Clustering Analysis Methods for GNSS Observations: A Data-Driven Approach to Identifying California's Major Faults

Robert Granat

City College of New York

Andrea Donnellan, Michael Heflin, Gregory Lyzenga, Margaret Glasscoe, Jay Parker

Jet Propulsion Laboratory, California Institute of Technology

Marlon Pierce and Jun Wang

Indiana University

John Rundle

University of California, Davis

Lisa Grant Ludwig

University of California, Irvine

Key Points

1. Unsupervised clustering methods provide a data-driven way of analyzing and partitioning GNSS observations of crustal deformation.
2. Deformation is distributed across the San Andreas fault system but is localized at the creeping section in central California.
3. The Southern San Andreas fault connects with the Eastern California Shear Zone rather than the rest of the San Andreas fault system.

Submitted to Earth and Space Science, 25 January 2021

Abstract

We present a data-driven approach to clustering or grouping Global Navigation Satellite System (GNSS) stations according to their observed velocities, displacements or other selected characteristics. Clustering GNSS stations has the potential for identifying useful scientific information, and is a necessary initial step in other analysis, such as detecting aseismic transient signals (Granat et. al., 2013). Desired features of the data can be selected for clustering, including some subset of displacement or velocity components, uncertainty estimates, station location, and other relevant information. Based on those selections, the clustering procedure autonomously groups the GNSS stations according to a selected clustering method. We have implemented this approach as a Python application, allowing us to draw upon the full range of open source clustering methods available in Python's scikit-learn package (Pedregosa et. al., 2011). The application returns the stations labeled by group as a table and color coded KML file and is designed to work with the GNSS information available from GeoGateway (Heflin et. al., 2020; Donnellan et al, 2021) but is easily extensible. We focused on California and western Nevada. The results show partitions that follow faults or geologic boundaries, including for recent large earthquakes and post-seismic motion. The San Andreas fault system is most prominent, reflecting Pacific-North American plate boundary motion. Deformation reflected as class boundaries is distributed north and south of the central California creeping section. For most models the southernmost San Andreas fault connects with the Eastern California Shear Zone (ECSZ) rather than continuing through the San Gorgonio Pass.

Index Terms

4302 Natural Hazards Geological

4306 Natural Hazards Multihazards

1998 Informatics Workflow

1294 Geodesy Instruments and Techniques

1209 Tectonic Deformation

Keywords

Clustering

Geodetic imaging

Tectonics

GNSS

Faults

Earthquake

1. Introduction

The use of unsupervised clustering methods for analysis and interpretation of geodetic data was discussed in (Granat, 2000), and the specific use of *k*-means and hierarchical agglomerative clustering methods to analyze GNSS velocity observations was presented in some detail in (Simpson et. al., 2012) and subsequent papers (Savage et. al., 2013; Thatcher et. al., 2016). Here we build upon this previous work to extend unsupervised clustering to GNSS observations of both velocity and displacement using a wide variety of clustering methodologies. This approach contrasts with that of (Meade and Hager, 2005) and (Johnson, 2013), which fit the California region GNSS velocities with translations and rotations of dozens of small pre-defined blocks based on prior mapped faults. Our method complements theirs, by exchanging rigid block definitions for data-driven clusters; their methods have an advantageous side product of reasonably determined fault slip rates at the boundaries.

We present a Python implementation of our approach, built upon the open source machine learning package scikit-learn (Pedregosa et. al., 2011), that provides considerable flexibility in experimental design. The software returns the GNSS stations labeled by group in both tabular form and as a color coded KML file for easy visualization in Google Earth or Google Maps. Our implementation is designed to work with the GNSS displacement and velocity information available from GeoGateway (Donnellan et. al., 2021), a map-based science gateway supported by NASA, but is easily extendable to other data sources or output formats. Velocities available from this source are estimated over the lifetime of each station, excluding detected breaks (Heflin et al, 2000).

2. Methods

The goal of any clustering method is to group the available observations such that the members of each group are objectively self-similar to each other while being different from members of all other groups. So-called *unsupervised* clustering methods attempt to do this with little or no *a priori* information about the "true" group membership. Our hypothesis is that there are natural groupings of crustal displacement reflective of underlying physical processes, and so to test we apply unsupervised clustering methods to GNSS observations.

When performing unsupervised clustering, certain key questions need to be answered. What is the metric used for self-similarity and cluster differentiation (e.g., Euclidean distance)? What features of the observations will be used, and will they be normalized (typically yes)? What is the expected number of clusters (usually referred to as "*k*"), or should the clustering method attempt to estimate the number? The answers to these questions inform our experimental design.

As a specific example, consider clustering cosesismic GNSS displacements using the *k*-means (Lloyd, 1982) algorithm. In *k*-means clustering, the algorithm attempts to

partition the n observations into k clusters in which each observation belongs to the cluster with the nearest mean. The final result is the partitioning of the observation space into Voronoi cells. The problem is computationally difficult (NP-hard), but efficient iterative heuristic algorithms have been developed that usually converge quickly to a local optimum (Elkan, 2003; Hammerly and Drake, 2015). When using k-means, our experimental setup is as follows:

1. Our similarity metric is Euclidean distance, set by our choice of algorithm.
2. We know that we need at least one partition of the data, but our choice of geographical area for the study also places a likely upper bound on the number of partitions, k_{max}. So we can run experiments for $k = 2,\ldots,k_{max}$.
3. We can include or exclude any direction of displacement observations for each station, although in general we usually include the horizontal while choosing to include or exclude the vertical direction depending on what sort of phenomena we wish to study.
4. We can choose to normalize the displacement observations with respect to one another; our implementation normalizes features to have zero mean and unit variance. Depending on the data, this may enhance the influence of small displacements on the resulting groupings.
5. We can include or exclude the geographical position information (latitude, longitude) of each station. Including geographical information biases the algorithm towards creating spatially continuous clusters, while excluding it means clusters are based solely on displacement. If we decide to include geographical information, we *must* normalize the observation vectors so that each feature can be considered unitless.

Since our software implementation is based on scikit-learn, we were able to easily include the following clustering options for use by investigators:

1. *K*-means clustering.
2. Affinity propagation clustering.
3. Mean-shift clustering.
4. Spectral clustering.
5. Agglomerative clustering.
6. DBSCAN.
7. Gaussian mixture modeling.
8. Variational inference Gaussian mixture modeling.

Wherever possible, we provide the user with reasonable default values for the necessary parameters used by each algorithm, although users may pass their own parameter values using optional command flags.

3. Experimental Results and Discussion

We carried out a series of experiments using this approach designed to identify major crustal deformation boundaries in California. We used as our experimental input velocity and displacement observations calculated using the data from the Global Navigation and Satellite Systems (GNSS) processed using the GIPSY software package (Heflin et al, 2020). Our velocity clustering results for California are in good general agreement with those presented previously in (Thatcher et. al., 2016), as shown in Figure 1. However, we extend our analysis to include several classes for long-term velocities, cosesismic offsets, post-seismic motions, and specific time intervals.

We employed two different clustering algorithms, *k*-means (discussed above) and Gaussian mixture modeling (McLachlan and Peel, 2004) for more robust results, and stepped up the number of classes for each from two to ten. We found that when we included horizontal and vertical motions in the analysis the results were a mixture of subsurface fluid motions that dominate the vertical motions and tectonic motions that

dominate the horizontal motions. Since we were primarily interested in tectonic motion in these experiments, we excluded vertical motion from our study to avoid contamination by strong subsidence signals in the vertical direction. The software allows for station geometry of the stations to be included or excluded in the analysis. This biases the results so that stations within a cluster tend to be adjacent. After performing several experiments, we discovered that for this data set, inclusion of the station locations caused the geometry to dominate the clustering results. We note that plate boundaries naturally produce fault parallel extended clusters, which would never be contiguous when geometry is included in the feature set. As a result, we excluded geometry information so as not to obscure the signal arising from motion.

We performed the following experiments perfomed. For the results presented here we used a Gaussian Mixture Model for horizontal components only with no geometric constraint. We solve for 2-10 classes for each. We used GeoGateway (Donnellan et al, 2021; Heflin et al, 2020) to compute a table of values for the following using a bounding box surrounding California (Latitude 37.1903, width 9.5335°, Longitude -119.1311, width 11.3379°).

1. Long-term velocities
2. Coseismic motions for the 2010 Mw7.3 El Mayor - Cucapah Earthquake (EMC; Rymer et al, 2010).
3. Coseismic motions for the 2019 Mw 6.4 and Mw7.1 Ridgecrest Earthquake Sequence (R; Liu et al, 2019)
4. Postseismic motions for the 2010 Mw7.3 El Mayor - Cucapah Earthquake for 18 days to 2 years after the event.
5. Postseismic motions for the 2019 Mw 6.4 and Mw7.1 Ridgecrest Earthquake Sequence for 17 months after the events (July 22, 2019 - December 1, 2020).

6. Modeled motions from 1 January 2000 - 1 April 2010 or before the El Mayor - Cucapah earthquake. Modeled motions remove annual terms and jumps in the time series.
7. Modeled motions from 10 April 2010 - July 1, 2019 or between the El Mayor - Cucapah and Ridgecrest earthquakes.
8. Modeled motion from 10 July 2019 - 1 December 2020 or after the Ridgecrest earthquake sequence.

In classifying long-term velocities there is one boundary for k=2, which generally follows the San Andreas fault system (Figure 1). The Eastern California Shear Zone (ECSZ) is identified when k≥3 or for three or more classes with two or more boundaries suggesting that the San Andreas fault system is the dominant Pacific - North American plate boundary and that the ECSZ is the next most dominant. We hand drew boundaries between classes k=2-10 (Figure 2). Overlaying all of the boundaries shows a distributed set of boundaries sub-parallel to the San Andreas fault system, the ECSZ/Owens Valley/Walker Lane and a north-south boundary further east, which may mark the stable North American Plate. The location of the latter two (eastern) boundaries is consistent and irrespective of the number of classes. Aside from two off-shore zones the class boundaries concentrate near the Salton Trough in the southernmost part of California. The Central Valley forms its own cluster. Cluster boundaries are less consistent in the southern Central Valley, but we attribute this in part of horizontal motion induced by subsidence.

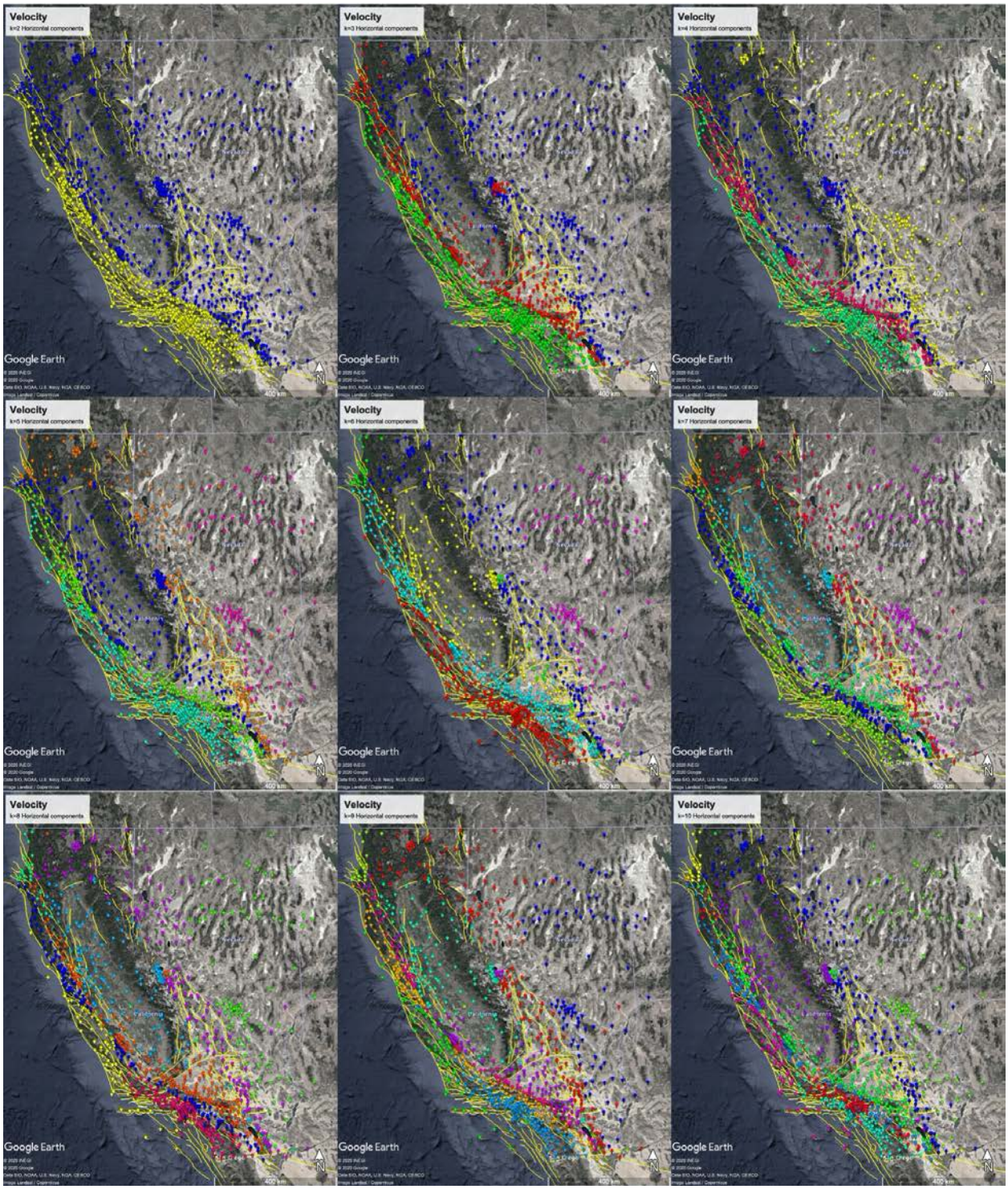

Figure 1. Clusters of velocities for number of clusters k = 2 to 10 starting row wise at top left. Yellow lines indicate faults from the Uniform California Earthquake Rupture Forecast model (UCERF-3; Field et al, 2014).

The boundary for k=2 is straighter than the San Andreas fault, likely indicating that it is controlled by deeper plate tectonic motions. From south to north, the boundary begins at and follows the Imperial fault, then follows the western side of the Salton Sea passing west of San Gorgonio Pass and east of Cajon Pass. From there the boundary follows the San Gabriel section of the San Andreas fault ~6 km north, passing east of Tejon Pass, and following the western edge of the San Joaquin Valley. The boundary joins the San Andreas fault at the south end of the Parkfield section and extends northwest through the creeping section. The boundary then joins the Calaveras fault veers slightly east in San Pablo Bay and joins the Maacama fault. The Maacama fault turns northward near Cloverdale as does the boundary, which makes a sharper northward bend and jogs east and north of Redwood Valley and Ukiah. At the north end of the Maacama fault the boundary joins the Garberville - Briceland fault before turning westward to the Pacific coast shore. Boundaries converge at the creeping section of the San Andreas fault regardless of the number of classes. The boundaries fan outward at the north and south end of the creeping section into sub-parallel boundaries. We interpret this to indicate distributed deformation north and south of the creeping section of the San Andreas fault with deformation concentrated at the creeping section. A set of boundaries extends south through offshore Channel Islands parallel to the coast.

The eastern margin of the Pacific-North American plate boundary is present for k≥5 and the location of the boundary is consistent for k=5-10. Originating in the Salton Trough the boundary extends northward from Bombay beach at the southernmost San Andreas fault. It extends northward east of the Mojave Block (Glazner et al, 2002), through the mountains east of Death Valley and into Nevada. The boundaries turn northward to join the 1954 Dixie Valley earthquake sequence ruptures (Caskey et al, 1996) and 1915 Pleasant Valley earthquake rupture (Jones 1915).

The ECSZ emerges at Bombay beach, departing from the San Andreas fault and trending northward through the Mojave block in a zone bounded by the 1992 M7.3 Landers (Hauksson et al, 1993) and 1999 M7.1 Hector Mine (Behr et al, 2000) earthquake ruptures. Rather than follow the Blackwater fault as previously suggested (i.e. Peltzer et al, 2001), the boundaries follow the Goldstone Lake fault (Glazner et al., 2002; Schermer et al., 1996) or Paradise fault 33 km to the east of the Blackwater fault, and joins with the Wilson Canyon fault about 20 km east of the 2019 M7.1 Ridgecrest rupture. From there it extends northward along the Owens Valley and Walker Lane. The boundaries divert around a 30–35 km radius area spanning the Mono Craters between Mammoth Lakes and Mono Lake. They continue along the Eastern Sierra to Carson City. Just north of Lake Tahoe the class boundaries follow a northwest trend to Mount Lassen and then continue west along the northern boundary of the Great Valley. Most of the boundaries converge at the coast due east of the Mendocino Triple Junction (Silver, 1971). North of Mendocino there is a boundary that separates the northern coast of California from the Coast Ranges 50 km inland.

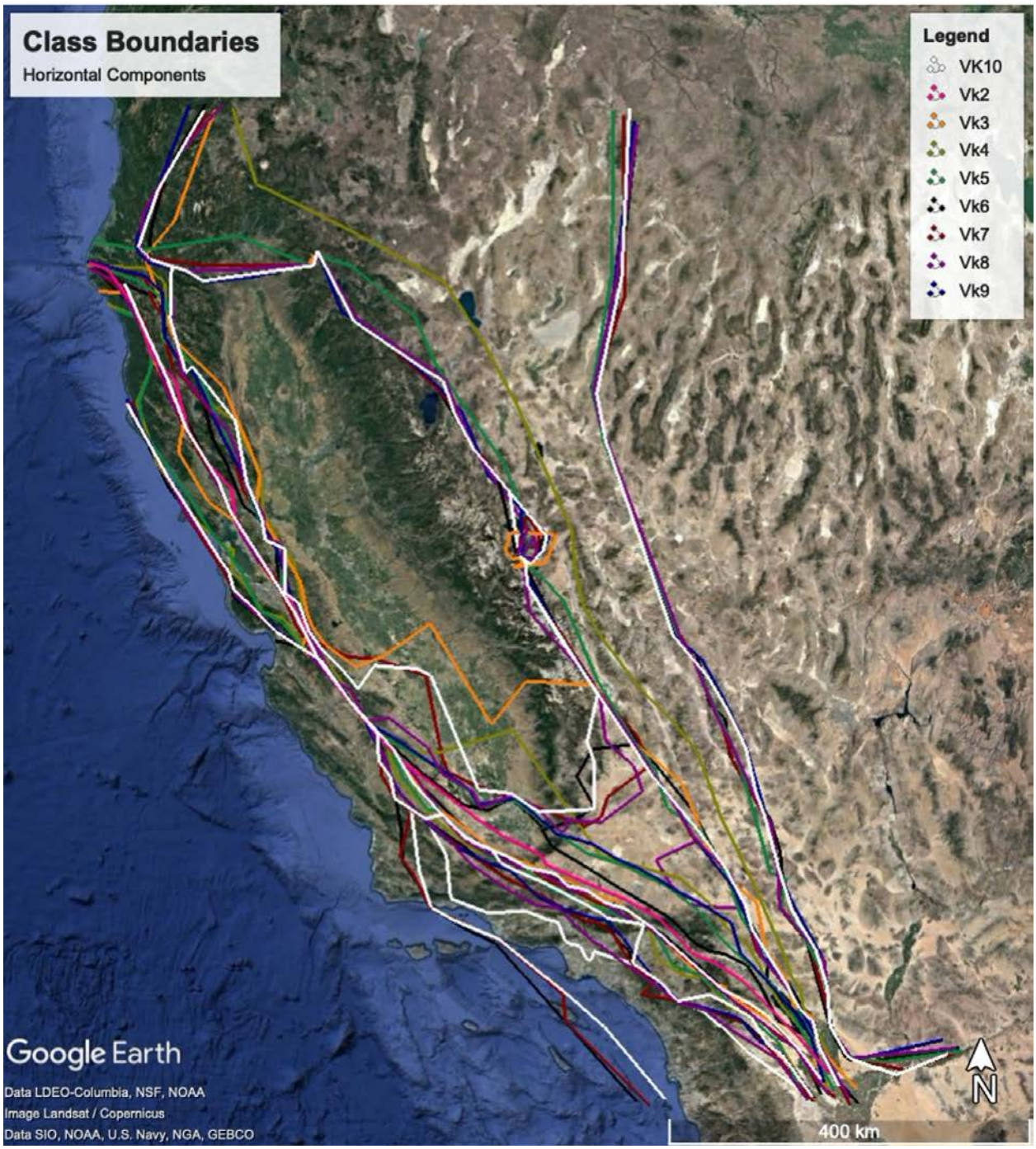

Figure 2. Hand drawn boundaries between clusters for long term velocities for number of clusters k = 2 to 10.

Clusters for the 2010 M7.3 El Mayor - Cucapah earthquake (EMC) show a distinct class near the earthquake rupture (Figure 3). Additional classes show different characteristics between the Salton Trough and Peninsular ranges with coseismic offsets detected into Mojave block and north. The 2019 Ridgecrest earthquake sequence, better centered in the middle of the GNSS network, shows far reaching cosesismic offsets to the GNSS stations (Figure 4). Stations within the Brawley Seismic Zone in the Salton Trough show breaks in the time series at the time of the Ridgecrest earthquakes. Classes for increasing k indicate local cosesismic offsets, classes that define the central California coast ranges and Transverse ranges, and that partition the eastern and western Mojave Desert along a zone mapped by Schermer et al. (1996). Showing the boundaries of each earthquake separately and combined (Figure 5) shows that stations are classified in groups that roughly follow the ECSZ and the boundaries take up more of the region to the east of both ruptures. Perhaps future earthquakes will affect stations along the San Andreas fault system to a greater extent.

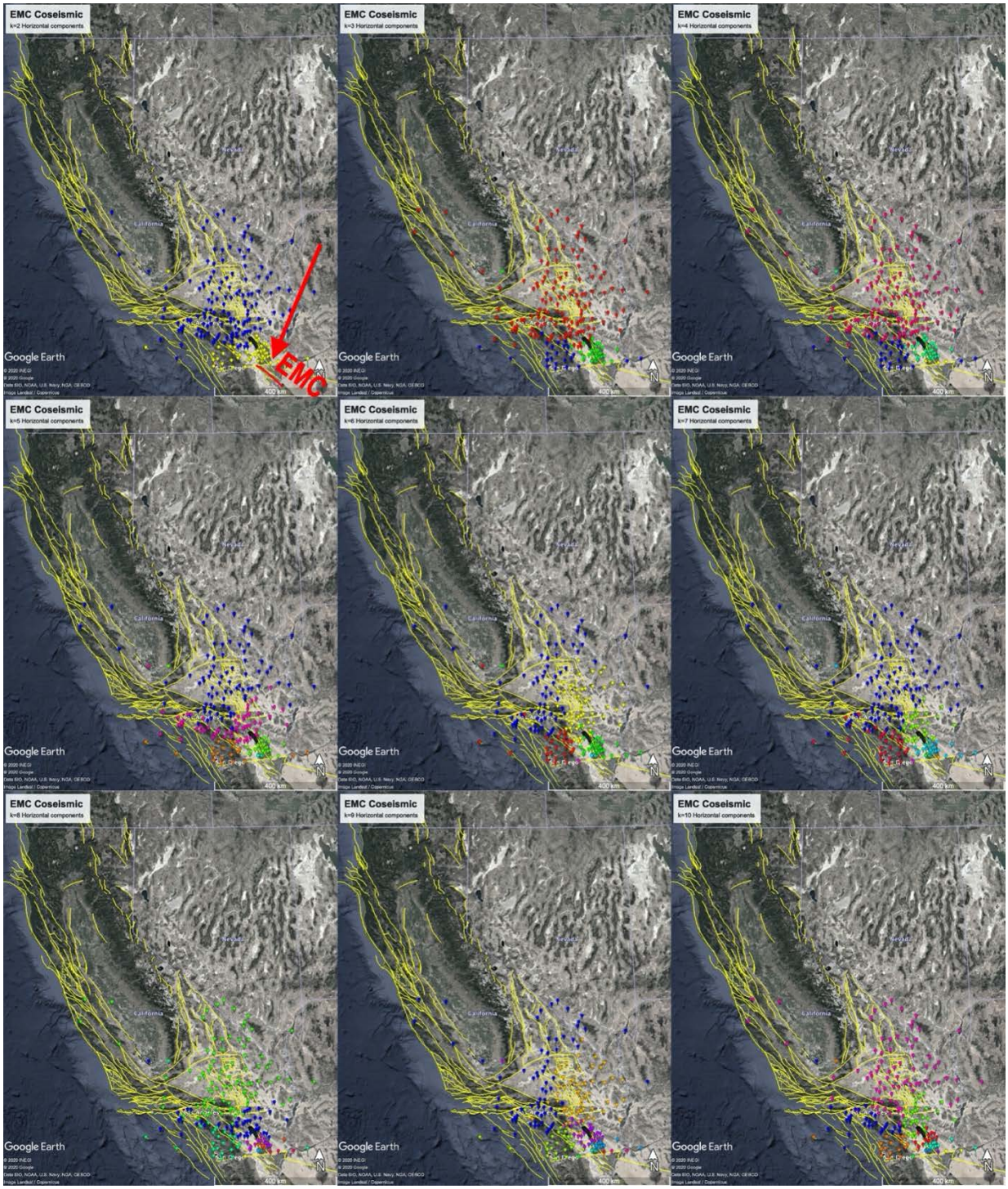

Figure 3. Clusters for cosesismic motions from the 2010 M7.3 El Mayor - Cucapah earthquake with the number of clusters $k = 2$ to 10 increasing row wise. Yellow lines show UCERF-3 faults. Red line shows the EMC rupture location. Red arrow points to the EMC rupture.

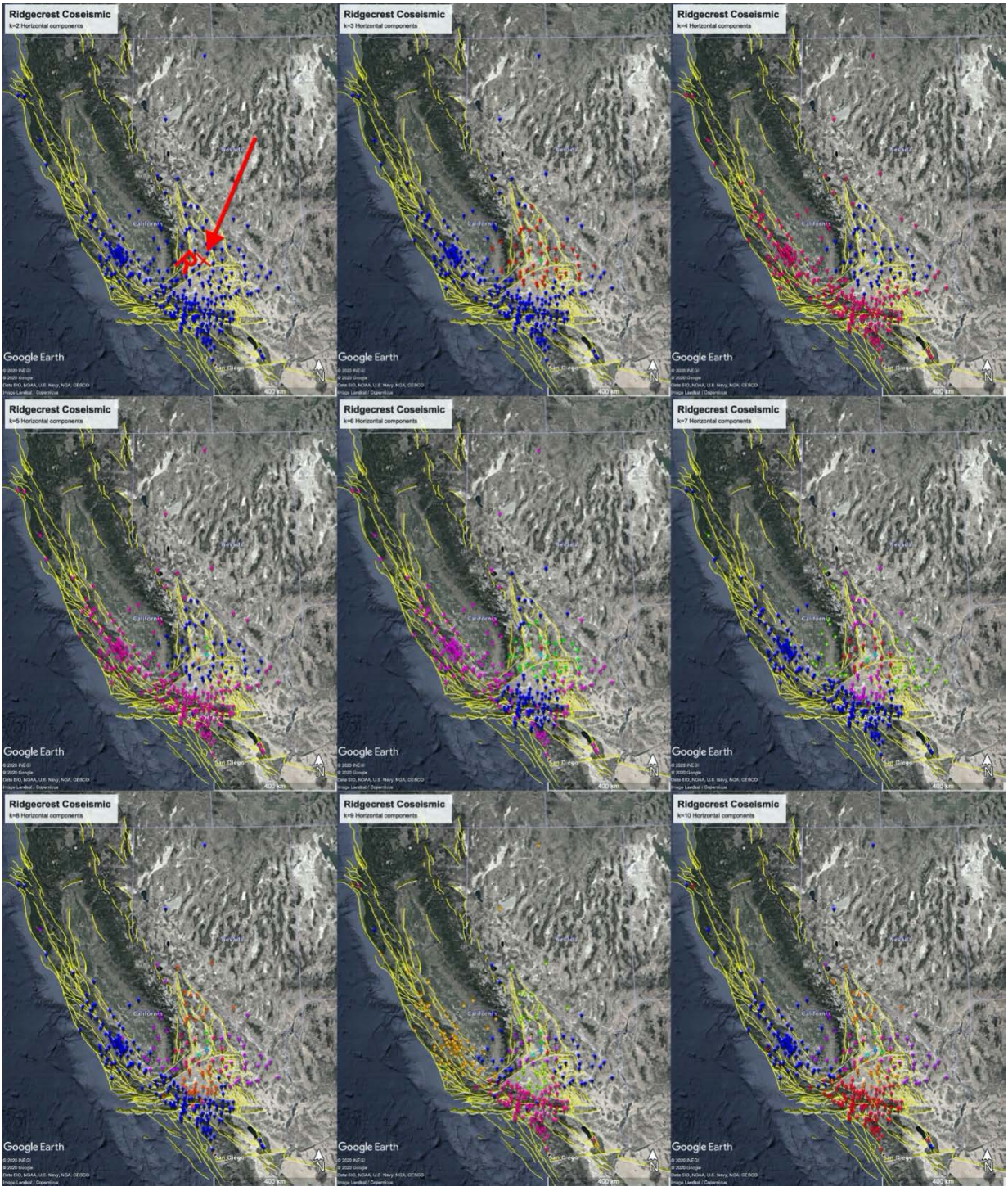

Figure 4. Clusters for cosesismic motions from the 2019 M6.4 and M7.1 Ridgecrest earthquake sequence with the number of clusters $k = 2$ to 10 increasing row wise. Yellow lines show UCERF-3 faults. Red line shows the EMC rupture location. Red arrow points to the EMC rupture.

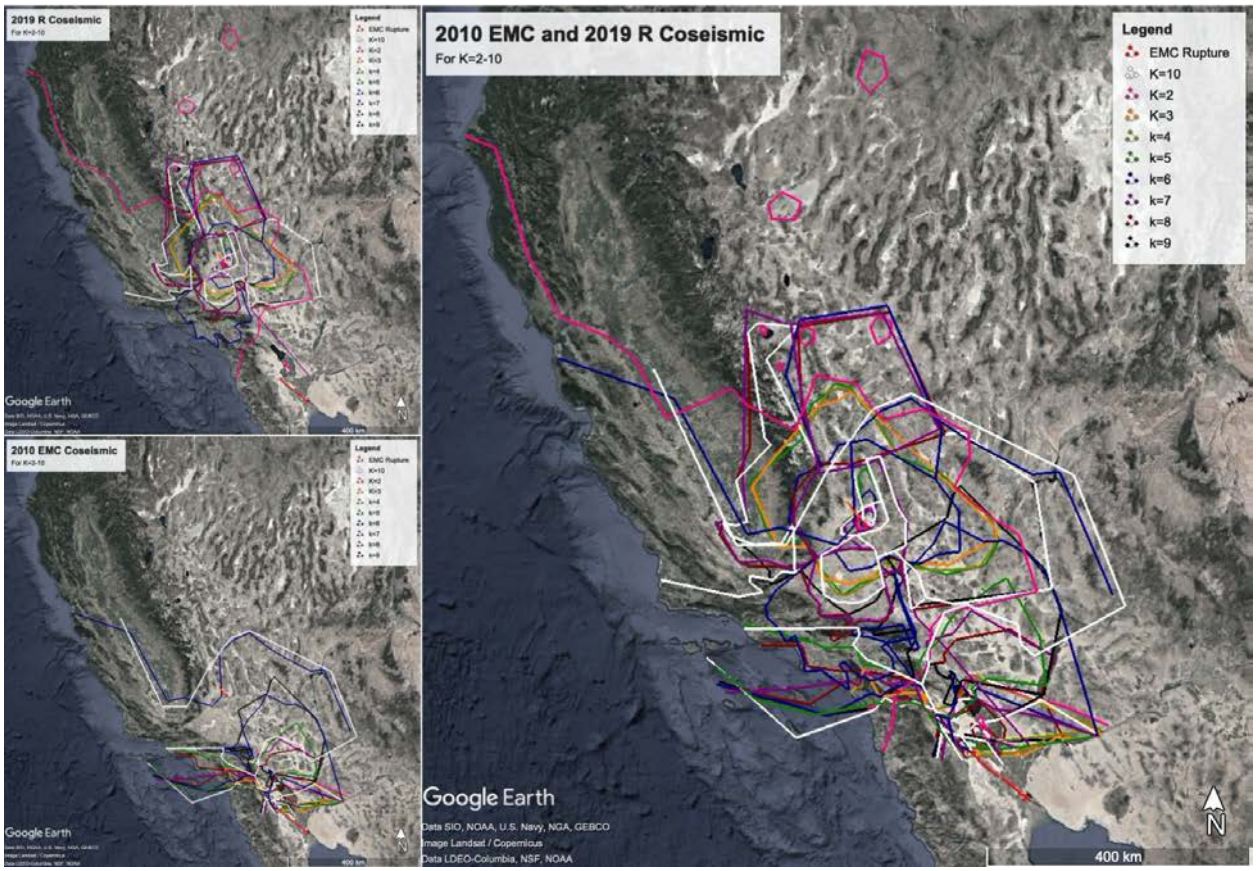

Figure 5. Coseismic boundaries for the EMC and R earthquakes. Top left shows classes k = 2 to 10 for R and the bottom left for EMC. Coseismic classes for both are shown in the right plot illustrating that large events of the last decade affected the ECSZ.

We next computed the classes for postseismic motions for each earthquake (Figures 6-7). GNSS station breaks are routinely identified in the time series (Heflin et al, 2000) and the GNSS tools sum the breaks over a given time period. The default is 2 years, however given the data of this writing we had to use 15 months for the Ridgecrest earthquake sequence. While the assumption is that breaks are associated with each event sequence it is likely that breaks can be identified due to other causes. As a result, the farther field stations that are identified are effectively noisy and more classes are not helpful to the interpretation. Using up to 5 classes is informative, however, and focuses attention on features local to the events (Figure 8). The Salton trough is clearly affected by the EMC

earthquake in the postseismic timeframe with sharp boundaries to the east and west that follow the general fault directions as well as the topographic fabric. The rupture is in the western third of the clusters, suggesting that there is more postseismic deformation on the northeast side of the EMC rupture. The R rupture shows a similar result with the fault located toward the west side of the central/local cluster groups. Interestingly, the local cluster groups are bounded by the Garlock fault to the south. While postseismic motions did occur south of the Garlock fault the clustering algorithm identifies those stations as having different feature characteristics.

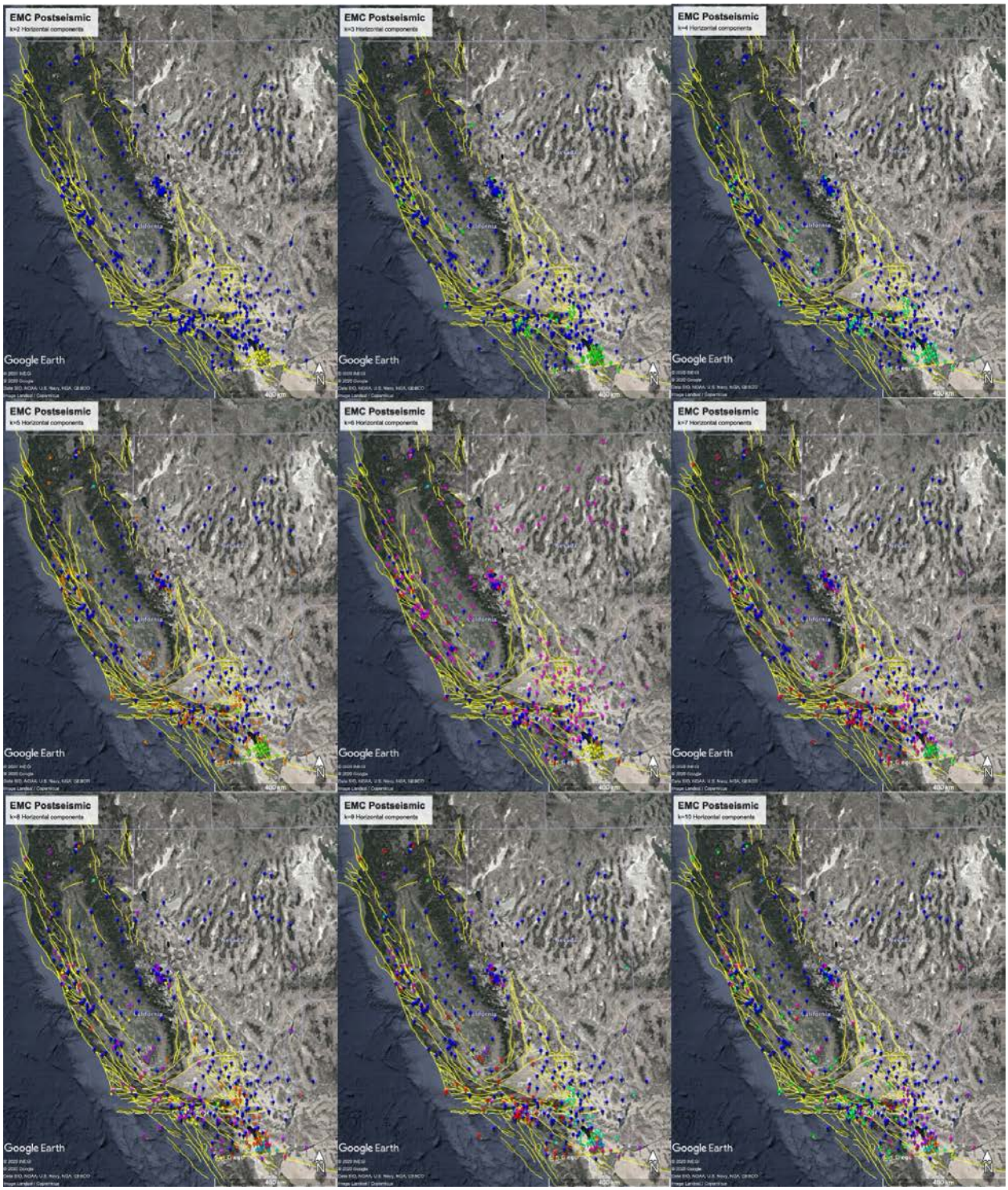

Figure 6. Classes for EMC cumulative postseismic motion for two years from 22 April 2010 with the number of clusters $k = 2$ to 10 increasing by rowe. Yellow lines are UCERF3 faults.

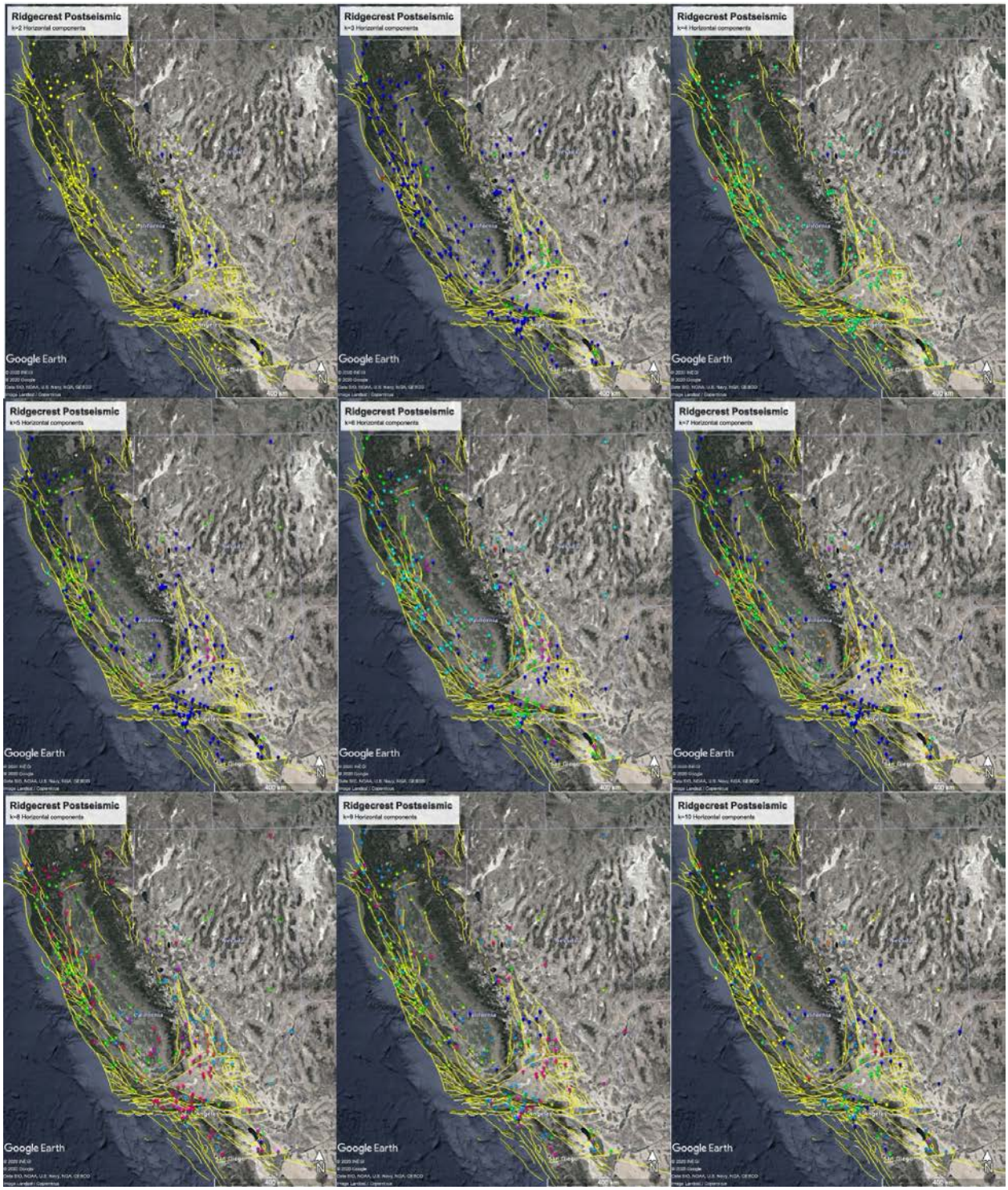

Figure 7. Classes for R cumulative postseismic motion from 23 July 2019 to 1 December 2020 with the number of clusters $k = 2$ to 10 increasing by row. Yellow lines are UCERF3 faults.

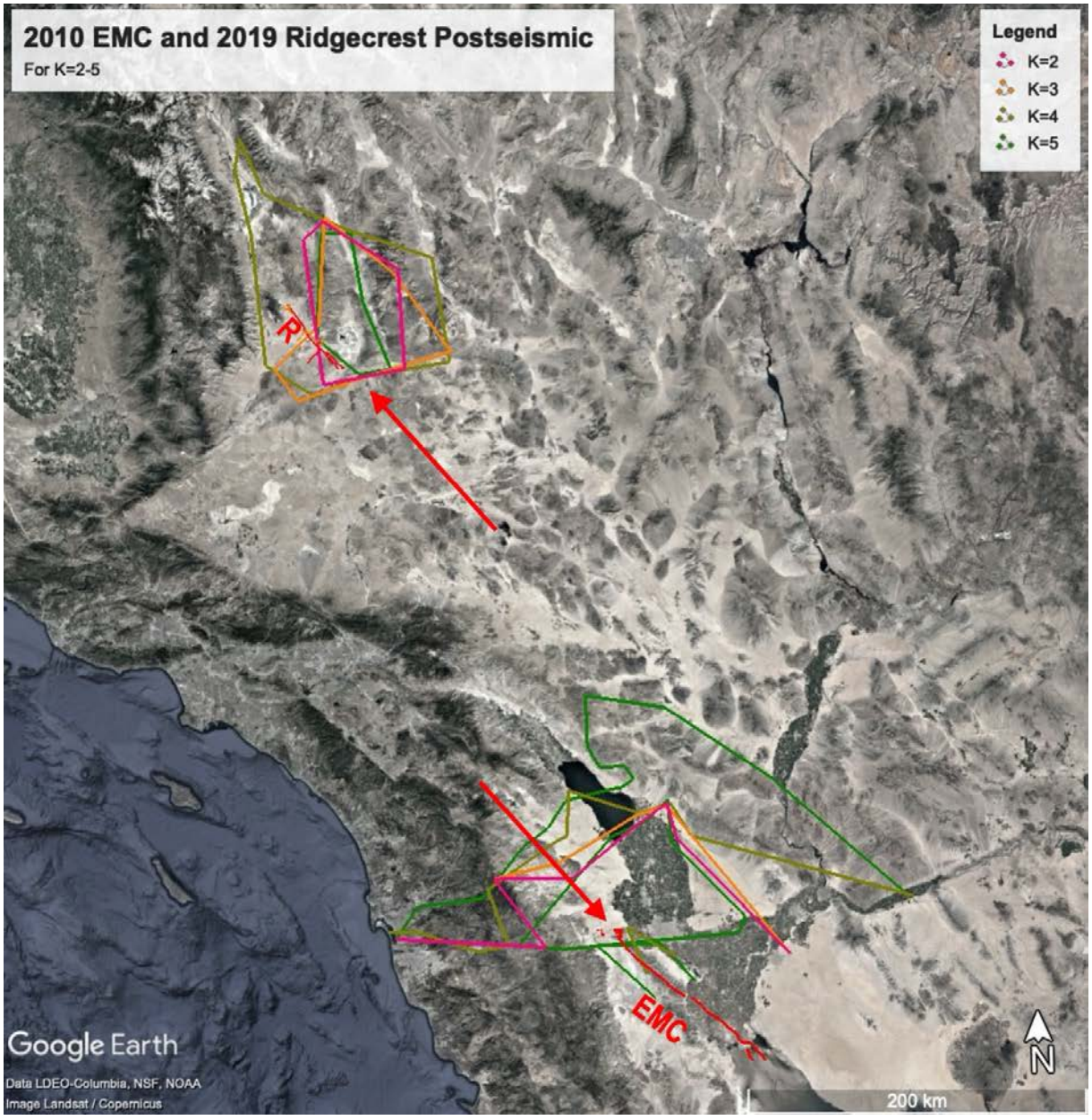

Figure 8. Boundaries for postseismic motions for EMC and R earthquakes that are found in close proximity to each respective rupture. Class boundaries distant to each earthquake are not shown as they inconsistent across the number of classes. The motions are subtle far field and the results are unlikely to be geophysically meaningful. Red lines show ruptures and red arrows point to the ruptures. Top of figure shows class boundaries for for R and the bottom shows boundaries for EMC for classes k =2 to 5 in both cases.

Finally, we divided the network time series into three time frames and computed the modeled displacements for each. The three time frames were 2000 to before the 2010 EMC earthquake, the time period between the 2010 EMC and 2019 R earthquakes, and the post R time frame (Figures 9-11). Qualitatively the results are not particularly different than for the long-term velocities. This suggests that the crust responds within a few years.

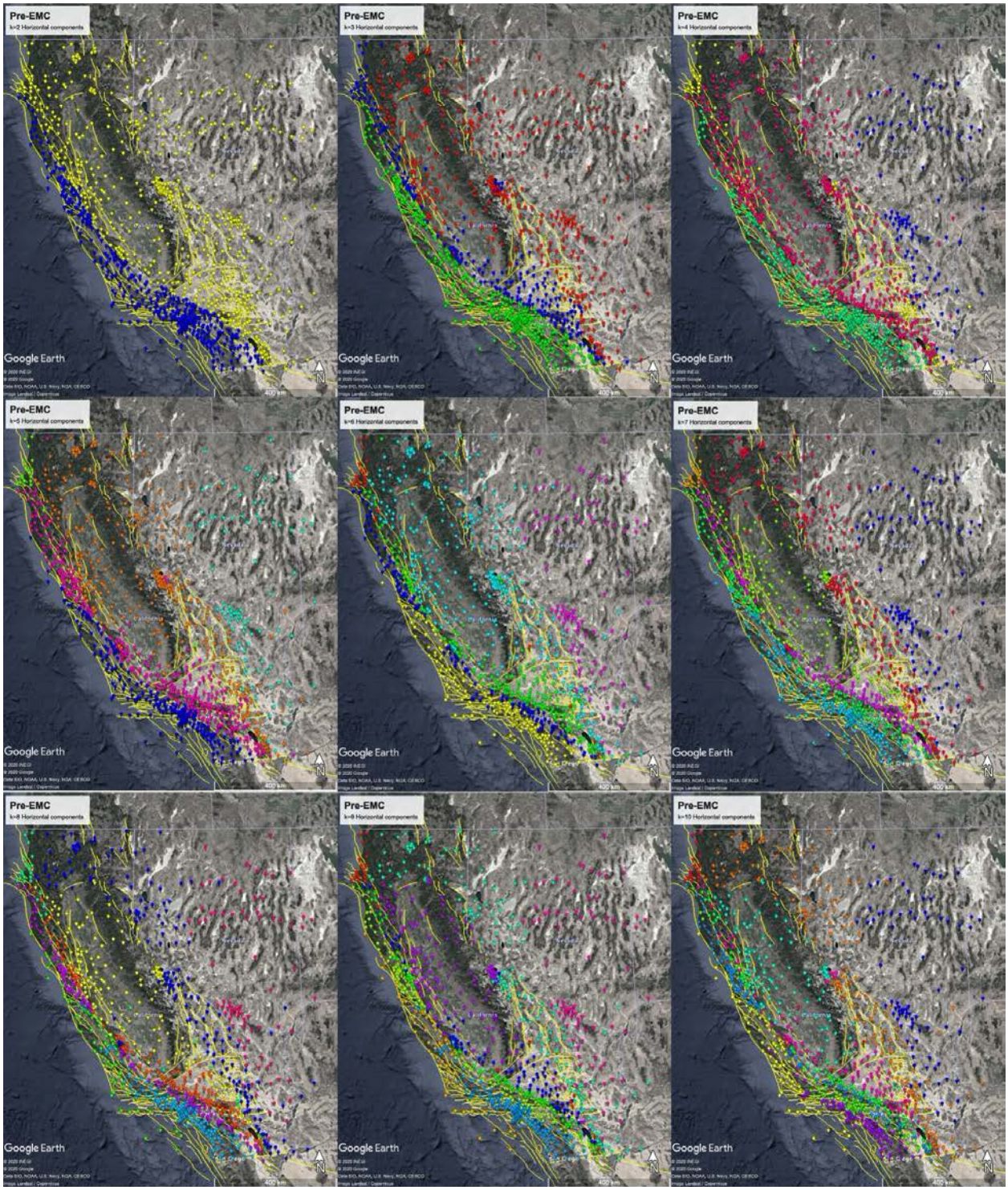

Figure 9. Clusters determined from modeled displacements in the decade before the EMC earthquake with the number of clusters k = 2 to 10 increasing row wise.

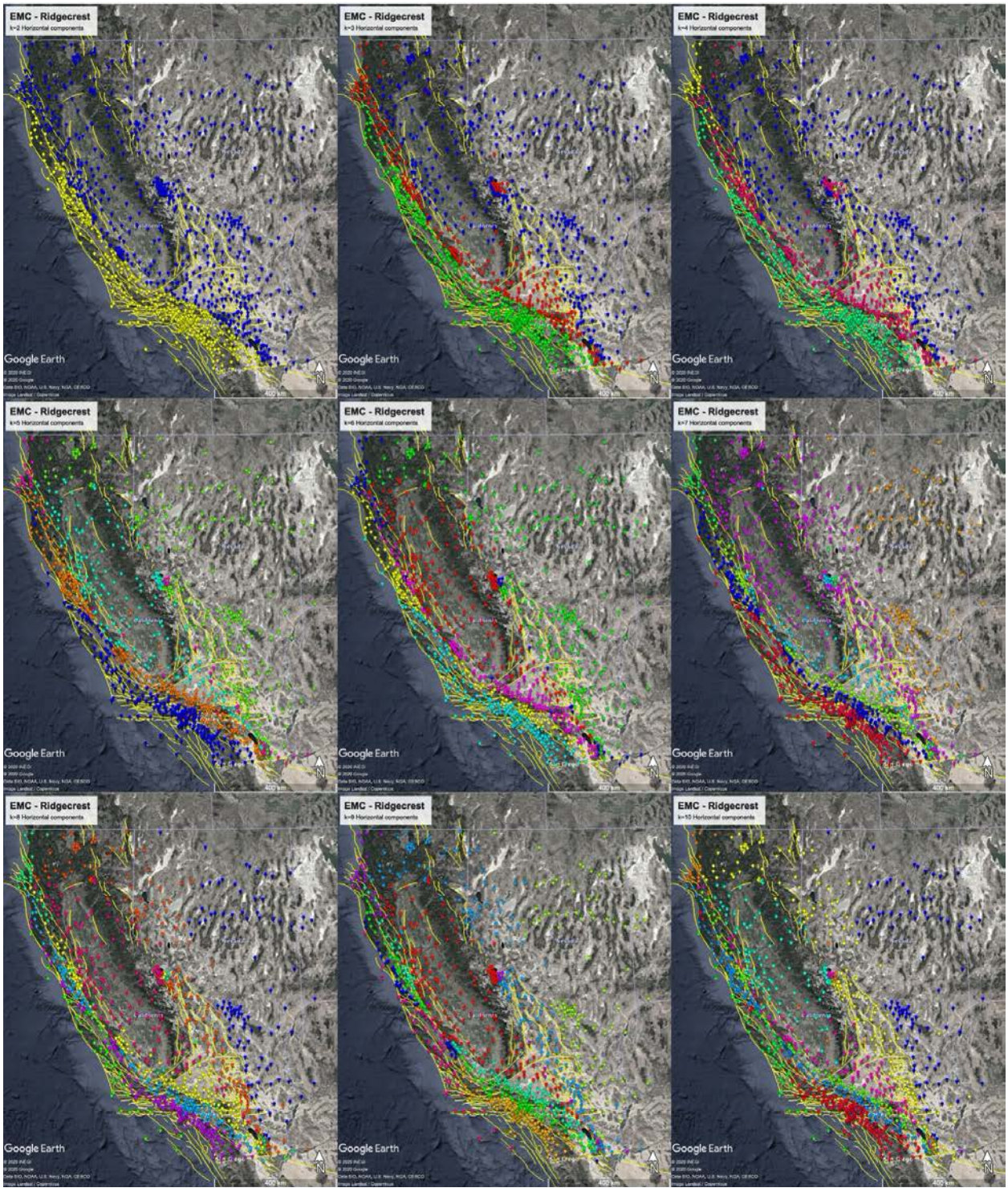

Figure 10. Clusters determined from modeled displacements in the decade between the EMC and R earthquakes with the number of clusters k = 2 to 10 increasing row wise.

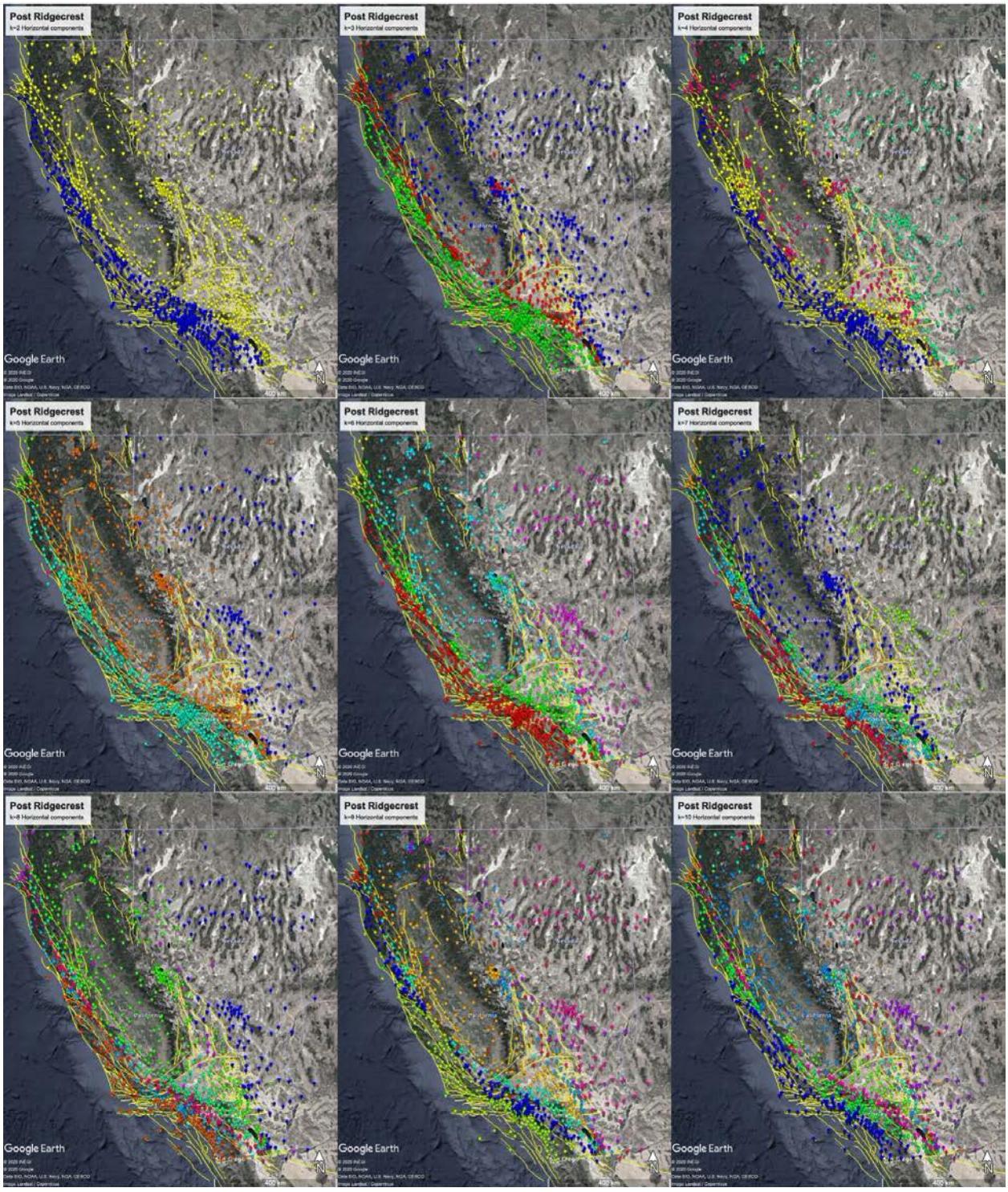

Figure 11. Clusters determined from modeled displacements for the 15 months after the R earthquake with the number of clusters k = 2 to 10 increasing row wise.

4. Conclusions

Our experiments indicate that not constraining the clustering algorithm by location or geometry results avoids domination by geometric features and allows for clusters that better correlate with faults and tectonic features, since the plate boundary naturally produces fault parallel extended clusters. Omitting the vertical results in clusters that better align with tectonic features and are not affected by subsidence associated with oil and groundwater withdrawal or recharge. Future work will explore the use of vertical results for study of non-tectonic crustal deformation signals.

The San Andreas fault system dominates crustal deformation of the Pacific - North American plate boundary in California. However, cluster boundaries for fewer clusters show a slightly straighter main plate boundary zone of deformation than the San Andreas fault. In particular, no cluster boundaries connect the fault north of the southern bend/San Gorgonio Pass across the Coachella Valley near Palm Springs to the Coachella segment of the San Andreas fault. Boundaries east of the Salton Sea connect with the ECSZ and boundaries associated with the San Andreas system are found from the center of the Salton Sea to the west. As the number of clusters increases there are an increasing number of cluster boundaries distributed to the west that can be associated with the San Jacinto fault, Elsinore fault, and faults offshore in the San Diego Trough. The finding that the classes cut across the San Andreas fault is consistent with observations of fragile geologic features near Cajon Pass, suggesting that joint ruptures of the San Andreas and San Jacinto faults propagate through the Pass (Grant-Ludwig et al., 2015; Lozos, 2016). A pattern of joint rupture implies that the San Andreas and San Jacinto faults are connected at seismogenic depths and the San Jacinto fault is the main branch of the San Andreas system in southern California south of Cajon Pass. Older branches of the San Andreas fault extending through the San Gorgonio Pass have complex geometry (Yule and Sieh, 2003) and are being actively offset by sinistral conjugate faults (Kendrick et al., 2015).

The creeping section of the San Andreas fault concentrates the boundaries, which fan out on either side. The weak Fransciscan rock facilitates creep (Irwin and Barnes, 1975). Presumably when the creeping section closes and there are fewer compliant rocks the zone of deformation will broaden in Central California. Further north as in southern California the boundaries become distributed as the number of clusters increases.

The ECSZ is a throughgoing boundary that begins in the eastern Salton Trough and trends north through the Mojave block. The ECSZ identified in this clustering methodology is further east than the Landers rupture and Blackwater fault. Instead, it follows the Goldstone Lake or Paradise fault bounding a set of east-west striking faults on the west side. Another persistent set of cluster boundaries bounds the same east-west striking faults on the east side as well as the eastern end of the Garlock fault. The presence of the boundaries and the fault geometry of the Mojave Desert area suggest that the east Garlock fault is transferred south along the right-lateral ECSZ.

For long-term velocities and coseismic motions for large earthquakes at least with Mw≥7 a range of clusters k from 2 to 10 is informative. For postseismic motions $k \leq 5$ is informative. It may be that more classes are not useful because fewer stations are typically affected by postseismic motions. In general, postseismic motions are fairly localized to within one fault dimension, but for the EMC and Ridgecrest earthquakes classes identified in postseismic motion are weighted more to the east of each rupture. The Garlock fault creates a clear barrier for Ridgecrest postseismic motion confining it to the north of the fault. Future experiments should test whether including geometry as a constraint improves the results.

Selecting decadal long subsets of the GNSS time series proved less useful other than to identify that classes computed from several years of data are similar to those for the long-term velocities. This suggests that most postseismic motion occurs within a few

years of earthquakes, which is similar to other observations (e.g. Ross et al, 2017, Donnellan et al, 2018, Shen et al, 1992, Pollitz et al, 2000).

Acknowledgments, Software and Data

This work was carried out at City College of New York and at the Jet Propulsion Laboratory, California Institute of Technology under contract with NASA. The work was funded by NASA's ACCESS, Advanced Information Systems Technology (AIST), Geodetic Imaging, and Earth Surface and Interior (ESI) programs. The data products can be accessed at http://geo-gateway.org. The clustering software used for these experiments can be accessed at https://github.com/GeoGateway/GNSS_clustering.

References

Behr, J., Bryant, B., Given, D., Gross, K., Hafner, K., Hardebeck, J., Hauksson, E., Heaton, T., Hough, S., Hudnut, K. and Hutton, K., 2000. Preliminary report on the 16 October 1999 M 7.1 Hector Mine, California, earthquake. *Seismological Research Letters*, *71*(1), pp.11-23.

Caskey, S.J., Wesnousky, S.G., Zhang, P. and Slemmons, D.B., 1996. Surface faulting of the 1954 Fairview Peak (MS 7.2) and Dixie Valley (MS 6.8) earthquakes, central Nevada. *Bulletin of the Seismological Society of America*, *86*(3), pp.761-787.

Donnellan, A., Parker, J., Heflin, M., Lyzenga, G., Moore, A., Ludwig, L.G., Rundle, J., Wang, J. and Pierce, M., 2018. Fracture Advancing Step Tectonics Observed in the Yuha Desert and Ocotillo, CA, Following the 2010 Mw7. 2 El Mayor-Cucapah Earthquake. *Earth and Space Science*, *5*(9), pp.456-472.

Donnellan, A., J. Parker, M. Heflin, M. Glasscoe, G. Lyzenga, M. Pierce, J. Wang, J. Rundle. L. Grant Ludwig, R. Granat, M. Mirkhanian, 2021, Improving Access to

Geodetic Imaging Crustal Deformation Data Using GeoGateway, Earth Science Informatics, DOI: 10.1007/s12145-020-00561-7.

Elkan, C., 2003. Using the triangle inequality to accelerate k-means. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)* (pp. 147-153).

Field, E.H., Arrowsmith, R.J., Biasi, G.P., Bird, P., Dawson, T.E., Felzer, K.R., Jackson, D.D., Johnson, K.M., Jordan, T.H., Madden, C. and Michael, A.J., 2014. Uniform California earthquake rupture forecast, version 3 (UCERF3)—The time-independent model. *Bulletin of the Seismological Society of America*, *104*(3), pp.1122-1180.

Glazner, A.F., Walker, J.D., Bartley, J.M. and Fletcher, J.M., 2002. Cenozoic evolution of the Mojave block of southern California. *Geologic evolution of the Mojave Desert and southwestern Basin and Range*, Geological Society of America Memoir *195*, pp.19-41.

Granat, R., Parker, J., Kedar, S., Dong, D., Tang, B. and Bock, Y., 2013. Statistical approaches to detecting transient signals in GPS: Results from the 2009–2011 transient detection exercise. *Seismological Research Letters*, *84*(3), pp.444-454.

Granat, R. and Donnellan, A., Unsupervised Learning for Geophysical Data Exploration. *2nd Int'l ACES Workshop*, Hakone, Japan, 2000.

Grant Ludwig, L., Brune, J. N., Anooshehpoor, A., Purvance, M. D., Brune, R. J. and Lozos, J. (2015). Reconciling precariously balanced rocks with large earthquakes on the San Andreas fault system, *Seismological Research Letters,* **86(5**), 1345-1353.

Hamerly, G. and Drake, J., 2015. Accelerating Lloyd's algorithm for k-means clustering. In *Partitional clustering algorithms* (pp. 41-78). Springer, Cham.

Hauksson, E., Jones, L.M., Hutton, K. and Eberhart-Phillips, D., 1993. The 1992 Landers earthquake sequence: Seismological observations. *Journal of Geophysical Research: Solid Earth*, *98*(B11), pp.19835-19858.

Heflin, M., Donnellan, A., Parker, J., Lyzenga, G., Moore, A., Ludwig, L.G., Rundle, J., Wang, J. and Pierce, M., 2020. Automated estimation and tools to extract positions, velocities, breaks, and seasonal terms from daily GNSS measurements: Illuminating nonlinear salton trough deformation. *Earth and Space Science*, *7*(7), p.e2019EA000644.

Irwin, W.P. and Barnes, I., 1975. Effect of geologic structure and metamorphic fluids on seismic behavior of the San Andreas fault system in central and northern California. *Geology*, *3*(12), pp.713-716.

Johnson, K.M., 2013. Slip rates and off-fault deformation in Southern California inferred from GPS data and models. *Journal of Geophysical Research: Solid Earth*, *118*(10), pp.5643-5664.

Jones, J.C., 1915. The pleasant valley, Nevada, earthquake of October 2, 1915. *Bulletin of the Seismological Society of America*, *5*(4), pp.190-205.

Kendrick, K.J, Matti, J.C. and Mahon, S.A., 2015. Late Quaternary slip history of the Mill Creek strand of the San Andreas fault in San Gorgonio Pass, southern California: The role of a subsidiary left-lateral fault in strand switching. *GSA Bulletin,* 127(5-6): 825-849. https://doi.org/10.1130/B31101.1

McLachlan, G. and Peel, D., 2004. *Finite mixture models*. John Wiley & Sons.

Liu, C., Lay, T., Brodsky, E.E., Dascher-Cousineau, K. and Xiong, X., 2019. Coseismic rupture process of the large 2019 Ridgecrest earthquakes from joint inversion of geodetic and seismological observations. *Geophysical Research Letters*, *46*(21), pp.11820-11829.

Lloyd, S., 1982. Least squares quantization in PCM. *IEEE transactions on information theory*, *28*(2), pp.129-137.

Lozos, J.C., 2016. A case for historic joint rupture of the San Andreas and San Jacinto faults. *Science advances*, *2*(3), p.e1500621.

Meade, B.J. and Hager, B.H., 2005. Block models of crustal motion in southern California constrained by GPS measurements. Journal of Geophysical Research: Solid Earth, 110(B3).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J., 2011. Scikit-learn: Machine learning in Python. *Journal of machine learning research*, *12*(Oct), pp.2825-2830.

Peltzer, G., Crampé, F., Hensley, S. and Rosen, P., 2001. Transient strain accumulation and fault interaction in the Eastern California shear zone. *Geology*, *29*(11), pp.975-978.

Pollitz, F.F., Peltzer, G. and Bürgmann, R., 2000. Mobility of continental mantle: Evidence from postseismic geodetic observations following the 1992 Landers earthquake. *Journal of Geophysical Research: Solid Earth*, *105*(B4), pp.8035-8054.

Ross, Z.E., Rollins, C., Cochran, E.S., Hauksson, E., Avouac, J.P. and Ben-Zion, Y., 2017. Aftershocks driven by afterslip and fluid pressure sweeping through a fault-fracture mesh. *Geophysical Research Letters*, *44*(16), pp.8260-8267.

Rymer, M.J., Treiman, J.A., Kendrick, K.J., Lienkaemper, J.J., Weldon, R.J., Bilham, R., Wei, M., Fielding, E.J., Hernandez, J.L., Olson, B.P. and Irvine, P.J., 2010. Triggered surface slips in southern California associated with the 2010 El Mayor-Cucapah, Baja California, Mexico, earthquake. *US Geological Survey Open-File Report*, *1333*.

Savage, J.C. and Simpson, R.W., 2013. Clustering of GPS velocities in the Mojave Block, southeastern California. *Journal of Geophysical Research: Solid Earth*, *118*(4), pp.1747-1759.

Savage, J.C. and Simpson, R.W., 2013. Clustering of velocities in a GPS network spanning the Sierra Nevada Block, the northern Walker Lane Belt, and the central Nevada Seismic Belt, California-Nevada. *Journal of Geophysical Research: Solid Earth*, *118*(9), pp.4937-4947.

Schermer, E.R., Luyendyk, B.P. and Cisowski, S., 1996. Late Cenozoic structure and tectonics of the northern Mojave Desert. *Tectonics,* 15(5), pp.905-932.

Shen, Z.K., Jackson, D.D., Feng, Y., Cline, M., Kim, M., Fang, P. and Bock, Y., 1994. Postseismic deformation following the Landers earthquake, California, 28 June 1992. *Bulletin of the Seismological Society of America*, *84*(3), pp.780-791.

Silver, E.A., 1971. Tectonics of the Mendocino triple junction. *Geological Society of America Bulletin*, *82*(11), pp.2965-2978.

Simpson, R.W., Thatcher, W. and Savage, J.C., 2012. Using cluster analysis to organize and explore regional GPS velocities. *Geophysical Research Letters*, *39*(18).

Thatcher, W., Savage, J.C. and Simpson, R.W., 2016. The eastern California shear zone as the northward extension of the southern San Andreas fault. *Journal of Geophysical Research: Solid Earth*, *121*(4), pp.2904-2914.

Yule, D., and Sieh, K., 2003. Complexities of the San Andreas fault near San Gorgonio Pass: Implications for large earthquakes, *Journal of Geophysical Research,* 108(B11), https://doi.org/10.1029/2001JB000451