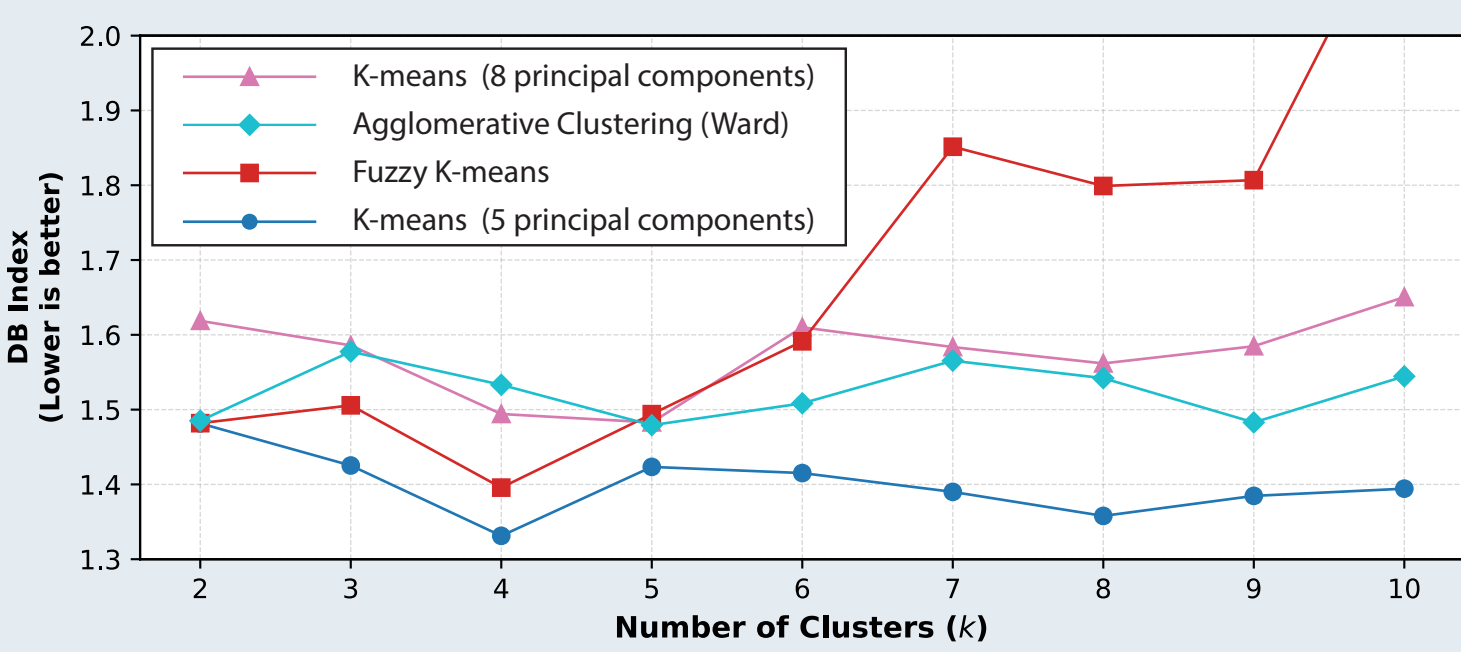


The principal components form a diffuse cluster elongated in some directions with no data break. We should not expect well-defined seismic velocities clusters but a range of valid solutions.

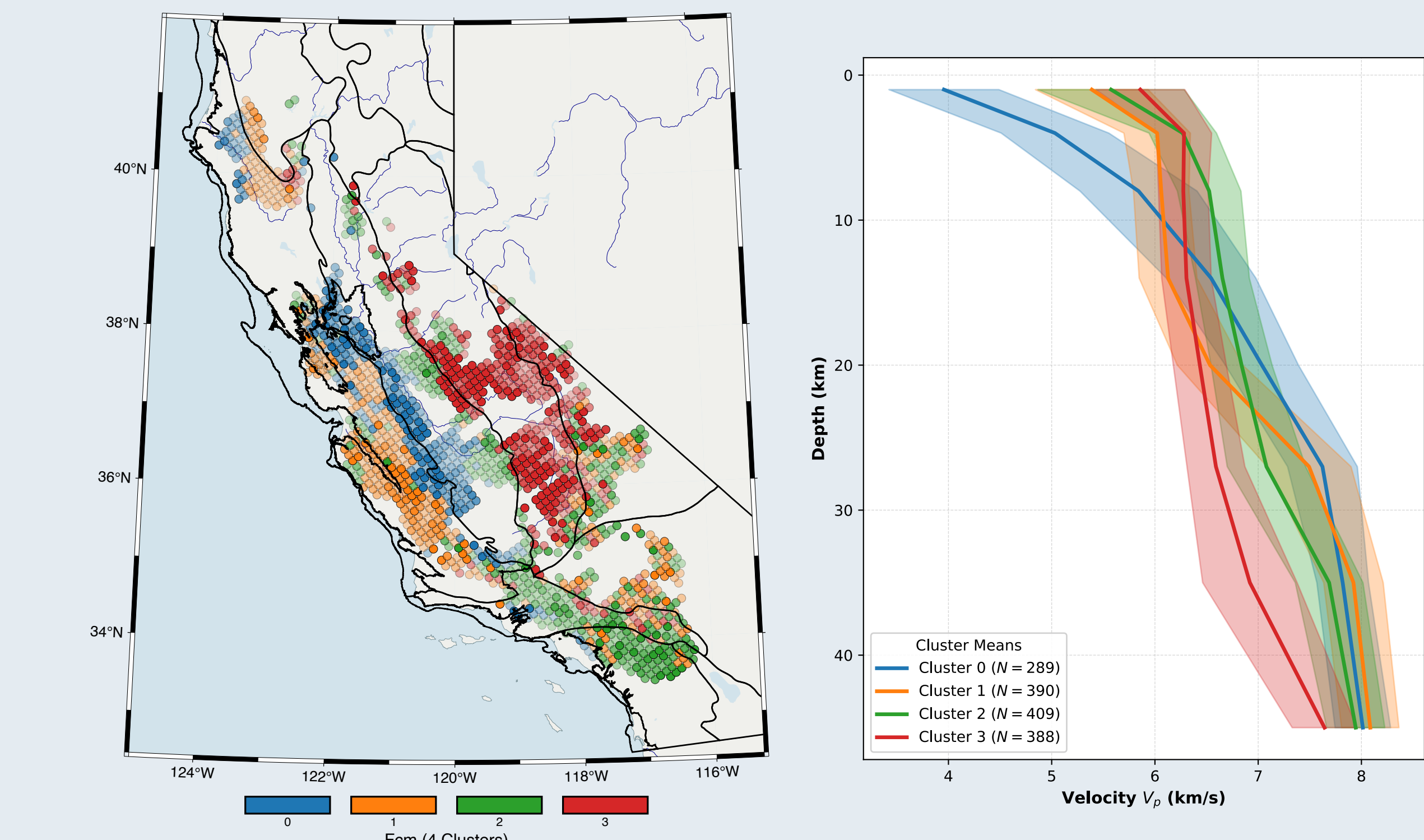
Clustering Solutions

We considered several clustering algorithms

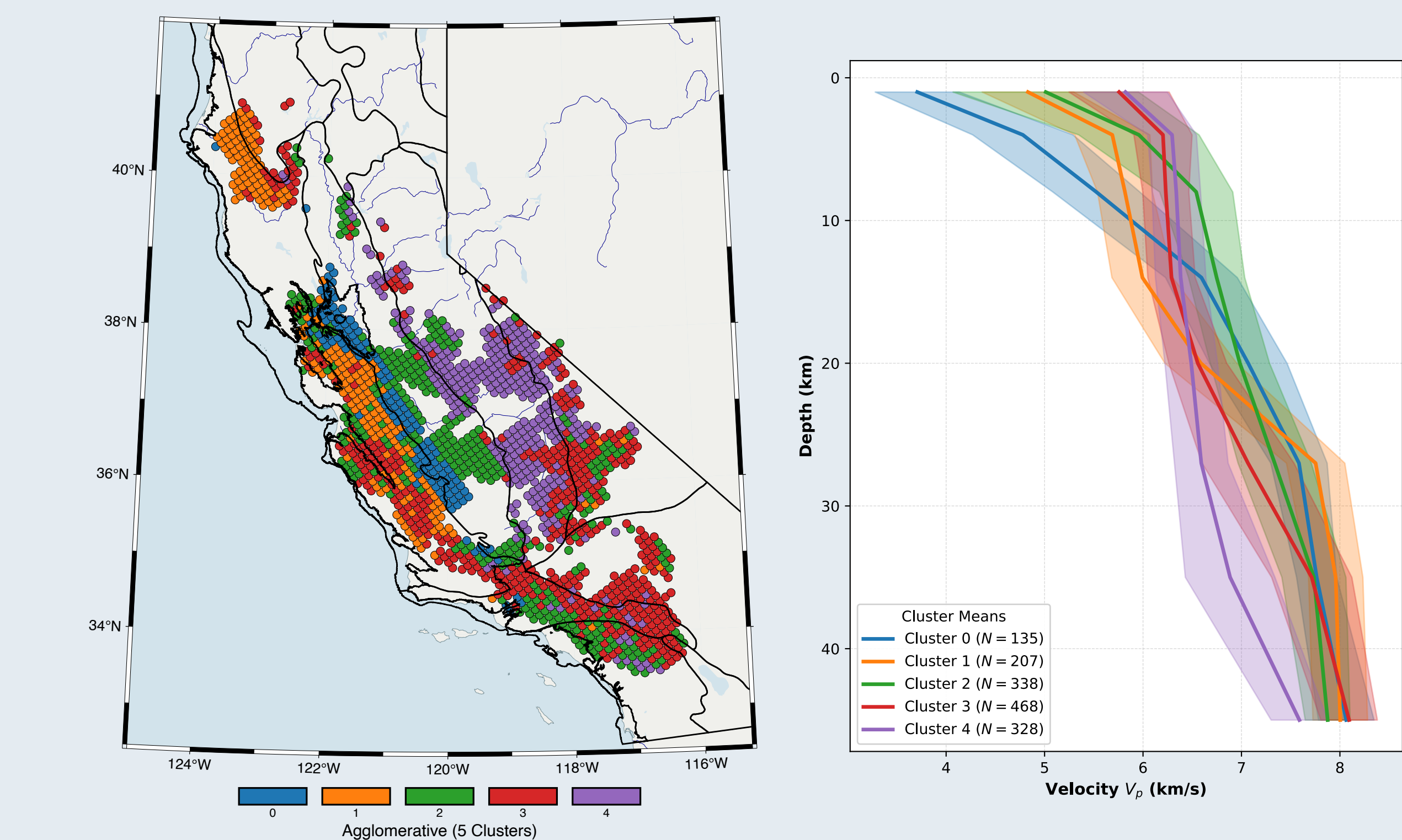
- Distance based algorithms
 - K-means: Points are assigned to the closest cluster centroid, which are redefined iteratively. The number of cluster must be specified a priori. We repeated K-means with 5 PCs and all 8 PCs.
 - Fuzzy K-means (FCM): Like K-means but each point is partitioned to every reach cluster based on the distance to the centroid.
 - Agglomerative clustering (Ward measure). Initially, each point is considered a clusters. Clusters are merged progressively based on the within-cluster variance.



The Davis-Bouldin Index proved to be the most useful diagnostic measure. Optimal solutions have 4 or 5 clusters. We illustrate below one example of each.



4-cluster solution obtained by FCM and average Vp profile in each cluster. On the map, transparency is linked to cluster membership entropy.



5-cluster solution obtained by Agglomerative Clustering (Ward method) and average Vp profile in each cluster.

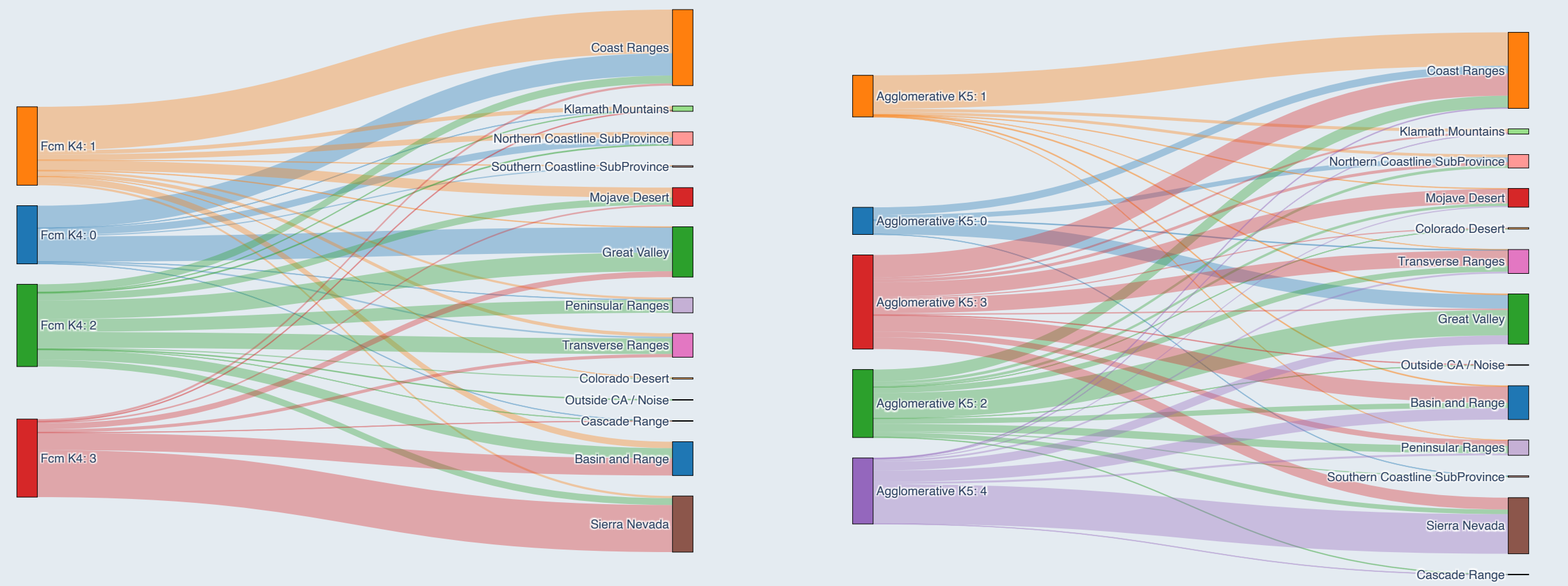
- Density-based algorithms.
 - DBSCAN: Each cluster is a gaussian blob with a high density core and a expending region. Some points are defined as noise but are not associated with any cluster. The solution depends on the minimum number of points required to form a dense region and eps, the local radius for expanding clusters.
 - HDBSCAN: Similar to DBSCAN but with a hierarchical structure of clusters merging into other clusters.

For each method, we report appropriate diagnostic measure as a function of number of clusters, whether specified a priori (K-means, FCM) or a result of the analysis (other methods). This information is used to identify the optimal solution for each clustering method.

Density-based algorithms return optimal solutions with only 2 clusters but 40 to 50% noise points. Therefore, we do not discuss these solutions further.

Geological Implications

In many cases, the seismic clusters correspond to geomorphologic provinces. The Sierra Nevada stands out consistently with a low velocity at 27 and 35 km depths. The Great valley is split in the East-West direction between the two clusters that have the greatest velocity gradients near the surface. The fuzzy cluster solution shows that this transition is gradual. The Coastal ranges and Peninsular Ranges define the remaining clusters.

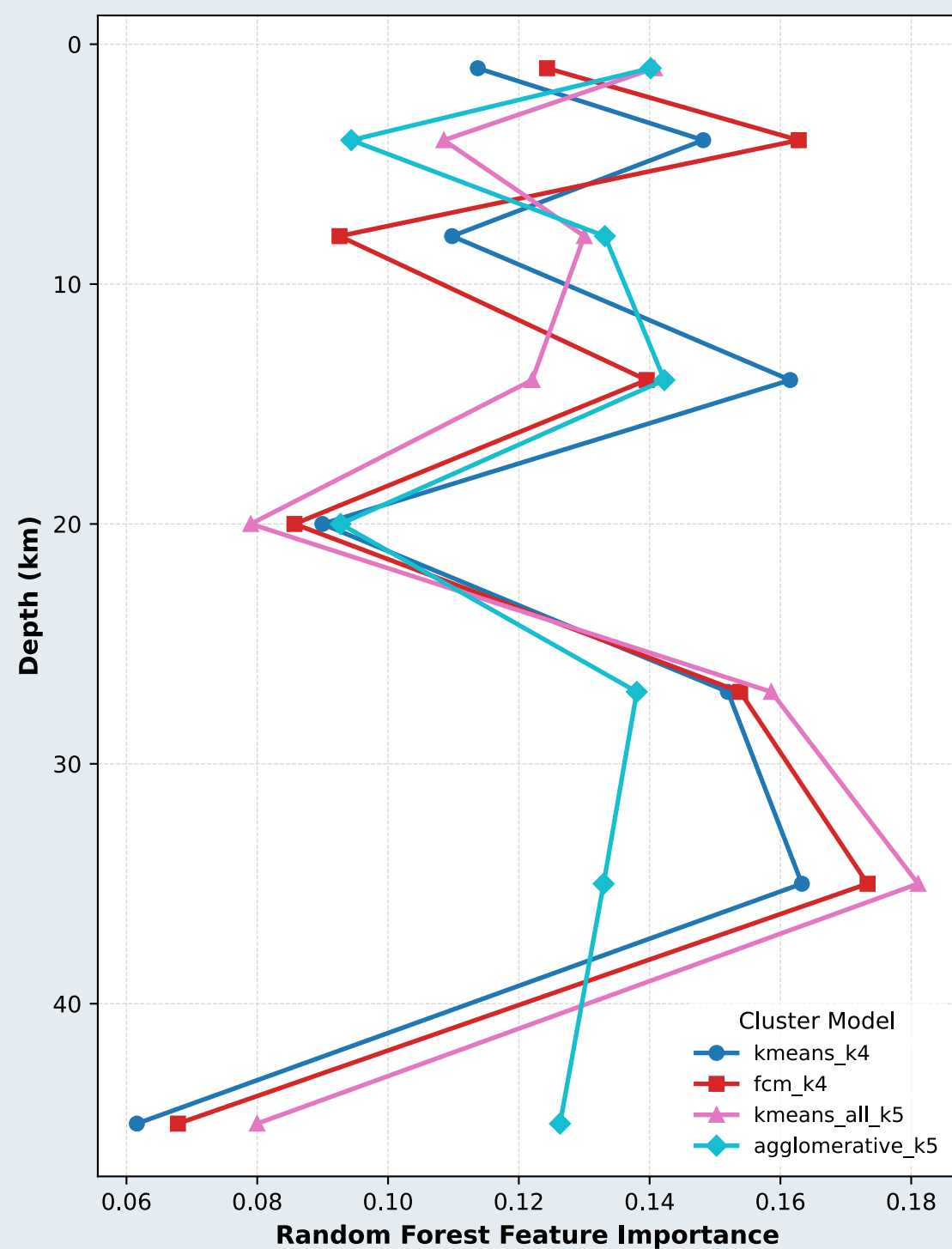


Correspondence between representative 4-cluster and 5-cluster solutions and California Geomorphic province.

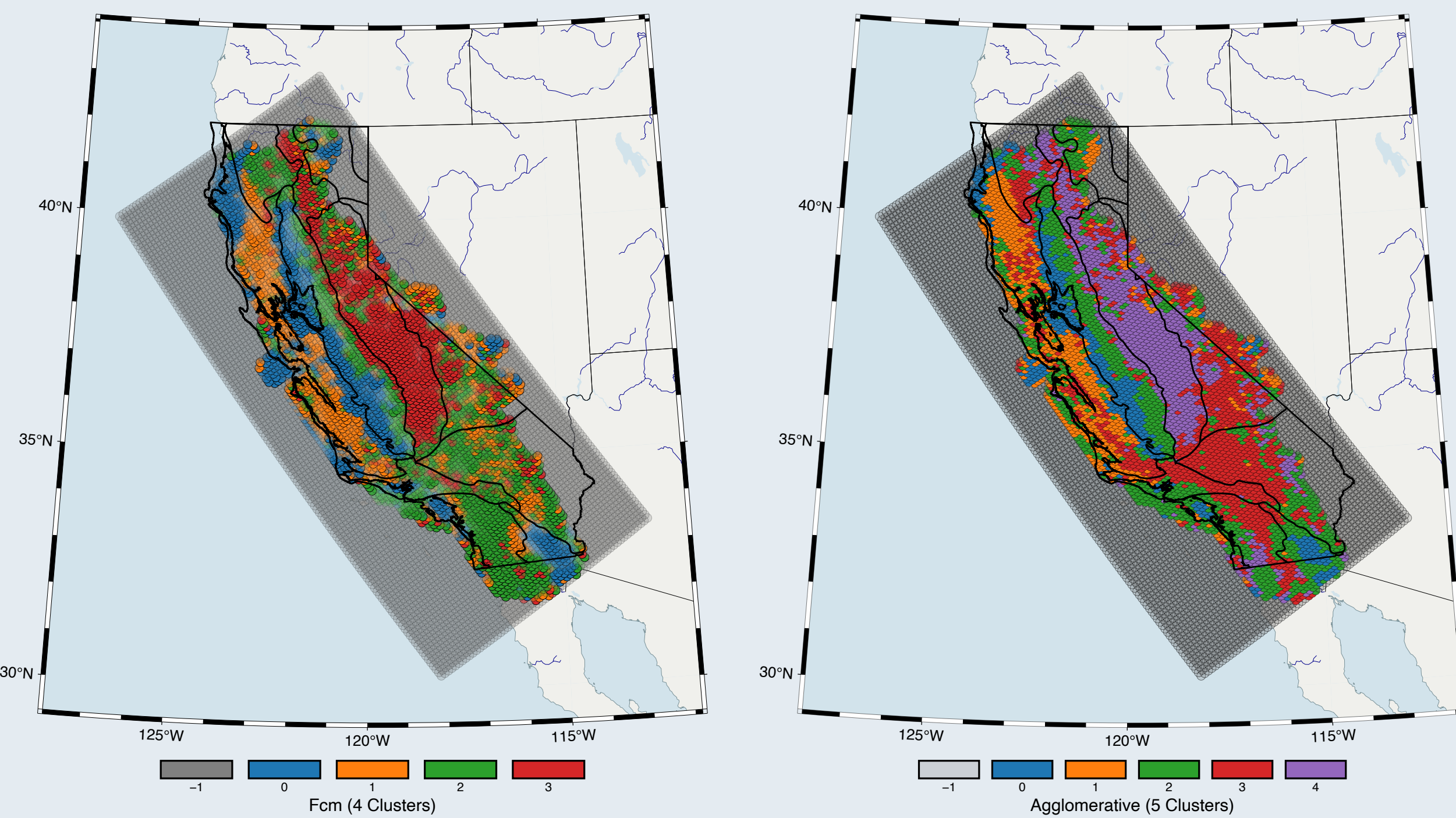
Although clustering is difficult in the absence of clear data breaks, the correspondence between these solutions and geomorphological provinces suggest that the identified clusters have geological meaning.

Decision trees and random forests help identify which depth intervals are key to defining clusters.

The 27 and 35 km depth intervals are important in each clustering solution, probably reflecting crustal thickness variations. The 14 km depth slice also appears as an important horizon. 4-cluster solutions utilize the 4-km depth but 5-cluster solutions are sensitive to the 1 and 8 km depth horizons instead, pointing to the importance of the upper crustal structure in the transverse ranges.



Generalization



The majority of the locations in the study area were ignored due to low DWS in at least one horizon. We can use the available information to infer membership to the clusters defined based on the complete profiles.

Doing so confirms that the entire Sierra Nevada is associated to a specific cluster and that the Peninsular ranges differ from the coastal ranges. The Basin-and-Range province and the Mojave desert appear similar to one another.

Acknowledgements

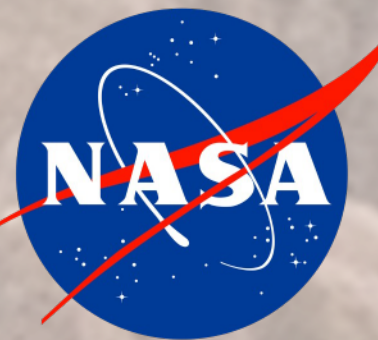
This research was supportrdd by the Statewide California Earthquake Center (Contribution No. 15452). SCEC is funded by NSF Cooperative Agreement EAR-2225216 and USGS Cooperative Agreement G24C0072. This project is hte basis of Arfan Rashid's Research Practicum, a year-long capstone project of the Science and Technology (S/T) Program at Eleanor Roosevelt High Schoole, Greenbelt, MD. We thank **Cliff Thurber** for sharing the uwlinca model an advice on data quality.



U.S. National
Science Foundation



U.S. DEPARTMENT OF
ENERGY



USC