SCEC Report

Implementing rapid, probabilistic association of earthquakes with source faults in the CFM for southern California

John H. Shaw (PI) & Andreas Plesch
Department of Earth and Planetary Sciences
Harvard University
20 Oxford St.
Cambridge, Massachusetts 02138
shaw@eps.harvard.edu

Egill Hauksson (PI), Men-Andrin Meier & Ellen Yu
Seismological Laboratory
California Institute of Technology
Pasadena, CA 91125
hauksson@gps.caltech.edu

Proposal Categories: Collaborative Proposal: Data Gathering and Products

Science Objectives: P3a, P2a, P3e

Duration: 1 February 2018 to 31 January 2018

Summary
During 2018, we have made progress on implementation of the earthquake association to a CFM fault (Eq2CFM) method through the SCEDC, and developed a series of enhancements to the data products that are provided (common fault name associations, fault maps, links to fault attributes [e.g., though USGS Qfault database]). The SCEDC will host this catalog locally and plans to submit this new catalog attribute to the ANSS ComCat national catalog. The SCEDC will also submit this product near-real time via email and post it to websites scedc.caltech.edu and NEIC to make it part of their earthquake.us.gov web pages and part of the ANSS ComCat. Two sample reports are included in Appendix.

This project has developed a new, statistically robust way to identify the fault (or sets of candidate faults) in the Community Fault Model (CFM) that generated an earthquake using information typically provided soon after these events occur (Evans, 2016; Evans et al., 2019). This effort effectively bridges the information provided by increasingly sophisticated near real-time seismograph networks with comprehensive 3D Community Fault Models (CFM's) developed by SCEC (Plesch et al., 2007). Our method of earthquake-to-fault association was developed using comprehensive earthquake hypocenter and focal mechanism datasets in California through 2016 (after Hauksson et al., 2012; Yang et al., 2012) and the southern California Community Fault Model (CFM) (Plesch et al., 2007), to assess what properties of earthquakes serve as the best predictors of the fault on which they occurred. We used a series of training datasets for earthquakes that were known to have occurred on faults within the model, and established that proximity (distance), focal mechanism (nodal plane orientation), and earthquake history (spatial and temporal clustering) can be combined in a robust way to assign probability that a given earthquake was associated with one or more source faults in the model (or on a fault not included in the model). Notably, these training datasets were comprised of earthquakes that occurred in the decade since the release of CFM 2.0, to ensure that they did not influenced the modeled fault geometries.

The method is implemented as an R script that calculates distances between earthquakes and CFM faults, and compares nodal plane orientations (see Figure 1). We have tested this approach on a series of previous CFM versions, and subsequently applied it to a current model version (5.2). For each earthquake in the catalog, the code outputs the five highest probabilities of association with a CFM fault, as well as the probability that the event is not associated with any source within the model. The majority of earthquakes (> 60%, above M 3) have a high probability of association with one or two faults in the model.

Objective earthquake-to-fault associations are of value as they provide an important measure of the activity of faults within the southern California plate boundary. The associations will facilitate detailed studies of whether and to what extent on-fault and off-fault earthquakes differ in their behavior. Moreover, as the method is implemented in near real-time, it will prove helpful in identifying clusters of small earthquakes that may be foreshocks of a larger, imminent event on a major fault. They will also assist in communicating objective information about the faults that source earthquakes to the scientific community and general public. Ultimately, more accurate identification of source faults for large earthquakes may help responders know which planned scenario is most like the situation they are facing. Such guidance should lead to more effective responses that can help save lives.

Method
Defining the fault that generated an earthquake is often not straightforward. Most moderate to small earthquakes, as well as larger blind events, do not involve direct surface rupture. Epicentral locations may be at significant distances from surface traces of dipping source faults. Moreover,

many earthquake sequences occur near fault junctures, or along fault zones that involve several closely spaced segments, providing several options for the earthquake source. In order to develop an objective approach to defining earthquake source faults, we evaluated a training dataset consisting of more than 600 earthquakes in southern California that had known associations with CFM faults (based on surface rupture, source inversions, or other seismologic, geodetic, or geologic studies) (Evans, 2016). In addition, we considered more than 20 events that were not associated with a fault in the model. We initially compared the properties of a subset of these events with known associations with a version of the CFM (2.0) completed in 2005, prior to the occurrence of these earthquakes. This addressed concerns related to the fact that CFM representations are developed, in part, using earthquake information.

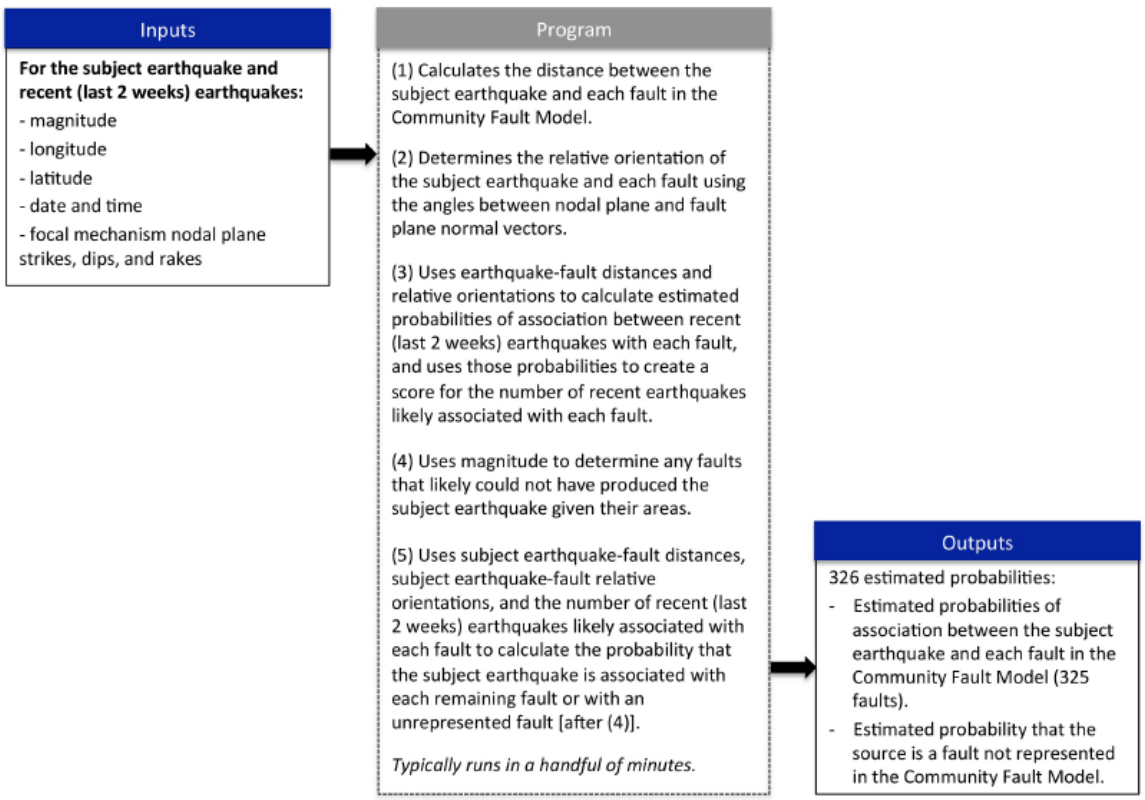

Figure 1: Earthquake-to-fault association workflow.

We design and evaluate four different models that use catalog data to predict the causative source fault(s) and optimize the model parameters in a maximum likelihood sense. First, we employed a method to calculate the distance between earthquake hypocenters and the closest part of faults within the CFM, which are represented as triangulated surfaces (tsurfs). For subsequent assessments, we only considered the faults that are within 20 km of the hypocenters as candidate source faults. Based on typical uncertainties in earthquake locations, we feel that this is a conservative assumption which does not exclude any viable source faults from consideration. We found that the distance to the true source fault is, on average, significantly lower than that with unassociated candidate faults, confirmed our expectation that distance is a useful predictor. The first, simplest model posits that the fault association only depends on the shortest distance between the fault and the hypocenter, with the probability decaying exponentially with distance. This Model 1 assigns the highest probability of association to 84% (as opposed to a 36% random probability) of events in our training dataset to the proper source fault. Figure 2 shows an example of applying this model to the Laguna Salada fault. Despite this success, many earthquakes were not assigned to the proper source fault in regions where two or more faults intersected forming a complex junction. Thus, we also explored comparisons of earthquake focal mechanisms with respect to the geometries of nearby faults represented in the model. For this assessment, we compared the minimum angle θ between the normal to the faults and the focal mechanism nodal planes, and used the plane of the two nodal planes that was closest to the orientation of the fault. We see that θ is, on average, smaller for true associations than between events and unassociated candidate faults. Thus we extended our statistical model to include both distance and orientation (Model 2). The revised model assigns the highest probability of association to 85% of events in our training dataset to the proper source fault, a 2% improvement over the distance-only assessment.

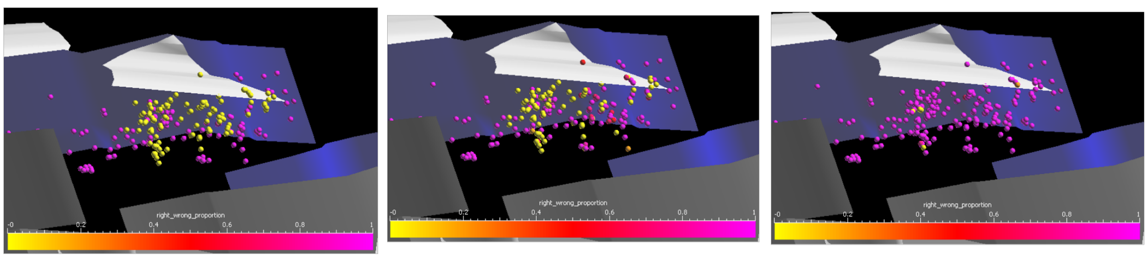

Figure 2: Effect of adding predictors to association model trained on CFM 2 faults, shown for the Laguna Salada fault; yellow events are miss-associated; (left) using only distance (Model 1); (middle) using distance and nodal planes (Model 2); (right) using distance, nodal planes and clustering (Model 3).

Finally, we explored whether earthquake clustering (in time and location) patterns provided value in determining earthquake to fault associations. Specifically, we assessed whether the probability of associating an earthquake with a source fault increased if previous earthquakes over a two week time period had a high probability of association with that same fault. To capture the clustering of earthquakes, we use Model 2 probabilities of association between earthquakes in the previous 2 weeks and that fault. This analysis showed that in our training datasets the probability of association with the correct source fault was significantly improved. Implemented as Model 3, this approach correctly associated 98% of the earthquakes with their true source fault, a significant improvement over Model 2.

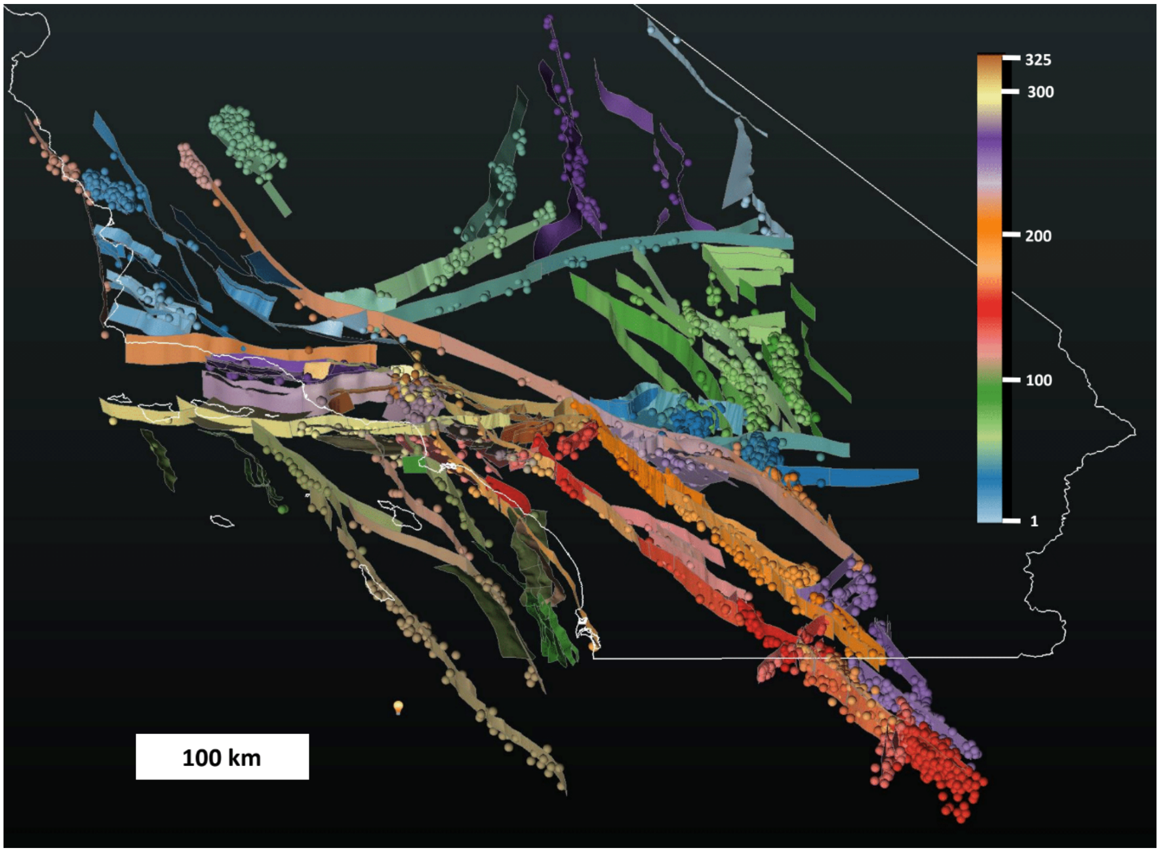

Figure 3: Primary association of the M>=3 catalog using the historical statistical model. Over 60% of events are associated. Colors refer to an internal fault ID for both events and fault representations.

Refinements and Implementation

To improve the clarity and richness of the model output, this past year we made a series of improvements and modifications to the script, its output, and associated databases. These included:

1: Providing common fault names via an internal fault ID and a compatible, where possible, USGS Quaternary Fault and Fold database ID. Previous versions of the script output names of CFM faults or fault segments, which are often complex and abbreviated descriptions of specific tsurf representations within the model. The new approach uses a unique fault ID to associate the CFM faults with common names derived from the full CFM fault name. The common fault name typically includes a section name for larger faults. As all CFM faults are not represented in the USGS database, the mapping to a Qfault ID is necessarily incomplete (251 out of the 325 faults could be mapped to a Qfault ID). All CFM faults were associated with common fault names defined in a look-up table, which is made available with the R evaluation script and used to compose automated reports.

2: Developing a set of fault traces that could be used to identify source faults on web-based map displays. This involved expanding and updating our fault trace database to a standard shapefile format. The shapefile now includes fault traces statewide and is based on CFM 5.2 in southern California.

3: Assessing the completeness of the CFM for different magnitude thresholds. As described above, over 60% of the magnitude 3 and higher earthquakes in the catalog have a primary association with a CFM fault. 17 % of those do not have a primary association, have a viable secondary fault association with a probability of more than 10%. 77 out of 120 M>=5 catalog events (64%) and 10 out of 14 M>=6 catalog events (72%) had a primary association with a CFM fault, Closer scrutiny of these four M>=6 events reveals that they are either located outside or at the margin of the fault model, or well below the maximum depth of the fault model. These locations are not expected to be associated with model faults and therefore the fault model appears largely complete at the M>=6 level. In the M>=5 magnitude range, there are unassociated events within the CFM model area indicating some degree of incompleteness of the fault model. In particular, a number of these events are located in the Coso geothermal area, in the Sierra Nevada. 6 out of the 43 M>=5 non-associated events have a significant secondary fault association with a probability of more than 10%.

4: The implementation at Caltech included writing a wrapper script to connect the Eq2CFM script into the AQMS seismic data processing system at Caltech. This enables the Eq2CFM script to run automatically when a M≥3.0 event with a focal mechanism is processed or post-processed. The wrapper scripts also sends out an email with an abbreviated message stating the probability of association with the nearest fault (see Appendix). The message format that was developed jointly by Caltech and Harvard will continue to evolve as we get more user feedback.

Conclusions

The project participants are now receiving emails with the Eq2CFM report of all M≥3.0 events in southern California. As we move forward, we will update the format of these reports to make them more easily understood by non-experts. We will also collect these reports and build a catalog that will be available through www.scedc.caltech.edu.

References & Publications

Evans, W., (2016), Determining Earthquake Source Faults in Southern California: Statistical Models for Historical and Future Earthquakes, Senior thesis in Statistics, Harvard University, Cambridge, MA.

Evans, W., A. Plesch, J.H. Shaw, N. Pillai, M.A. Meier, and E. Hauksson, (2019), A statistical method for associating earthquakes with their source faults in southern California, Bulletin of the Seismological Society of America, (in prep).

Hauksson, E., W. Yang, and P. M. Shearer, (2012), Waveform relocated earthquake catalog for southern california (1981 to June 2011), *Bull. Seismol. Soc. Am.*, *102*(5), 2239–2244.

Plesch, A., J. H. Shaw, C. Benson, W. A. Bryant, S. Carena, M. Cooke, J. Dolan, G. Fuis, E. Gath, L. Grant, E. Hauksson, T. Jordan, M. Kamerling, M. Legg, S. Lindvall, H. Magistrale, C. Nicholson, N. Niemi, M. Oskin, S. Perry, G. Planansky, T. Rockwell, P. Shearer, C. Sorlien, M. P. Süss, J. Suppe, J. Treiman, and R. Yeats, (2007), Community Fault Model (CFM) for Southern California, Bulletin of the Seismological Society of America, Vol. 97, No. 6, doi: 10.1785/012005021

Yang, W., E. Hauksson, and P.M. Shearer, P. M., (2012), Computing a large refined catalog of focal mechanisms for southern California (1981–2010): Temporal stability of the style of faulting. BSSA, 102(3), 1179-1194.

Yu, E., P. Acharya, A., Bhaskaran, S.-L. Chen, J. R. Andrews, V. Thomas, E. Hauksson, and R. W. Clayton, (2018), Cloud Computing and Big Data – Using the Southern California Earthquake Data Center (SCEDC) and the Southern California Seismic Network (SCSN) Products and Services for Earthquake Research, (SCEC Poster Presentation 057)

Appendix: Two Sample Reports

SCEC CFM fault associations for event 38213159
1 message

Thu, Jan 31, 2019 at 7:38 AM

Event Information

Event:	38213159
Time:	2019/1/31 3:13:36.09
Location:	35.365, -117.871
Depth (km/miles):	2.7/1.7
Magnitude:	3.17

Earthquakes can occur both near or on major known faults, and in places where no clear fault zones are known. The location and focal mechanism of this earthquake suggest the below association with modeled faults in the Community Fault Model (CFM) provided by the Southern California Earthquake Center (SCEC). This information is subject to change as more up-to-date data become available.

CFM Fault Association Probability

Most Likely
Garlock fault (96.0%)

Alternates
Not associated with a CFM modeled fault (4.0%)

Glossary:

CFM Fault:	SCEC CFM and closest segment if available
Probability:	The probability in percent the earthquake is associated with this fault
SCSN:	Caltech/USGS Southern California Seismic Network

SCEC CFM fault associations for event 38211175

Wed, Jan 30, 2019 at 7:32 AM

Event Information

Event:	38211175
Time:	2019/1/30 2:40:12.28
Location:	34.408, -116.892
Depth (km/miles):	1.9/1.2
Magnitude:	3.56

Earthquakes can occur both near or on major known faults, and in places where no clear fault zones are known. The location and focal mechanism of this earthquake suggest the below association with modeled faults in the Community Fault Model (CFM) provided by the Southern California Earthquake Center (SCEC). This information is subject to change as more up-to-date data become available.

CFM Fault Association Probability

Most Likely
Helendale; South Lockhart fault segment (89.2%)

Alternates
Not associated with a CFM modeled fault (8.7%)
Other CFM faults (2.1%)

Glossary:

CFM Fault:	SCEC CFM and closest segment if available
Probability:	The probability in percent the earthquake is associated with this fault
SCSN:	Caltech/USGS Southern California Seismic Network