

JGR Solid Earth

RESEARCH ARTICLE

10.1029/2019JB017641

Key Points:

- We develop an automated procedure to derive a catalog of earthquake locations and sizes from raw waveform data
- We derive a new catalog for the San Jacinto fault zone region with 108,800 events between 2008 and 2016
- We interpret variable mechanical properties of fault-zone regions from seismicity patterns

Supporting Information:

- Supporting Information S1
- Movie S1
- Movie S2

Correspondence to:

M. C. A. White,
 malcolm.white@usc.edu

Citation:

White, M. C. A., Ben-Zion, Y., & Vernon, F. L. (2019). A detailed earthquake catalog for the San Jacinto fault-zone region in southern California. *Journal of Geophysical Research: Solid Earth*, 124, 6908–6930. <https://doi.org/10.1029/2019JB017641>

Received 5 MAR 2019

Accepted 31 MAY 2019

Accepted article online 6 JUN 2019

Published online 22 JUL 2019

A Detailed Earthquake Catalog for the San Jacinto Fault-Zone Region in Southern California

Malcolm C. A. White¹ , Yehuda Ben-Zion¹ , and Frank L. Vernon² 

¹Department of Earth Sciences, University of Southern California, Los Angeles, CA, USA, ²Scripps Institution of Oceanography, University of California San Diego, San Diego, CA, USA

Abstract We develop an automated processing procedure to derive a new catalog of earthquake locations, magnitudes, and potencies and analyze 9 years of data between 2008 and 2016 in the San Jacinto fault-zone region. Our procedure accounts for detailed 3-D velocity structure using a probabilistic global-search location inversion and obtains high-precision relative event locations using differential travel times measured by cross-correlating waveforms. The obtained catalog illuminates spatiotemporal seismicity patterns in the fault zone with observations for 108,800 earthquakes in the magnitude range -1.8 to 5.4 . Inside a focus region consisting of an 80-km by 50-km rectangle oriented parallel to the main fault trace, we estimate a 99% detection rate of earthquakes with magnitude 0.6 and greater and detect and locate about 60% more events than those present in the Southern California Seismic Network catalog. The results provide the most complete catalog available for the focused study region during the analyzed period and include both deeper events and very shallow patches of seismicity not present in the regional catalog. The seismicity exhibits a variety of complex patterns that contain important information on deformation processes in the region. The fraction of event pairs with waveforms having cross-correlation coefficients ≥ 0.95 is only about 3%, indicating diverse processes operating in the fault zone.

1. Introduction

Catalogs of earthquake locations, occurrence times, and sizes are foundational for numerous studies of earthquake physics, crustal dynamics, and structural seismology. They define the existence and geometry of seismogenic faults at depth, provide information on spatiotemporal seismicity patterns, and are the basis of many derivative studies including body-wave tomography and analyses of seismic hazard. More detailed earthquake catalogs directly increase the available information on brittle-failure processes, crustal structures, and related topics.

The San Jacinto fault zone (SJFZ) is the most seismically active fault zone in Southern California (Hauksson et al., 2012; Ross et al., 2017; Sanders & Kanamori, 1984) during the modern instrumental period and has ruptured in $15 \text{ M} > 7$ earthquakes over the past 4,000 years (Rockwell et al., 2015). The fault zone accommodates 10–21 mm/year of the plate motion in Southern California and has accumulated 25 km of slip (Fialko, 2006; Sharp, 1967). The SJFZ has significant scientific and societal relevance because its high seismic activity can provide detailed information on seismogenic processes in a complex active plate boundary region, while posing significant seismic hazard to large urban areas.

Since 1932, the Southern California Seismic Network (SCSN) has been the authoritative agency for monitoring seismicity in Southern California; the agency currently operates over 400 seismic stations and regularly updates a catalog of analyst-reviewed earthquake parameters (Hutton et al., 2010), referred to here as *SCSN_catalog_2010*. A series of studies over the past two decades (Hauksson, 2000; Hauksson & Shearer, 2005; Hauksson et al., 2012; Lin et al., 2007; Shearer et al., 2005) relocated seismicity from the SCSN catalog from 1981, using 3-D velocity information, a source-specific station-term relocation method, and a double-difference relocation based on waveform cross-correlations. These studies culminated in a waveform-relocated catalog (Hauksson et al., 2012), referred to here as *HYS_catalog_2011*, which has been extended to later years and provides the standard earthquake catalog for Southern California.

The SCSN catalog covers the entire Southern California region and omits some data from more focused temporary local networks that make negligible difference for routine processing on a regional scale. The

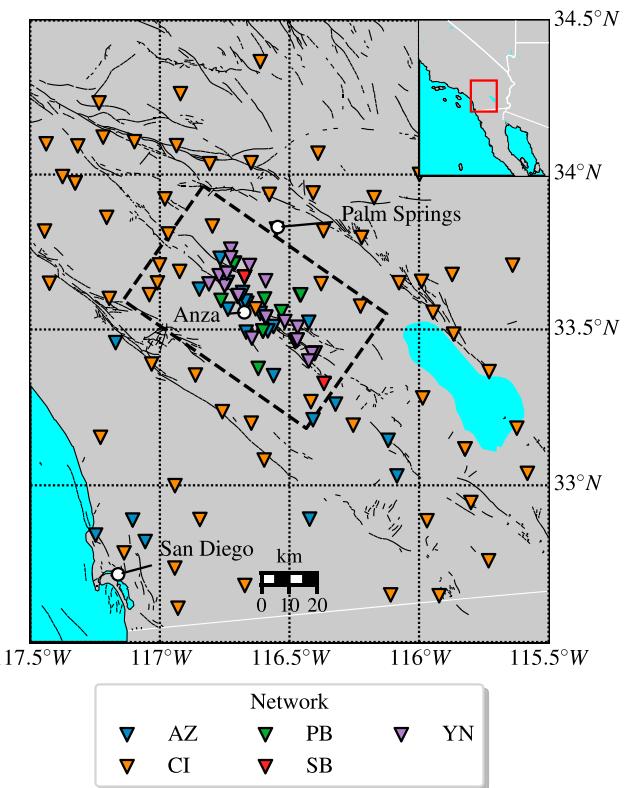


Figure 1. Map of seismic stations (inverted triangles color coded by operating network; see legend at bottom left) used in this study, surface traces of known Quaternary faults (thin black lines), and our focus region (black, dashed rectangular outline). The cities of San Diego, Anza, and Palm Springs are shown for reference. The region mapped in the large-scale map corresponds to the solid, red rectangular outline in the small-scale map (top right).

SJFZ Experiment Network (Vernon & Ben-Zion, 2010) includes 35 stations near the central portion of the SJFZ (Figure 1) that allow derivation of a more detailed earthquake catalog in that region. In the present study, we use data from the SJFZ experiment and additional four networks operating near the SJFZ to obtain a high-resolution earthquake catalog for the region marked by a black, dashed rectangle in Figure 1. In the following sections 2 and 3, we describe in detail the data and methods used to derive the catalog. The resulting catalog is used in section 4 to illuminate seismogenic structures in the SJFZ, and the observed features are discussed in section 5 in relation to tomographic results and some mechanical models for the region.

2. Data

2.1. Seismic Stations Around the SJFZ

Five seismic networks operating in the SJFZ region with complementary capabilities are used in this study (Figures 1 and 2): the Anza Network (AZ; Vernon, 1982), the SJFZ Experiment Network (YN; Vernon & Ben-Zion, 2010), the Plate Boundary Observatory Borehole Network (PB), the SCSN (CI; California Institute of Technology & United States Geological Survey Pasadena, 1926; Southern California Earthquake Center, 2013), and the UC Santa Barbara Engineering Seismology Network (SB). The AZ network is the longest running network of those that focus on the SJFZ, operating since 1982 and archiving continuous waveforms since 1998. The YN network is the densest network of those focusing on the SJFZ, with 65 stations deployed as either stand-alone stations or in fault-normal linear arrays. The PB network contributes high-quality data from eight high-frequency geophones installed in boreholes between 5 and 230 m deep. The CI network is the most extensive network, covering the whole of Southern California and providing a regional backbone for this study. The SB network operates two down-hole arrays in the SJFZ, from which we use data from one instrument per array.

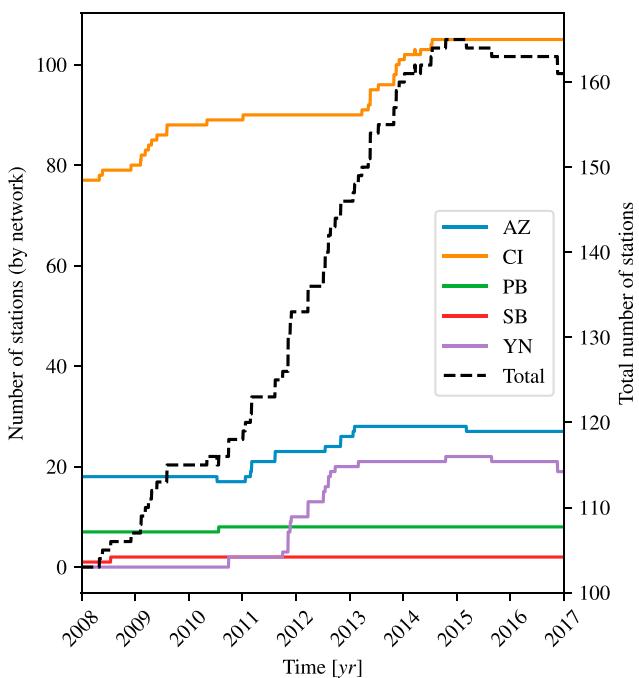


Figure 2. The number of stations used in this study as a function of time. The colored, solid lines represent the number of stations stratified by operating network (see legend; right middle) and are scaled relative to the left vertical axis. The black, dashed line represents the total number of stations and is scaled relative to right vertical axis. The increasing number of stations between 2008 and 2014 is chiefly the result of new station installation. The rapid expansion of the YN network starting in late 2011 is particularly important for this study.

2.2. Choosing and Downloading the Data

Because our aim is to catalog SJFZ microseismicity, and the AZ and YN networks target this region, we build an aggregate network called here the Augmented Anza Network (AZ+) using stations from these as the core (Figure 1). We include data from all five networks previously mentioned and choose stations to evenly cover the fault-zone area. The SCSN archived continuous waveforms at maximum sample rates of 40 samples per second (s.p.s.) prior to 2008, so we begin our analysis in 2008 when 100 s.p.s. continuous SCSN data archives became available.

Many stations have colocated instruments (e.g., broadband seismometers and strong-motion accelerometers), and digital data are sometimes archived at multiple sample rates. Because this study aims to detect small seismic events, we preferentially select a single data stream for each station based on two criteria: (i) instrument type and (ii) sample rate. Data from broadband seismometers are most preferred, followed in order of decreasing preference by intermediate-period seismometers, high-frequency geophones, and strong-motion accelerometers (see Table S1 in the supporting information for instruments used). The highest sample-rate data available (up to 250 s.p.s.) are preferred, and no data with sample rate lower than 40 s.p.s. are used in this study; we omit data with sample rates lower than 40 s.p.s. because the high-frequency signals from the small earthquakes we target are inadequately sampled below this threshold. Wherever possible, we use low-preference data streams to fill gaps in preferred data streams that are longer than 24 hr.

The final 12.5 TB data set comprises waveform data from 116 stations. Most of the data were downloaded from archives at UC San Diego; however, some gaps were filled with data downloaded with the Seismogram Transfer Program (used for data from the CI network) and the BREQ_FAST web resource provided by the Incorporated Research Institutions for Seismology's Data Management Center (used for all other networks). All data from the CI network used in this study can be downloaded using Seismogram Transfer Program, and all other data can be downloaded from the Incorporated Research Institutions for Seismology's Data Management Center using BREQ_FAST.

2.3. Velocity Model

The three-dimensional velocity structure in the region is modeled with a hybrid framework derived from Fang et al. (2016), referred to as FANG16, and the Southern California Earthquake Center's Community

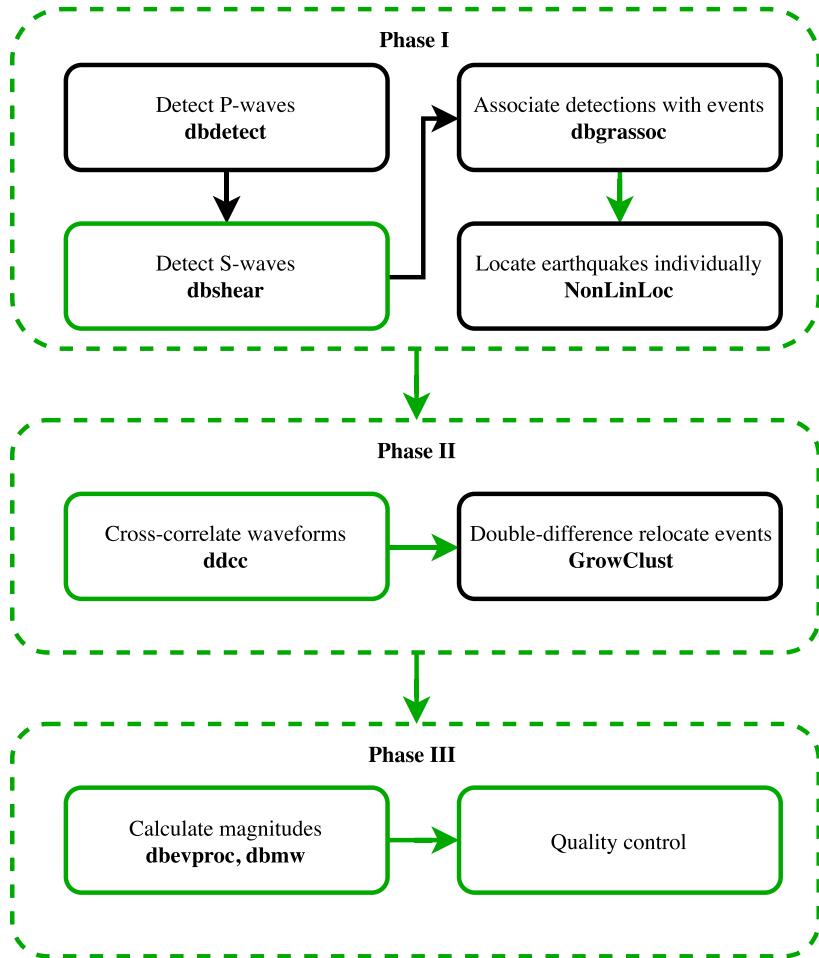


Figure 3. Schematic diagram outlining the automated processing procedure. Steps are divided into three phases: phase I—detecting and locating earthquakes; phase II—relocating earthquakes; and phase III—calculating event sizes and postprocessing. Each step gives a pithy description of its purpose, and the name of software components used are provided in **bold**. Components that we develop or contribute to are indicated in green. Program names beginning with **db** come from the Antelope software package (BRTT Inc.), and all others are open-source. All software used was obtained for free, either under an academic license or as open-source software.

Velocity Model Harvard version 15.1, referred to as CVM-H15.1. The FANG16 model was obtained directly from its authors and CVM-H15.1 was obtained using the Southern California Earthquake Center Unified Community Velocity Model software Small et al. (2017). We combine these two models because FANG16 resolves detailed structure within and around the SJFZ but lacks resolution along the edges of our study area; CVM-H15.1 complements FANG16 by resolving regional velocity structure well but lacks the detail of FANG16 inside the focus region of our study.

The hybrid velocity model (Figures S1 and S2) is defined on the same set of grid nodes as FANG16 and is derived by computing a weighted average of FANG16 and CVM-H15.1. Complementary weight functions are defined using a 2-D Gaussian distribution-oriented parallel to the fault zone (Figure S1) so that FANG16 is heavily weighted in the SJFZ and CVM-H15.1 is heavily weighted near the edges of the study area. The uppermost layer of FANG16 is located at -1.5 km (i.e., 1.5 km above sea level), and CVM-H15.1 is not defined for negative depths, so we set the -1.5 -km layer for CVM-H15.1 equal to 0.9 times the 0.0-km layer. An additional constant velocity layer is added to the final model at 30.0-km depth using velocities from the IASPEI91 model (Kennett & Engdahl, 1991): 6.5 and 3.75 km/s for P and S waves, respectively.

2.4. Analyst-Reviewed Control Data Set

To validate the performance of different processing parameters, we compare intermediate results with an unpublished, 40-day, analyst-reviewed catalog of phase arrival times and earthquake locations observed on

Table 1
Sample-Rate Dependent Processing Parameters Used for Detecting P Waves

	≥ 80 s.p.s.	<80 s.p.s.
Preprocessing filter	10-Hz high-pass	4-Hz high-pass
Short-term average window length	0.1 s	0.25 s
Long-term average window length	10.0 s	10.0 s

Note. The extremely narrow window used for ≥ 80 s.p.s. data is more sensitive to the targeted short-duration, impulsive signals from microseismicity than the wider window used for lower sample-rate data. The parameters were chosen to ensure short-term averages were computed using at least 10 samples; however, this criteria was relaxed for the case of 80 s.p.s. data streams, of which there were only four used.

data of the AZ+ network, referred to as *ANZA_review*. This catalog contains 1,200 events, 36,608 *P* wave arrival times, and 32,590 *S* wave arrival times, each reviewed by an analyst with the goal of registering every readily identifiable phase arrival and locating all associated earthquakes.

3. Methods

3.1. Overview

The processing procedure used in this study is presented in three main phases (Figure 3), each broken down into individual steps in the following subsections: phase I—detecting and locating earthquakes; phase II—relocating earthquakes; and phase III—calculating event sizes (potencies and magnitudes) and postprocessing. For reproducibility, we strive to use open-source or free software; major software components and notes for accessing them are specified in Table S2.

3.2. Phase I—Detecting and Locating Earthquakes

3.2.1. Detecting *P* Waves

A short-term average to long-term average ratio (STA/LTA) detection algorithm (Allen, 1982) targets *P* wave arrivals by operating on vertical-component data with sample-rate dependent parameters (Table 1; the full set of parameters used are listed in Tables S3 and S4). If the sample rate is greater than or equal to 80 s.p.s., waveform data are preprocessed with a 10-Hz high-pass filter, and window lengths of 0.1 and 10.0 s are used for the STA and LTA, respectively. Data with a lower sample rate are preprocessed with a 4-Hz high-pass filter, and window lengths of 0.25 and 10.0 s are used for the STA and LTA, respectively. The STA/LTA algorithm registers a detection when the ratio of the two averages (each calculated as the root mean square signal amplitude) exceeds an onset threshold of 5 and remains above a secondary threshold of 2 for at least 20 s. The long-term average is held constant from the time the ratio exceeds the onset threshold until it falls below the secondary threshold.

We chose these parameters for detecting *P* waves after comparing results from two tests using different parameters selected based on experience. In the first test, we preprocessed all data uniformly, regardless of sample rate, with a 1- to 10-Hz band-pass filter and used window lengths of 1.0 and 10.0 s for the STA and LTA, respectively. In the second test, we processed data as described in the preceding paragraph. Validating the results of both tests against *ANZA_review* showed that the chosen sample-rate dependent parameters yield the greatest number of and most precise detections.

3.2.2. Detecting *S* Waves

Each detected *P* wave triggers a search for the corresponding *S* wave using a multistep algorithm described by Ross et al. (2016). This algorithm is based on five statistical quantities, each derived from three-component data in a sliding window: (i) the signal rectilinearity, (ii) the signal incidence angle, (iii) the STA/LTA ratio, (iv) the signal kurtosis, and (v) the first difference of the kurtosis. We preprocess waveforms with a 3- to 20-Hz band-pass filter. The algorithm derives signal rectilinearity and incidence angle from the covariance matrix (Jurkevics, 1988) in a 3-s sliding window and uses them to construct a polarizing filter designed to isolate energy with high rectilinearity and near-vertical incidence angle. The algorithm then applies this filter to preprocessed, horizontal-component data and registers an initial detection using an STA/LTA algorithm with parameters like those for detecting *P* waves. A procedure based on tandem analysis of the kurtosis and kurtosis rate(-of-change) signals then refines the initial detection. We use the same algorithm and

parameters as those presented in Ross et al. (2016) in an earlier application to data from the SJFZ region. The full set of used parameters is provided in Table S5.

3.2.3. Associating Phase Detections With Events

Phase detections are associated with events by examining the spatial coherency of the detections. Specifically, the procedure checks that a minimum of five unique stations have detections at times that agree, within a tolerance of 1.5 s, with the arrival times modeled by a 1-D travel-time calculator. For each subset of 20 consecutive detections, the procedure searches a 3-D spatial grid for a tentative source location that passes this check. If such a source location is found, the procedure creates a unique event ID and associates it with the coherent phase detections. If more than one location passes this initial check, the procedure associates the arrivals with the location that minimizes the root mean square of the difference between observed and modeled arrival times.

We chose the minimum five-station threshold after comparing results from three tests, where we varied the minimum station threshold between four and six. In these tests, events were tagged as genuine if their location matched (within a prescribed space-time tolerance) an event in *ANZA_review* and were tagged as false if they did not. A five-station threshold recovers 81% of the events in *ANZA_review*, with a 34% false-detection rate, and increasing the station threshold to 6 reduces the event-recovery rate to 75%. Most of the false detections with the adopted five-station threshold are culled in later processing steps. The full set of used parameters is provided in Table S6.

3.2.4. Locating Events Individually

After associating phase detections with events, the events are individually located using the algorithm **NonLinLoc** (Lomax et al., 2009), which implements a probabilistic, global-search for locating earthquakes in heterogeneous 3-D velocity structure. We use our hybrid 3-D velocity model and the finite difference method of Podvin and Lecomte (1991) for solving the Eikonal equation to calculate travel-time-lookup tables for each station. All events are located using these travel-time-lookup tables with default algorithm parameters for locating local events (Table S7). We report 68% (1σ) confidence intervals produced by **NonLinLoc** as absolute location uncertainties; these are estimated assuming arrival-time uncertainties can be modeled by a Gaussian distribution with mean 0.1 and 0.2 s for *P* and *S* waves, respectively.

3.3. Phase II—Relocating Earthquakes

3.3.1. Cross-Correlating Waveforms

To relocate clusters of events, we measure differential travel times for pairs of closely spaced events by cross-correlating waveforms. Each event is cross-correlated with each of its 200 nearest neighbors that occurred later in time (correlating only with later events avoids redundant calculations), and each component of data is correlated independently for each unique station-phase pair with at least one arrival associated with the event pair.

To cross-correlate a pair of traces, the travel time of the arrival from the earlier event (or the later event if no arrival registered for the earlier one) is calculated by subtracting the origin time from the arrival time. This travel time is then added to the origin time of the later event (or earlier event when appropriate) to obtain the arrival time corresponding to zero travel-time difference between the events. The traces are then band-pass filtered between 2 and 15 Hz, windowed around these arrival times (0.25 s before and 0.75 s after for *P* waves; 0.25 s before and 1.25 s after for *S* waves) and cross-correlated with a maximum offset of ± 0.5 s. The offset that produces the maximum absolute cross-correlation coefficient is recorded as the differential travel time (i.e., double difference), and measurements with a coefficient less than 0.6 are discarded.

We use the Adaptable Seismic Data Format (Krischer et al., 2016) to store waveform data, instead of a conventional format like miniSEED or SAC, for its superior scalability with respect to I/O resource consumption. This is an important technical consideration for cross-correlating a data set of this size; using a methodology designed to work with miniSEED data overwhelmed the I/O resources of the high-performance computing cluster we use, making the calculation untenable.

3.3.2. Relocating With a Double-Difference Algorithm

The differential travel times obtained by cross-correlating waveforms are input, along with an average 1-D velocity model, to the **GrowClust** (Trugman & Shearer, 2017) double-difference algorithm for relocating earthquakes. A minimum of six differential travel times are required to relocate each event pair. Aside from

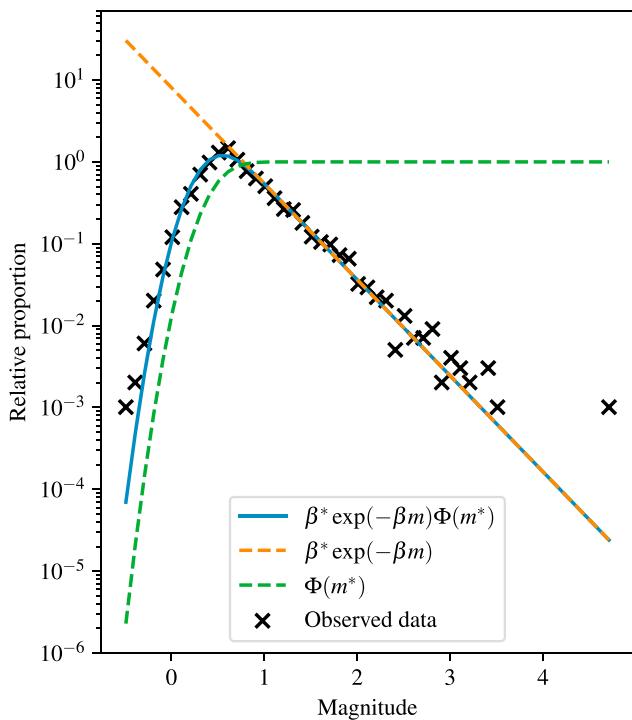


Figure 4. Example fit of an exponentially-modified Gaussian PDF (blue curve; equation (1)) to an observed frequency-magnitude distribution (black X's; data taken from *HYS_catalog_2011* for a 0.1° square centered at 33.5° N, 116.5° W) using maximum-likelihood estimation. The distribution is decomposed into two components: (i) an exponential component (orange, dashed line) representing ideal Gutenberg-Richter statistics and (ii) a Gaussian CDF component (green, dashed line) representing network sensitivity. We define the magnitude of completeness as the 99th percentile of the PDF underlying the Gaussian CDF; that is, the event magnitude with a 99% probability of being detected.

time in length. The noise level is calculated as 1.414 times the standard deviation of 10.0 s of data preceding the predicted *P* wave arrival time, and maximum amplitude measurements are discarded if they are less than three times this value.

3.4.3. Calculating Moment Magnitudes, M_w

Moment magnitudes, M_w , are estimated using a spectral-based method implemented by the **dbmw** program (Ross et al., 2016) and parameters listed in Table S10. This method estimates scalar seismic potencies and moments from the low-frequency asymptote of observed displacement spectra and then uses the scaling relation of Hanks and Kanamori (1979) to get the moment magnitude.

3.4.4. Estimating Minimum Magnitude of Catalog Completeness, M_c

The minimum magnitude of completeness for the catalog, M_c , is estimated using a method based on the work of Ogata and Katsura (1993). This involves modeling the observed frequency-magnitude distribution using an exponentially modified Gaussian probability density function (PDF) (Figure 4) defined by equation (1):

$$f_M(m; \mu, \sigma, \lambda) \equiv \lambda \exp\left(\frac{\lambda}{2}(2\mu + \lambda\sigma^2)\right) \exp(-\lambda m) \Phi\left(\frac{m - (\mu + \lambda\sigma^2)}{\sigma}\right), \quad (1)$$

where $\Phi\left(\frac{m - (\mu + \lambda\sigma^2)}{\sigma}\right)$ is a Gaussian cumulative distribution function (CDF) with mean $(\mu + \lambda\sigma^2)$ and standard deviation σ and λ is the decay-rate of the exponential component of the distribution. The exponential component in equation (1) models the classical frequency-magnitude distribution following Gutenberg-Richter statistics, and the Gaussian CDF component acts as a “thinning operator,” representing network sensitivity as a function of magnitude. We estimate parameters μ , σ , and λ by numerically maximizing the log-likelihood function (i.e., using the maximum-likelihood estimator implemented by

the minimum number of differential travel times per event pair, the used parameters (Table S8) are as those provided by the example distributed with the source code. In an earlier stage of this project, the **HypoDD** algorithm (Waldhauser & Ellsworth, 2000) was used. It was replaced by **GrowClust** because limits on computer memory and the number of tunable parameters make **HypoDD** cumbersome for a data set of this volume.

3.4. Phase III—Calculating Event Sizes and Postprocessing

3.4.1. Quality Control

To obtain a catalog that maximizes the ratio of genuine to false events detected, all events not relocated by **GrowClust** are validated by a 5-point quality-control test and retained only if they satisfy all five criteria. Events relocated by **GrowClust** bypass the quality-control test because it is unlikely that a false event will correlate sufficiently well with another event for **GrowClust** to relocate it. All events in the final catalog are either relocated by **GrowClust** or satisfy the following:

1. event has a horizontal error (reported by **NonLinLoc**) less than 3 km,
2. event has a vertical error (reported by **NonLinLoc**) less than 3 km,
3. event is recorded by at least one station within 10-km epicentral distance,
4. event has at least three *S* wave arrival detections, and
5. event has an *S* to *P* phase-picks ratio of at least 1/4.

3.4.2. Calculating Local Magnitude, M_L

Local magnitudes, M_L , are calculated using Richter's original method (Richter, 1935), as implemented by the **dbevproc** program (see Table S9 for parameters used) and accompanying module **Mlrichter.pm** of the Antelope software package (version 5.8, distributed by Boulder Real Time Technologies Inc.). The maximum signal amplitude of horizontal component data is measured, on integrated velocity data with simulated Wood-Anderson instrument response, in a window that starts 10.0 s before the predicted *P* wave arrival and is twice the predicted *S*-minus-*P*

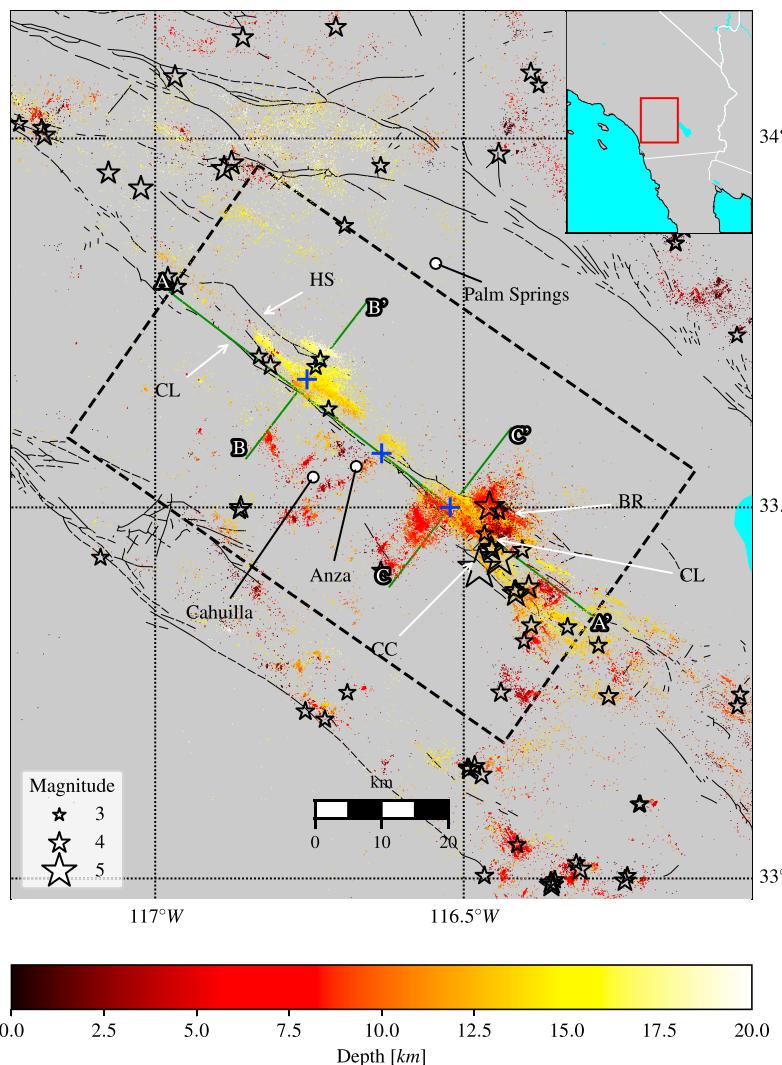


Figure 5. Map of 108,800 cataloged events in and around the focus region, with color showing depth (scale at bottom). Stars represent moderate events ($M \geq 3.5$; scale at bottom left) taken from the United States Geological Survey (USGS) catalog—USGS catalog data are used for moderate events because our processing procedure is tuned for detecting microseismic events. Surface traces of vertical transects A-A', B-B', and C-C'—plotted in Figures 10, 11, and 12, respectively—are shown as solid green lines; the center of each transect is marked by a blue +. Locations of Clark (CL), Hot Springs (HS), Coyote Creek (CC), and Buck Ridge (BR) faults are annotated with white arrows. Towns of Palm Springs, Anza, and Cahuilla are marked by white circles with black edge. The Hot Springs area (surrounding center of transect B-B') is characterized by deep seismicity (>13 km) primarily distributed inside two elongate volumes, which trend subparallel to the trend of and are displaced to the NE of the main fault trace—few moderate, and no $M \geq 4$, events occurred here. The Trifurcation area (surrounding the center of transect C-C') is characterized by networks of discrete surfaces with conjugate orientations at intermediate depths (1–13 km) and deep seismicity (>13 km) resembling structure in the Hot Springs area—deep seismicity is chiefly located between the Clark and Buck Ridge fault traces. The largest events inside the focus region occurred in the Trifurcation area.

`scipy.stats.rv_continuous.fit`; Jones et al., 2001) and define the minimum magnitude of catalog completeness as the 99th percentile of the Gaussian PDF underlying Φ .

Spatial variations in M_C are analyzed by assigning each node of a grid with 0.1° node-spacing the value of M_C estimated for all events within 0.1° epicentral distance of that node; nodes with fewer than 200 events within 0.1° are not assigned an M_C value.

3.4.5. Analyzing Similar-Event Chains

To investigate the degree of similarity between events, we build similar-event chains based on waveform similarity using a simple chaining rule like the one used by Peng and Ben-Zion (2005): If event A is similar

Table 2
SJFZ Catalog 2008–2016 Detection Counts at Different Steps in the Processing Procedure

	Before associating	After associating	After QC	After DD relocating
P detections	11,391,410	2,159,717	1,683,038	1,575,515
S detections	4,286,019	1,666,272	1,268,476	1,195,924
Events	0	160,867	108,800	102,719

Note. SJFZ = San Jacinto fault zone. *Associating*, *QC*, and *DD relocating* refer to the procedures described in sections 3.2.3, 3.4.1, and 3.3.2, respectively.

to event B, and event B is similar to event C, then events A, B, and C form a similarity chain. Events A and C will not necessarily satisfy the criteria used to assess similarity but are considered indirectly similar. Two events are considered similar in this context if the maximum cross-correlation coefficient (for either P or S waves; measured as described in section 3.3.1) exceeds a chosen threshold (we analyze results for various thresholds between 0.70 and 0.95) at a minimum of five unique stations.

4. Results

4.1. Basic Features of the Catalog

The performed analysis provides a new catalog, named *SJFZ_catalog_2008–2016*, of earthquake parameters for 160,867 events. During the quality-control stage, we discard 52,067 of these events as either spurious or poorly constrained, leaving 108,800 high-quality locations (Figure 5) with 105,762 (local/Richter) magnitudes, 97,709 seismic potencies (scaled and reported as moment magnitudes), 1,683,038 P wave arrival times, and 1,268,476 S wave arrival times (Table 2). About 63% of the retained events are inside the focus

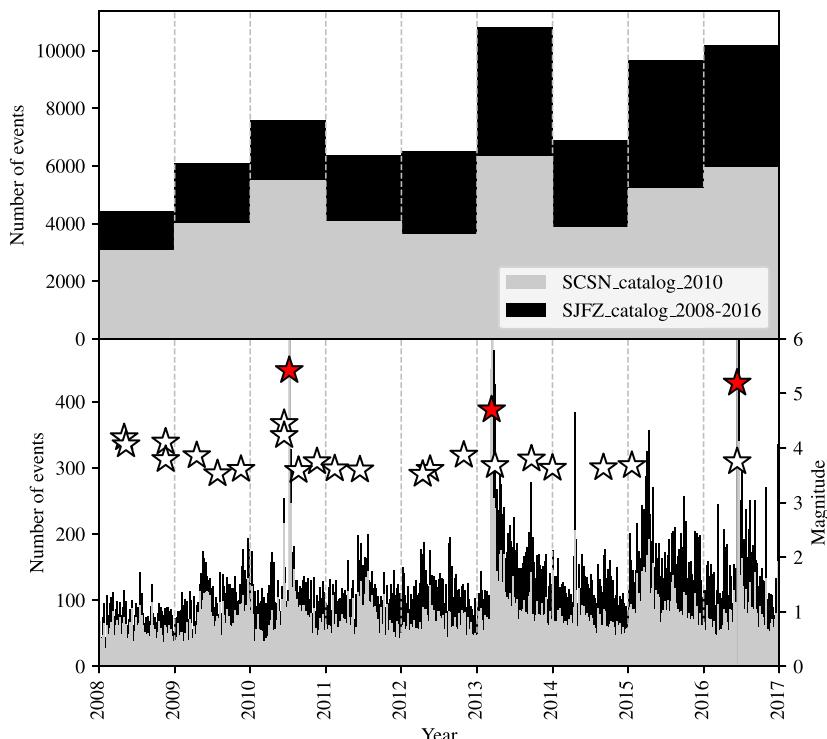


Figure 6. Events detected inside the focus region as a function of time in 1-year bins (top panel) and 1-week bins (bottom panel) for *SCSN_catalog_2010* (gray bars) and *SJFZ_catalog_2008–2016* (black bars). Over 16 events are detected in *SJFZ_catalog_2008–2016* for every 10 events in *SCSN_catalog_2010*. Moderate events ($M \geq 3.5$; taken from the USGS catalog) are shown as white stars in the bottom panel (scaled relative to the right vertical axis). The three largest events to occur inside the focus region (2010 M_w 5.4 Borrego Springs, 2013 M_L 4.7 Anza Borrego, and 2016 M_w 5.2 Borrego Springs; shown as red stars) are observed in the bottom panel as a brief spike in event detection rates followed by a decay pattern typical of aftershock sequences. Other spikes that are temporally uncorrelated with significant mainshocks (e.g., early 2014 and early 2015) are visually inspected and confirmed to be swarm activity.

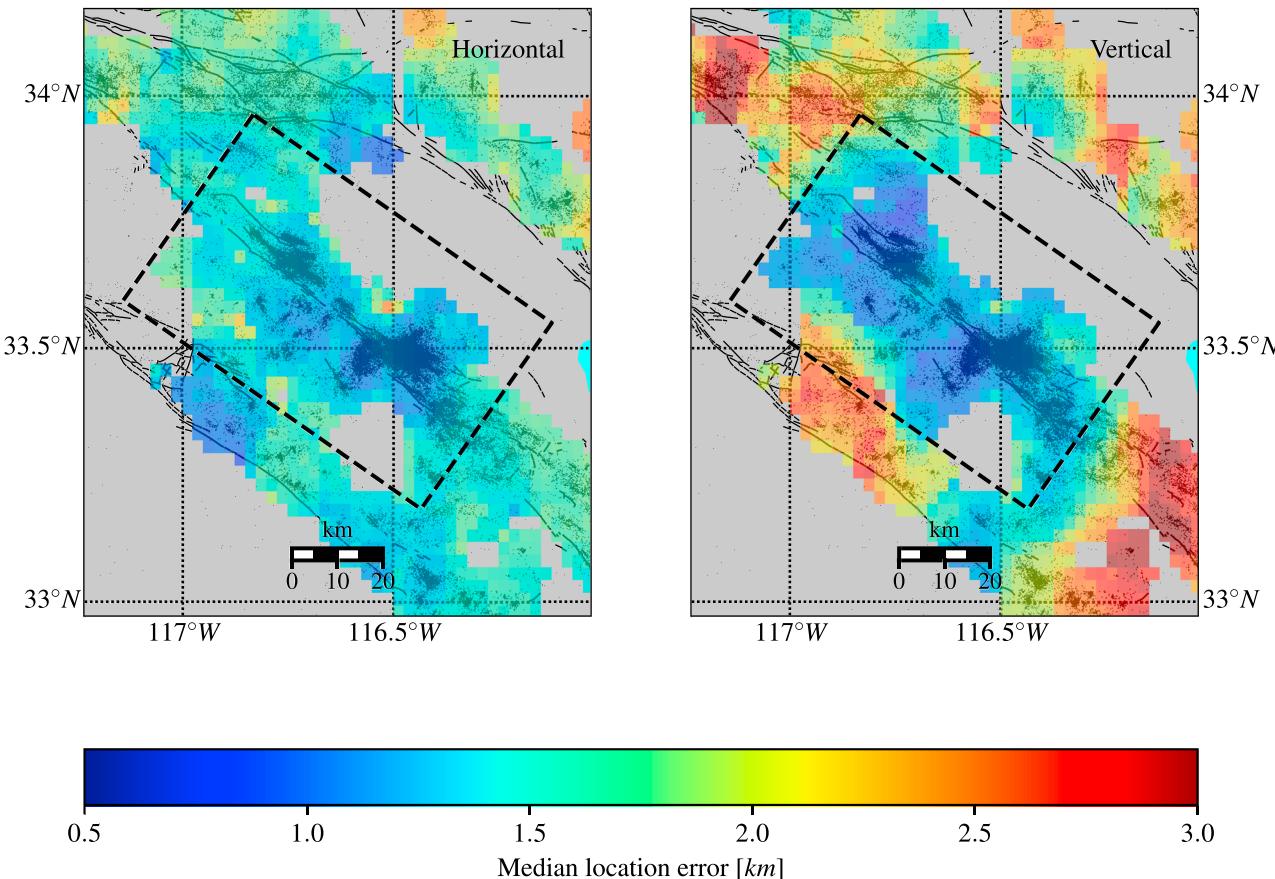


Figure 7. Spatial distribution of median absolute location errors (68% confidence intervals reported by **NonLinLoc**). The side length of each cell is 0.025° , and the cell color represents the median value for all events within 0.05° epicentral distance of the cell's center; cells with fewer than 50 associated events are assigned a null value. Inside the focus region, the median horizontal and vertical errors are 1.2 and 1.1 km, respectively, and outside of the focus region they increase to 1.8 and 2.2 km, respectively. Inside the focus region—where station density is high, many events are deep, and the velocity model is well constrained—the median vertical error is smaller than the median horizontal error. Outside the focus region, the median vertical error is larger than the median horizontal error. The dashed rectangular outline indicates our focus region, as in Figure 1.

region, with 16 events for every 10 events in *SCSN_catalog_2010* for the same region (Figure 6), and 40% of the retained events are within the Trifurcation area (TR).

Aftershocks of the three largest events in the focus region during the examined period (red stars in Figure 6) increase considerably the event detection rates, but additional spikes of seismicity rates that are temporally uncorrelated with significant events are also observed. Visual inspections of the results indicate that most of these features are produced by earthquake swarms. A short-lived (7-day) swarm ~ 3 km NNW of Cahuilla produces the largest such uncorrelated spike (early 2014), a long-lived (100-day) swarm ~ 3 km SW of the Coyote Creek (CC) fault produces the second to largest such spike (early 2015), and additional short-lived swarms are observed near the Coyote Creek and Clark (CL) faults. It is also interesting to note the variable event rates that follow the occurrence of earthquakes with $M_L \sim 4$, likely reflecting variable stress conditions.

Absolute event locations are well constrained, both horizontally and vertically, inside the focus region (Figure 7; Table 3), but vertical uncertainties increase rapidly outside the footprint of the core network coverage. Contrary to conventional assumption, the median vertical uncertainties inside the focus region are smaller than horizontal, though the opposite is true outside the focus region. The magnitudes of detected events range from -1.8 to 5.4 , and the catalog is complete above $M_L 0.6$ inside the focus region of this study. The *SCSN_catalog_2010* is complete above $M_L 0.9$ inside the focus region, so *SJFZ_catalog_2008-2016* outperforms *SCSN_catalog_2010* in the focused region. On the other hand, *SCSN_catalog_2010* outperforms the derived *SJFZ_catalog_2008-2016* in the outer region (Figure 8).

Table 3
Basic Characteristics of SJFZ_catalog_2008-2016 in Different Geographical Regions

	Study area	Focus region	HS	TR
# of events	108,800	68,642	14,814	43,006
Hor. error	1.4 km	1.2 km	1.3 km	1.1 km
Vert. error	1.5 km	1.1 km	1.1 km	1.1 km
M_C	0.8	0.6	0.7	0.5

Note. # of events is the total number of events in the catalog; Hor. error and Vert. error are median absolute horizontal and vertical location errors, respectively (68% confidence intervals reported by NonLinLoc); and M_C is the magnitude of completeness estimated by the 99th percentile of the Gaussian component of the best fitting exponentially-modified Gaussian PDF.

Event pairs in *SJFZ_catalog_2008-2016* generally exhibit weak waveform similarity (Table 4); at least 6 of every 10 events in the catalog are sufficiently unique (in terms of waveform signature) that they are omitted from every similarity chain. Fewer than 3% of events are considered highly similar (maximum cross-correlation coefficient ≥ 0.95 at five or more stations), and, regardless of the cross-correlation coefficient threshold chosen, a relatively small number of similarity chains contain most of the similar events. These similarity chains form highly localized structures in the core fault zone surrounded by diffuse clouds of dissimilar events, are strongly correlated with spatiotemporal event clusters, and give useful information on the principle seismogenic structures in the fault zone (Figures 9–12).

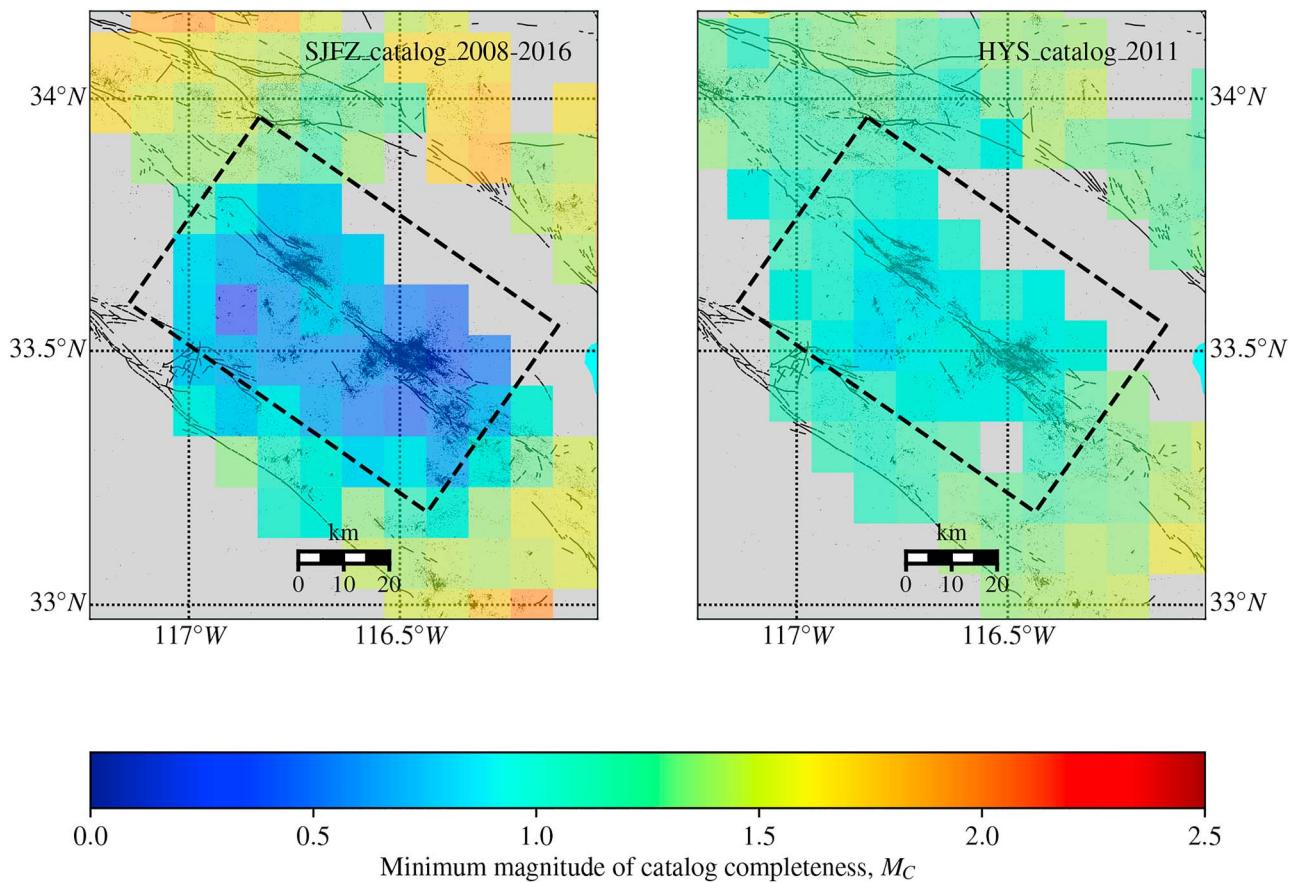


Figure 8. Spatial distribution of minimum magnitude of catalog completeness for *SJFZ_catalog_2008-2016* (left) and *HYS_catalog_2011* (right) estimated by the 99th percentile of the Gaussian component of the best fitting (in the maximum-likelihood sense) exponentially modified Gaussian PDF (equation (1)). The side length of each cell is 0.1° , and the cell color represents the magnitude of completeness for all events within 0.1° epicentral distance from the cell's center; cells with fewer than 200 associated events are assigned a null value. *SJFZ_catalog_2008-2016* and *HYS_catalog_2011* are complete above M_L 0.6 and 0.9, respectively, inside the focus region and above M_L 1.9 and 1.5, respectively, outside the focus region (and within the mapped area).

Table 4
Summary of similar events in SJFZ_catalog_2008-2016

CC thresh	N _{sim}	N _{chain}	N _{chain5}	N _{sim5}	Percent
0.95	2,880	511	121	1918	2.6
0.90	8,425	870	265	6,951	7.7
0.85	14,359	920	280	12,772	13.2
0.80	20,550	849	259	19,073	18.9
0.75	27,049	682	204	25,831	24.9
0.70	33,953	488	143	33,089	31.2

Note. CC thresh is the cross-correlation coefficient threshold applied for chain linking, N_{sim} is the total number of events belonging to similarity chains, N_{chain} is the total number of similarity chains, N_{chain5} is the number of similarity chains with at least five events, N_{sim5} is the number of events belonging to similarity chains with at least five events, and “percent” is the percent of events in the final catalog that belong to similarity chains. The number and length of similarity chains increases as the cross-correlation threshold decreases, until a threshold of 0.85 is reached, at which point the length of chains continues to increase, but the number of chains begins to decrease: relaxing the linking criteria beyond the critical threshold of 0.85 causes formerly independent chains to merge.

4.2. Structural Features in the Catalog

4.2.1. General Structural Features

We divide the SJFZ into four distinct, quasi-tabular regions based on observed characteristic patterns of seismicity (Figures 10–12). Three characteristic patterns are defined for this purpose: (i) seismic quiescence; (ii) spatiotemporally localized, discrete manifolds of seismicity; and (iii) spatiotemporally diffuse clouds of seismicity.

Region I consists of the uppermost ~1 km of crust and is nearly devoid of seismicity, although a noteworthy cluster of events is observed near the SE end of the TR. This shallow cluster is correlated spatially with a

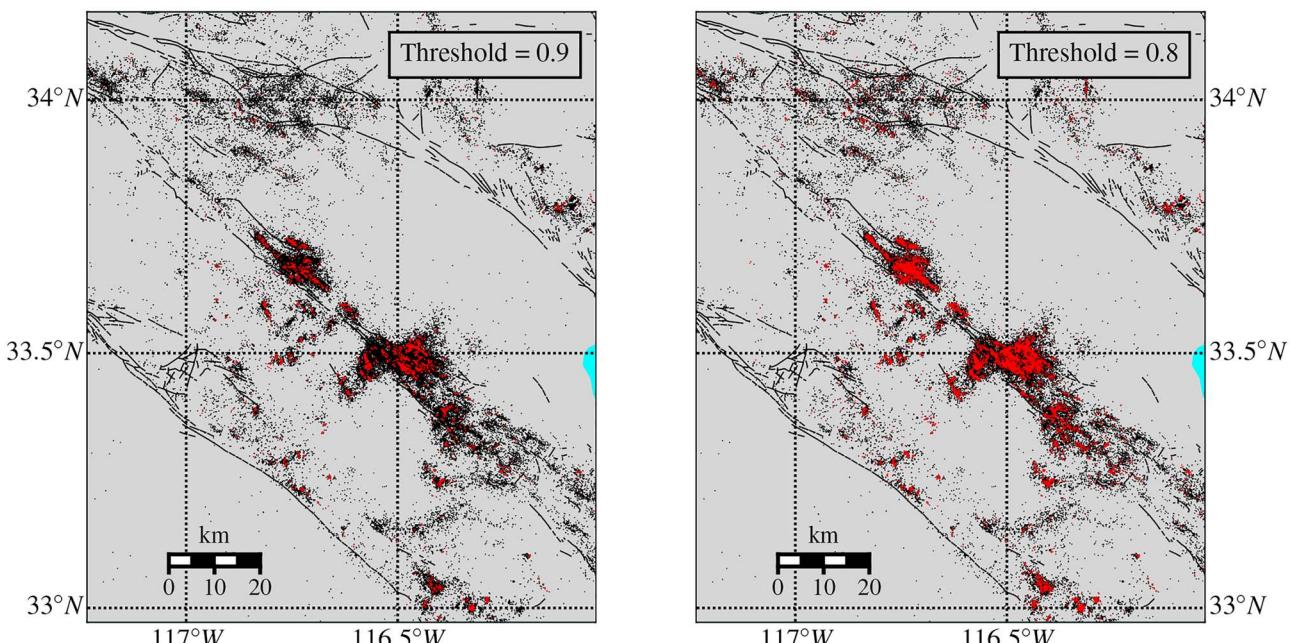


Figure 9. Map view of events belonging to similarity chains (red points) obtained for cross-correlation coefficient thresholds of 0.9 (left) and 0.8 (right); isolated events are shown as black points. Similar events tend to occur near the center of diffuse clouds of surrounding dissimilar events and form compact lineations. The relatively compact lineations formed by similarity chains (compared to diffuse surrounding events) may be an artifact of the double-difference relocation method or may be evidence of highly localized zones of repeated failure accommodating significant deformation.

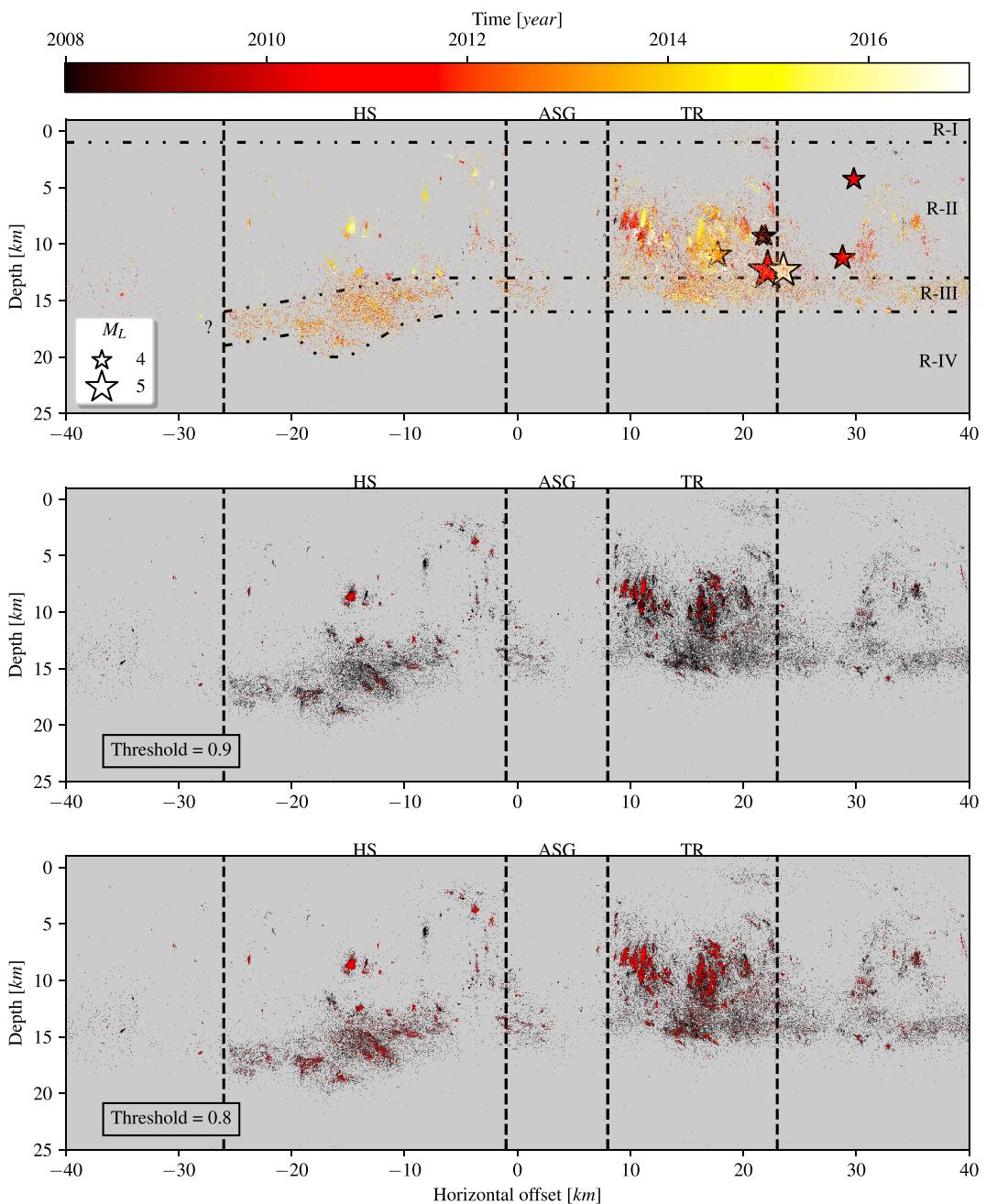


Figure 10. Vertical transects (see also Figures 11 and 12) along A-A' (Figure 5) of all seismicity within ± 15 km of the nominal fault plane collapsed onto the plane (by orthogonal projection). In the top panel, events are color coded by origin time (scale at top), significant events ($M \geq 4$; taken from USGS catalog) are shown as stars (magnitude scale at the right of panel) color coded by origin time, and boundaries of the four mechanically distinct regions we interpret are outlined by thin black dash-dotted lines and labeled R-I through R-IV. Events belonging to similarity chains with cross-correlation coefficient thresholds of 0.9 (middle panel) and 0.8 (bottom panel) are shown as red points and independent events are shown as black points (bottom two panels). Vertical dashed lines indicate the extent of the Hot Springs area (HS), the Anza Seismic Gap (ASG), and Trifurcation area (TR), as we define them. Regions I and IV are characterized by an absence of seismicity, region II is characterized by dense space-time clusters of events with similar waveforms distributed on discrete quasi-planar surfaces, and region III is characterized in this projection by a ribbon of spatiotemporally diffuse seismicity.

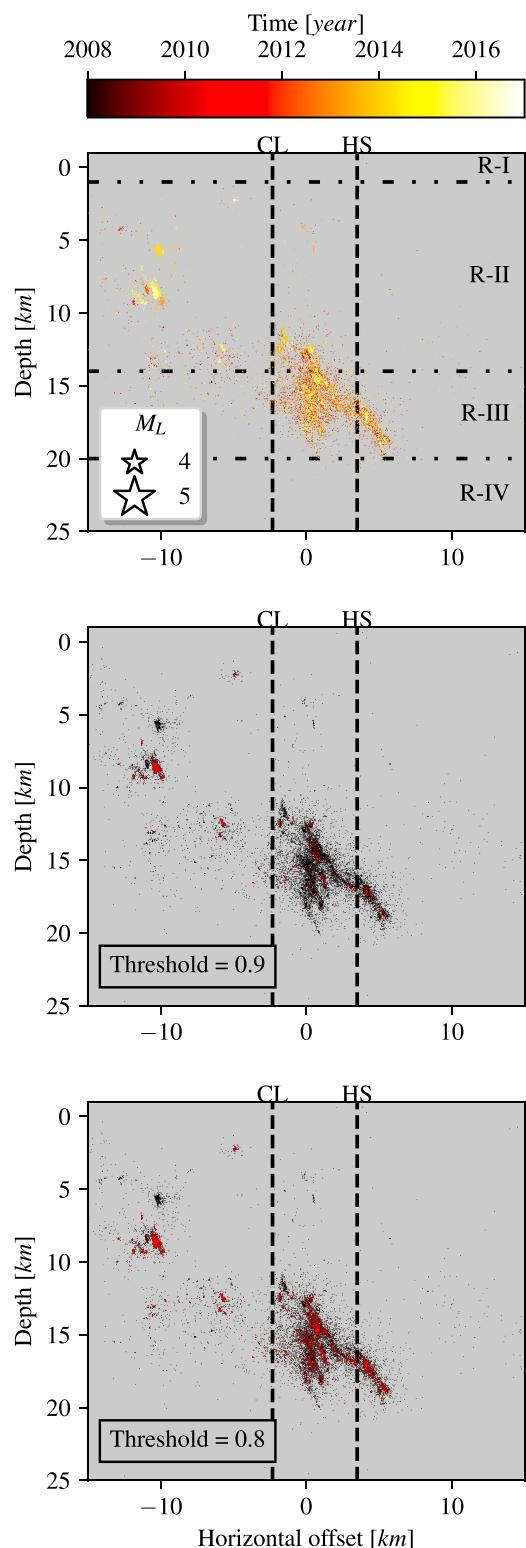


Figure 11. Vertical transects as in Figure 10 along B-B' through the Hot Springs (HS) area (Figure 5). The vertical projections of the Clark (CL) and HS faults' surface traces are marked by dashed vertical lines. There are no events $M > 4$ in this transect, and most seismicity along this transect is in region III. Region III seismicity exhibits greater waveform similarity than corresponding seismicity in Trifurcation area (Figure 12), and, from this perspective, appears to form limbs of volumetrically distributed seismicity which converge at low angles near the top of the region.

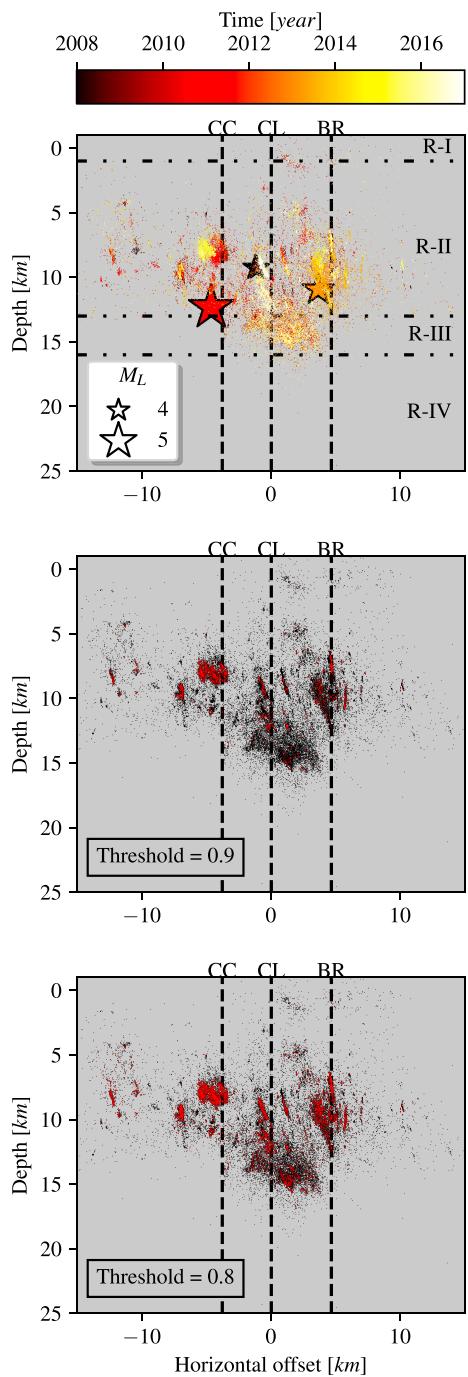


Figure 12. Vertical transects as in Figure 10 along C-C' through the Trifurcation area (Figure 5). Vertical projections of the Coyote Creek (CC), Clark (CL), and Buck Ridge (BR) faults' surface traces are marked by dashed vertical lines (the location of the Coyote Creek fault is extrapolated by continuing the mapped segment of the fault to the point of intersection with the transect). Region II seismicity along this transect reveals numerous discrete surfaces that were active during brief periods of intense seismic activity—quintessential event clusters—and these event clusters exhibit high waveform similarity (middle and bottom panels). The lateral extent of region II seismicity narrows with increasing depth and converges to a ~9-km-wide zone of diffuse region III seismicity at ~13-km depth.

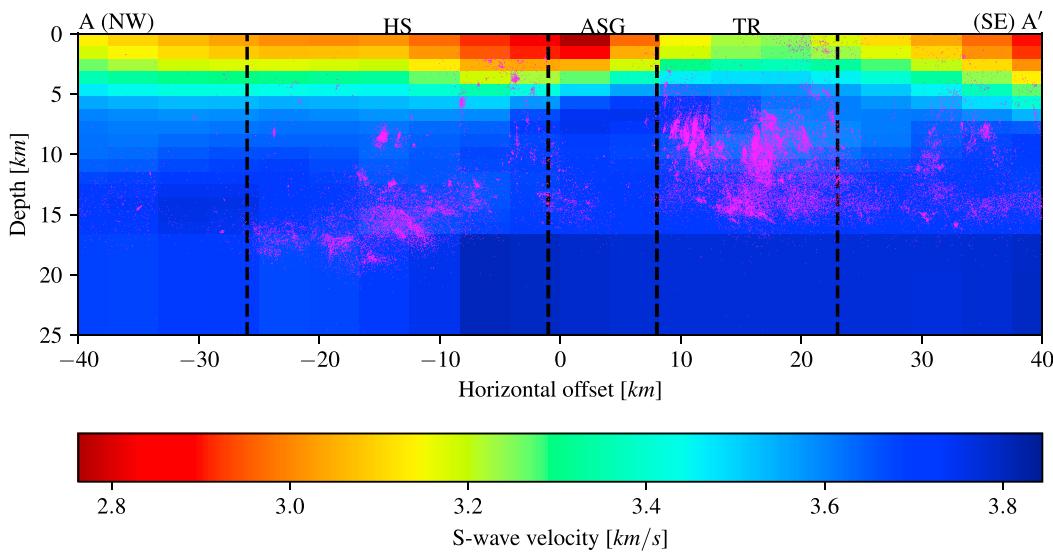


Figure 13. Vertical transect along A-A' (Figure 5) with earthquakes plotted as magenta points and the FANG16 S wave velocity model as background. Seismogenic structures are more fully illuminated in SJFZ_catalog_2008–2016 than in HYS_catalog_2011 (Figure S7), and the depth extent of the seismogenic zone between 20- and 40-km horizontal offset is greater in SJFZ_catalog_2008–2016 than in HYS_catalog_2011.

strong bilateral fault-parallel S wave-velocity gradient in FANG16-S (Figure 13) and the SW slope of the Santa Rosa Mountains.

Region II extends from the bottom of region I to ~13- to 15-km depth—the lower bound of this region varies along strike—and contains 62% of the observed seismicity in and around the SJFZ (an 80-km-long, 30-km-wide rectangular swath-oriented parallel to the main fault trace plotted in Figure 10). Seismicity in this region clusters tightly in space and time, forming angular networks of discrete surfaces with conjugate orientations along with significant gaps of seismicity (Figures 10, 12, and 14). All seven $M \geq 4$ events that occurred during the observation period nucleated in this region.

Region III populates a ~3- to 6-km ribbon beneath region II with abundant, spatiotemporally diffuse, and low-magnitude earthquakes (37% of observed seismicity in the fault zone, no $M \sim 4$ events, and only four $3 \leq M \leq 4$ events). The downward bulge near -15-km horizontal offset in Figure 10 is correlated somewhat with Thomas Mountain and the associated topographic ridge (Figure S7). Basement lithology may control this feature (Magistrale & Zhou, 1996), and it may also be related to the dip of the Moho in this region (Lewis et al., 2000; Ozakin & Ben-Zion, 2015; Zhu & Kanamori, 2000).

Region IV includes everything beneath region III and is almost completely devoid of seismicity.

The lower bound of the seismogenic zone generally shallows toward the SE, with the deepest seismicity observed in the Hot Springs area (HS). Separating the deep seismicity of HS to the NW from the generally shallower seismicity in the highly-productive TR area to the SE is the Anza Seismic Gap (ASG) of microseismicity (Sanders & Kanamori, 1984). The ASG area is associated with a highly localized section of the SJFZ with paleoseismic evidence of moderate and large events (Rockwell et al., 2015), a high-velocity contrast across the fault (Allam & Ben-Zion, 2012; Share et al., 2019), and a strong overlying velocity gradient in FANG16-S (Figure 13).

4.2.2. Hot Springs Area

The Hot Springs area is predominantly populated by seismicity in region III (Figure 11); 78% of HS seismicity is in region III. The deepest earthquakes observed in the SJFZ, and the only $M \geq 3.5$ earthquakes (four events) observed in region III, are in HS. The earthquakes in HS region III include anastomosing NE dipping limbs of diffusely distributed seismicity that converge at low angles near the top of the region (Figure 11; Movie S1) where seismicity rates quickly drop. Region III seismicity in HS exhibits stronger waveform similarity than corresponding region III seismicity in TR (Figures 10–12). Region II seismicity is sparse in HS and is mostly located ~10 km SW of the main fault segments.

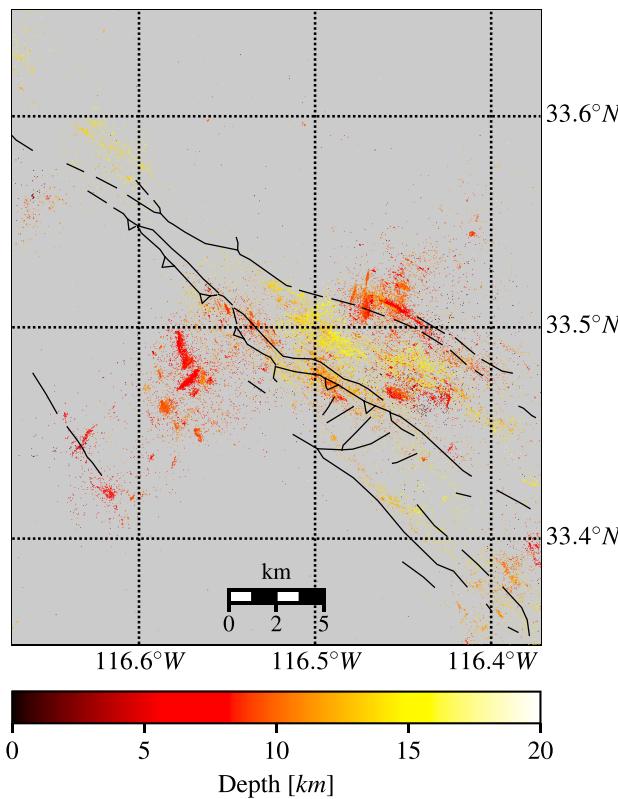


Figure 14. Map of seismicity in the Trifurcation area showing intermediate-depth “shatter networks”—networks of quasi-planar fracture surfaces with conjugate orientations. Shatter networks are distributed suborthogonally to the trend of the main fault trace and produce most of the similar waveforms observed.

4.2.3. Trifurcation Area

The TR (Figure 12) is the most seismically active portion of the SJFZ, with abundant seismicity in regions II and III; 66% of core fault-zone seismicity is in TR. Seismicity in TR is asymmetrically distributed between regions II (75%) and III (24%) and depicts “shatter networks”—principal sets of commonly oriented discrete fracture surfaces coexisting with suborthogonal conjugate surfaces—distributed along a trend suborthogonal to the main fault trace (Figure 14). Region II seismicity in TR includes spatiotemporally dense event clusters, comprised of events with highly similar waveforms, that are dipping to the NE with angles that become shallower below about 10 km. Region III seismicity in TR, however, is laterally distributed across a ~9-km swath below the three major faults in the area and exhibits greater waveform diversity than the overlying region II seismicity.

Certain events occurring on conjugate surfaces in TR produce highly similar waveforms (cross-correlation threshold ≥ 0.9), and similarity chains with a threshold of 0.8 linking events spanning > 10 km and all three major TR faults are observed (Figure 15). Waveforms produced by events from such chains with large separation, however, show only moderate similarity (maximum cross-correlation coefficient of ~ 0.5), and the coherent energy of these coincides with the highest-amplitude signal. Coherent, high-amplitude signals like this may provide information on regional processes that unify events belonging to similarity chains.

4.2.4. Anza Seismic Gap

The Anza Seismic Gap (Figure 10), as considered here, is a region nearly devoid of seismic activity between TR and HS. It extends ~ 9.5 km in the along-fault direction in region II, spans the entire seismogenic zone, and narrows to ~ 4 km at the base of region III. Its SE boundary is vertical and is met with intense intermediate-depth TR activity while its NW boundary dips steeply to the SE and is met with modest shallow-to-intermediate-depth HS activity. Region III seismicity rates are significantly reduced outside ASG on either side.

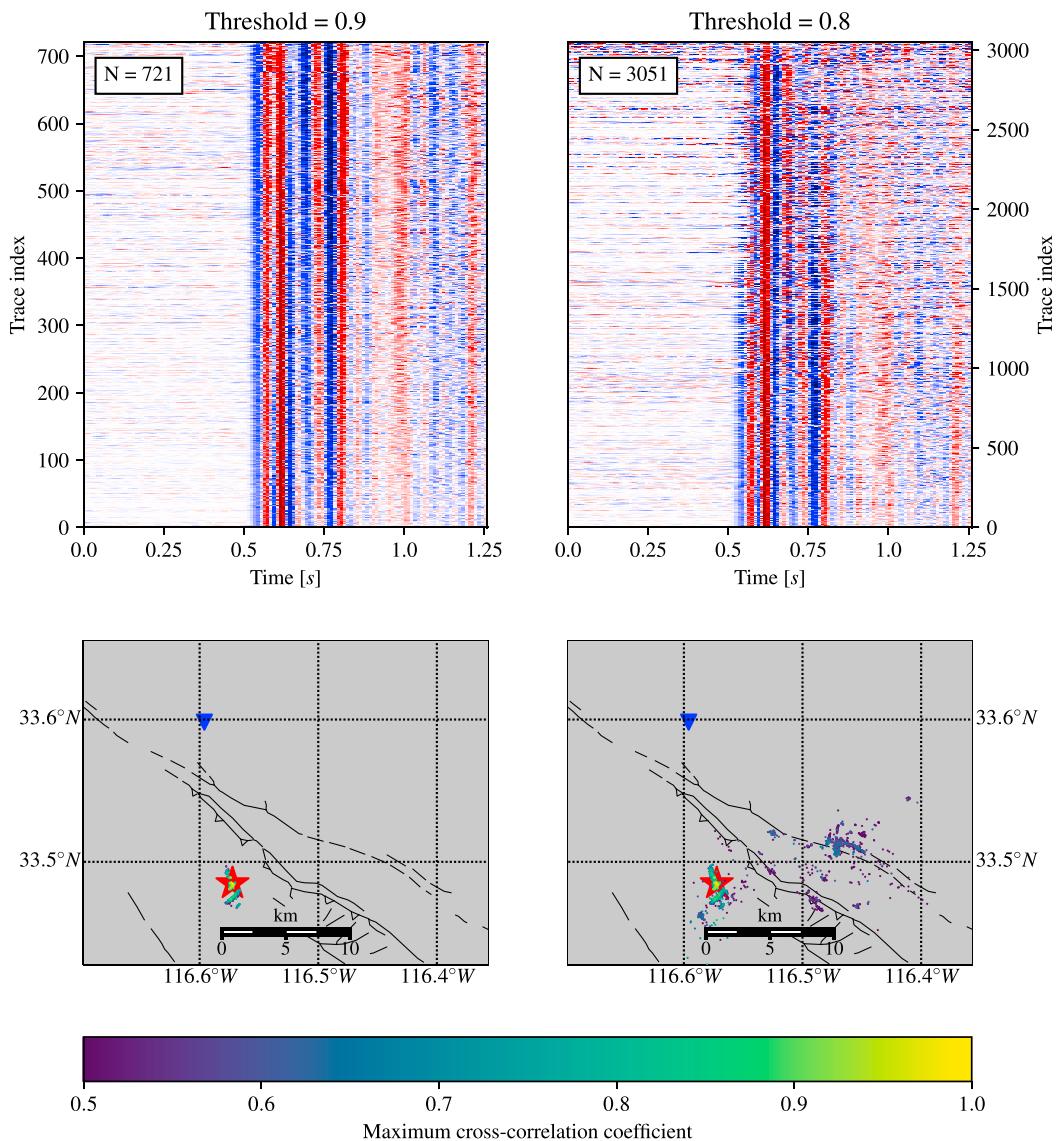


Figure 15. P-wave waveforms (top two panels) and event location maps (bottom two panels) for the largest similarity chains obtained with cross-correlation coefficient thresholds of 0.9 (left two panels) and 0.8 (right two panels). Similar P-wave waveforms observed at station PB.B082 (inverted blue triangle on location maps) are cross-correlated and aligned with a template event trace (marked by a red star on location maps; trace index 0 on waveform plots), and then sorted in order of decreasing similarity (similarity with template decreases as trace index increases in waveform plots). Event markers in location maps are color-coded by maximum cross-correlation coefficient with respect to the template trace. A modest reduction of the similarity-chain cross-correlation threshold from 0.9 to 0.8 results in a continuous chain of similar events linking events that are separated by >10 km epicentral distance and span all three major fault traces. Despite the continuity of the similarity chain in this instance, waveforms from event pairs with large inter-event separation are only moderately similar (maximum cross-correlation coefficient ~ 0.5).

5. Discussion

5.1. Characteristics of the Catalog

The developed new catalog complements an ongoing suite of investigations targeting the SJFZ: It provides an enhanced regional seismological framework to contextualize highly detailed studies of internal SJFZ structures (Qiu et al., 2017, 2018; Share et al., 2017); it provides an improved data set for repeating tomographic inversions at higher resolution (Allam & Ben-Zion, 2012; Allam et al., 2014; Fang et al., 2016); and it provides a more complete catalog for testing different mechanical models of the deforming lithosphere (e.g., Ben-Zion & Lyakhovsky, 2006; Cheng et al., 2018; Wdowinski, 2009). The resolving power and coverage of the derived results bridge the gap between the internal fault-zone studies at specific sites, and the

lower-resolution high-coverage tomographic inversions. At present, *SJFZ_catalog_2008-2016* is the most complete set of observed seismic activity in the SJFZ for 2008–2016, and continuing operation of key stations in the SJFZ Experiment Network will allow the catalog to be extended beyond 2016.

We compare *SJFZ_catalog_2008-2016* with the currently best regional catalog, *HYS_catalog_2011*, for Southern California. *HYS_catalog_2011* is derived from *SCSN_catalog_2010*, so the catalog derived here has the same relative increase in the number of events (about 60% more events than *HYS_catalog_2011*). Significant along-strike structural features of the two catalogs are generally similar (Figures 13 and S3); however, their expression in *SJFZ_catalog_2008-2016* is fuller. Despite general similarities, seismicity shallows toward the SE more rapidly in *HYS_catalog_2011*. This bears important implications on the depth of the seismogenic zone, and thus maximum likely rupture dimensions in the SJFZ. Our results are likely to be more reliable in our focus region (rectangle in Figure 1) for two reasons: We locate events using (i) increased station density and (ii) a higher-resolution velocity model. *SJFZ_catalog_2008-2016* provides a better framework than *HYS_catalog_2011* for most studies inside the SJFZ; however, *HYS_catalog_2011* remains the standard regional catalog around the SJFZ. Additionally, we recommend using the *HYS_catalog_2011* if results are needed prior to 2008, or if high-accuracy (i.e., analyst-reviewed) arrival times are needed.

SJFZ_catalog_2008-2016 is a fully automated catalog targeting microseismic events, and it is impractical to manually review each phase detection or even each event. This will become increasingly relevant as time passes; seismic networks will mature, technology will advance, and methodology will improve, resulting in larger earthquake catalogs. Although *SJFZ_catalog_2008-2016* is comparable to (and in some regards better than) *SCSN_catalog_2010* and *HYS_catalog_2011*, it is a categorically different catalog because of its fully automated nature. Despite extensive efforts to tune our processing procedure, the employed automated detection algorithms are not as good as trained human analysts and can make imprecise or erroneous observations. An occasional secondary phase (e.g., reflection or *P*-to-*S* converted phase) may trip the *S* wave detector. Fault-zone head wave arrivals at near-fault stations (e.g., Allam et al., 2014; Ben-Zion, 1989; Ross & Ben-Zion, 2014) or short bursts of noise may be mislabeled as a *P* wave arrival. STA/LTA algorithms tend to pick arrivals later than the true arrival time, and this type of error is greater for emergent arrivals than impulsive. By comparing 16,597 *P* and 12,015 *S* wave arrival times from *SJFZ_catalog_2008-2016* with corresponding arrival times in *ANZA_review*, we estimate that the average errors associated with determining phase arrival times are 0.07 and 0.14 s for *P* and *S* waves, respectively—average errors are estimated by computing the mean arrival-time residual relative to analyst-reviewed observations. New phase-picking techniques using machine-learning algorithms (e.g., Meier et al., 2018) may approach the accuracy of trained human analysts and be incorporated into future automated workflows of the type used in this work.

The absolute event-location uncertainty for individual events in *SJFZ_catalog_2008-2016* is likely greater than in *SCSN_catalog_2010* and its derivative *HYS_catalog_2011*, because of arrival-time errors. However, the relative event locations in *SJFZ_catalog_2008-2016* may be more precise than in *HYS_catalog_2011*, because similar automated procedures were used for cross-correlating waveforms from both catalogs, and *SJFZ_catalog_2008-2016* leverages increased station density to constrain locations. Furthermore, if we treat absolute arrival-time errors as symmetrically distributed (this is an approximation that ignores the effect of skewness), they add random errors to event locations; that is, we would expect them to increase the diffusivity of event clusters without imparting a systematic shift to the cluster centroid. If the cluster centroid is accurate, as expected since we use a detailed 3-D velocity model, the double-difference procedure should (at least partially) correct random location errors by pulling events back to their relative position in the cluster. If the location of an event is accurate relative to its cluster centroid, and the cluster centroid is accurate in an absolute sense, then the event location is also accurate in an absolute sense.

The percent of events producing highly similar waveforms in the SJFZ (Figures 9–12) is far lower than observed for seismicity associated with highly localized sections of large displacement faults. Most notably, these include the creeping section of the San Andreas fault (e.g., McGuire & Ben-Zion, 2005; Nadeau et al., 1994; Rubin & Gillard, 2004) and the Calaveras fault in northern California (e.g., Schaff et al., 2002). Clusters of events with highly similar waveforms in other locations such as the Marmara region of the North Anatolian fault were assumed to reflect aseismic creep (e.g., Bohnhoff et al., 2017). It is possible that some of the highly similar chains of similar events in the SJFZ are associated with some aseismic creep, although the geometrical character of the similar-event clusters observed here are significantly more complex than in the studies mentioned above. The chains of events producing highly similar waveforms in the SJFZ may

be associated with some forms of earthquake swarms as reported for events in Japan (e.g., Ito, 1990). Some although not all the similar-event clusters overlap with reports of earthquake swarms in the SJFZ (e.g., Vidale & Shearer, 2006).

The weak waveform similarity of events in SJFZ implies that many of the new events that we detect will not be recovered by common template-matching techniques because it is unlikely that the waveforms from these events will be similar to any template waveforms in existing catalogs. Common template-matching techniques are particularly well suited for detecting very small events with waveforms that closely resemble waveforms of previously detected (template) events, but a diverse library of initial templates is needed to maximize the efficacy of template-based processing. Diversifying the available library of templates in SJFZ is a key contribution of *SJFZ_catalog_2008-2016*. The existence of continuous similarity chains that link distant dissimilar events (Figure 15) suggests that a recursive adaptation to the template-matching technique, as in Frank and Abercrombie (2018), may be well suited for the SJFZ.

The automated procedure that we developed to derive *SJFZ_catalog_2008-2016* begins with raw waveform data and network meta-data. Because the procedure does not depend on any a priori information about seismicity (e.g., template waveforms) in the target region, it is applicable to diverse data sets and is particularly well suited for detailed reconnaissance of new field areas.

5.2. Characteristics of the Fault Zone

The brittle portion of a fault zone, as reflected by microseismicity, is generally correlated with heat flow and rheological properties of the rocks (e.g., Ben-Zion, 2008; Sibson, 1982; Scholz, 2002). The absence of seismicity in most of region I is consistent with similar observations at many other locations. This is typically interpreted as reflecting velocity-strengthening frictional properties in the top few kilometers (e.g., Blanpied et al., 1991; Rice, 1993), but can also reflect relatively thick fault zone at shallow depth (Hillers et al., 2006) and the general low confining pressure inhibiting the shallow crust from accumulating high elastic strain energy. There is, however, a noteworthy cluster of shallow events at the SE end of TR, which correlates with the Clark and Buck Ridge faults, a lateral S wave velocity gradient in FANG16-S (Figure 13), and the SW slope of the Santa Rosa Mountains. The increased confining pressure at shallow depths caused by the Santa Rosa Mountains may sufficiently increase the brittle strength of the underlying crust to allow detectable events to nucleate.

Region II corresponds with the geodetically-inferred locking depth along the SJFZ (Fialko, 2006; Wdowinski, 2009). Strain energy in this region is released mostly through brittle fracture, although the region in the Anza Seismic Gap is a clear exception, and it is interesting to note that Inbal et al. (2017) detected triggered aseismic slip below and around the ASG. The seismicity in region II is associated with seismic quiescence punctuated by brief events of intense deformation on networks of quasi-planar surfaces. All three $M > 4$ events occur within 2 km of the lower boundary of region II.

Region III is where the brittle-strength and the ductile-strength of the comprising rocks are similar and produce a brittle-to-ductile transition zone (Abolfathian et al., 2019; Kohlstedt et al., 1995; Sibson, 1986). Strain energy accumulated in region III is released during both brittle and ductile failure events. Accelerated rock healing and partial ductile failures limit the rate of brittle-failure events and development of highly localized fracture surfaces, thus producing diffuse clouds of distributed seismicity. Since some strain energy in this region is released aseismically, only four $M \geq 3$ events occur in region III (all inside HS).

The ductile strength of rocks in region IV is sufficiently lower than the brittle strength that hypocenters do not occur here, although rupture zones and early aftershocks of large earthquakes may occur in this region (e.g., Ben-Zion & Lyakhovsky, 2006; Jiang & Lapusta, 2017; Rolandone et al., 2004).

6. Conclusions

We integrate 9 years of data (2008–2016) from five seismic networks operating near the SJFZ, then develop and apply an automated processing procedure to derive a new catalog—called *SJFZ_catalog_2008-2016*—of earthquake locations, magnitudes, and potencies from raw waveform data. Our catalog contains 108,800 earthquakes in the magnitude range -1.8 to 5.4 , each located using a detailed 3-D velocity model, and is complete above $M_L 0.6$ inside our focus region. The obtained catalog has the most detailed information on seismicity in the focused study region between 2008 and 2016. The events contained in the catalog may be used as templates to detect additional events (e.g., Gibbons & Ringdal, 2006; Shelly et al., 2016). This can

increase the size of the catalog by a factor of 10 or more (e.g., Ross et al., 2017, 2019) and may be the subject of a follow-up work. The derived catalog illuminates seismogenic structures in the SJFZ in unprecedented detail and includes both deeper and shallower seismicity than in *HYS_catalog_2011*. The brittle-to-ductile transition zone is evident in our results as a ~3- to 6-km ribbon of spatiotemporally diffuse seismicity at the bottom of the seismogenic zone.

Acknowledgments

The data set used in this study is a result of decades of hard labor by the many individuals and teams operating seismic networks in our study area. The operators and maintainers of data archives and data retrieval tools at the Southern California Earthquake Data Center (SCEDC; <http://scedc.caltech.edu>) and IRIS-DMC were also a critical part of this project. The employed data from the AZ, PB, SB, and YN networks can be obtained from the IRIS-DMC (<https://ds.iris.edu/ds/nodes/dmc>), and the employed data from the CI network can be obtained using SCEDC's STP interface. Data Set S2 comprises the presented catalog of earthquake locations, location uncertainty estimates, and magnitudes in white-space-delimited ASCII format. We thank Zach Ross and Chris Johnson for discussions and contributions in earlier phases of the work. We also thank Natalia Ruppert and an anonymous reviewer for their constructive feedback during the review process. We acknowledge BRTT Inc. for generously providing free software licenses. The study was supported by the National Science Foundation Grants EAR-1818589 (USC), EAR-1722561 (USC), and EAR-0908042 (UCSD).

References

- Abolfathian, N., Martínez-Garzón, P., & Ben-Zion, Y. (2019). Spatiotemporal variations of stress and strain parameters in the San Jacinto fault zone. *Pure and Applied Geophysics*, 176, 1145–1168.
- Allam, A., & Ben-Zion, Y. (2012). Seismic velocity structures in the southern California plate-boundary environment from double-difference tomography. *Geophysical Journal International*, 190(2), 1181–1196.
- Allam, A., Ben-Zion, Y., Kurzon, I., & Vernon, F. (2014). Seismic velocity structure in the Hot Springs and Trifurcation areas of the San Jacinto fault zone, California, from double-difference tomography. *Geophysical Journal International*, 198, 978–999. <https://doi.org/10.1093/gji/ggu176>
- Allam, A., Ben-Zion, Y., & Peng, Z. (2014). Seismic imaging of a bimaterial interface along the Hayward Fault, CA, with fault zone head waves and direct P arrivals. *Pure and Applied Geophysics*, 171(11), 2993–3011. <https://doi.org/10.1007/s00024-014-0784-0>
- Allen, R. (1982). Automatic phase pickers: Their present use and future prospects. *Bulletin of the Seismological Society of America*, 72(6), 173–186.
- Ben-Zion, Y. (1989). The response of two joined quarter spaces to SH line sources located at the material discontinuity interface. *Geophysical Journal International*, 98(2), 213–222.
- Ben-Zion, Y. (2008). Collective behavior of earthquakes and faults. *Reviews of Geophysics*, 46, RG4006. <https://doi.org/10.1029/2008RG000260>
- Ben-Zion, Y., & Lyakhovsky, V. (2006). Analysis of aftershocks in a lithospheric model with seismogenic zone governed by damage rheology. *Geophysical Journal International*, 165(1), 197–210. <https://doi.org/10.1111/j.1365-246X.2006.02878.x>
- Blanpied, M. L., Lockner, D. A., & Byerlee, J. D. (1991). Fault stability inferred from granite sliding experiments at hydrothermal conditions. *Geophysical Research Letters*, 18(4), 609–612.
- Bohnhoff, M., Wollin, C., Domigall, D., Küperkoch, L., Martínez-Garzón, P., Kwiatek, G., & Malin, P. E. (2017). Repeating Marmara Sea earthquakes: Indication for fault creep. *Geophysical Journal International*, 210(1), 332–339. <https://doi.org/10.1093/gji/ggx169>
- California Institute of Technology, & United States Geological Survey Pasadena (1926). Southern California Seismic Network. <https://doi.org/10.7914/SN/CI>
- Cheng, Y., Ross, Z. E., & Ben-Zion, Y. (2018). Diverse volumetric faulting patterns in the San Jacinto fault zone. *Journal of Geophysical Research: Solid Earth*, 123(6), 5068–5081. <https://doi.org/10.1029/2017JB015408>
- Fang, H., Zhang, H., Yao, H., Allam, A., Zigone, D., Ben-Zion, Y., et al. (2016). A new algorithm for three-dimensional joint inversion of body wave and surface wave data and its application to the Southern California plate boundary region. *Journal of Geophysical Research: Solid Earth*, 121, 3557–3569. <https://doi.org/10.1002/2015JB012702>
- Fialko, Y. (2006). Interseismic strain accumulation and the earthquake potential on the southern San Andreas fault system. *Nature*, 441(7096), 968–971.
- Frank, W. B., & Abercrombie, R. E. (2018). Adapting the matched-filter search to a wide-aperture network: An aftershock sequence and an earthquake swarm in Connecticut. *Bulletin of the Seismological Society of America*, 108(1), 524–532. <https://doi.org/10.1785/0120170190>
- Gibbons, S. J., & Ringdal, F. (2006). The detection of low magnitude seismic events using array-based waveform correlation. *Geophysical Journal International*, 165(1), 149–166. <https://doi.org/10.1111/j.1365-246X.2006.02865.x>
- Hanks, T. C., & Kanamori, H. (1979). A moment magnitude scale. *Journal of Geophysical Research*, 84(B5), 2348. <https://doi.org/10.1029/JB084iB05p02348>
- Hauksson, E. (2000). Crustal structure and seismicity distribution adjacent to the Pacific and North America plate boundary in southern California. *Journal of Geophysical Research*, 105(B6), 13,875–13,903.
- Hauksson, E., & Shearer, P. (2005). Southern California hypocenter relocation with waveform cross-correlation, part 1: Results using the double-difference method. *Bulletin of the Seismological Society of America*, 95(3), 896–903.
- Hauksson, E., Yang, W., & Shearer, P. M. (2012). Waveform relocated earthquake catalog for Southern California (1981 to June 2011). *Bulletin of the Seismological Society of America*, 102(5), 2239–2244.
- Hillers, G., Ben-Zion, Y., & Mai, P. M. (2006). Seismicity on a fault controlled by rate- and state-dependent friction with spatial variations of the critical slip distance. *Journal of Geophysical Research*, 111, B01403. <https://doi.org/10.1029/2005JB003859>
- Hutton, K., Woessner, J., & Hauksson, E. (2010). Earthquake monitoring in Southern California for seventy-seven years (1932–2008). *Bulletin of the Seismological Society of America*, 100(2), 423–446.
- Inbal, A., Ampuero, J.-P., & Avouac, J.-P. (2017). Locally and remotely triggered aseismic slip on the central San Jacinto Fault near Anza, CA, from joint inversion of seismicity and strainmeter data. *Journal of Geophysical Research: Solid Earth*, 122, 3033–3061. <https://doi.org/10.1002/2016JB013499>
- Ito, A. (1990). Earthquake swarm activity revealed from high-resolution relative hypocenters—Clustering of microearthquakes. *Tectonophysics*, 175(1–3), 47–66. [https://doi.org/10.1016/0040-1951\(90\)90129-V](https://doi.org/10.1016/0040-1951(90)90129-V)
- Jiang, J., & Lapusta, N. (2017). Connecting depth limits of interseismic locking, microseismicity, and large earthquakes in models of long-term fault slip. *Journal of Geophysical Research: Solid Earth*, 122, 6491–6523. <https://doi.org/10.1002/2017JB014030>
- Jones, E., Oliphant, T., & Peterson, P. (2001). SciPy: Open source scientific tools for Python. Retrieved from <http://www.scipy.org/>, accessed 2019-06-17.
- Jurkevics, A. (1988). Polarization analysis of three-component array data. *Bulletin of the Seismological Society of America*, 78(5), 1725–1743.
- Kennett, B. L. N., & Engdahl, E. R. (1991). Traveltimes for global earthquake location and phase identification. *Geophysical Journal International*, 105, 429–465.
- Kohlstedt, D. L., Evans, B., & Mackwell, S. J. (1995). Strength of the lithosphere: Constraints imposed by laboratory experiments. *Journal of Geophysical Research*, 100, 17,587–17,602. <https://doi.org/10.1029/95JB01460>
- Krischer, L., Smith, J., Lei, W., Lefebvre, M., Ruan, Y., de Andrade, E. S., et al. (2016). An adaptable seismic data format. *Geophysical Journal International*, 207(2), 1003–1011.

- Lewis, J. L., Day, S. M., Magistrale, H., Eakins, J., & Vernon, F. (2000). Regional crustal thickness variations of the Peninsular Ranges, southern California. *Geology*, 28(4), 303–306.
- Lin, G., Shearer, P. M., & Hauksson, E. (2007). Applying a three-dimensional velocity model, waveform cross correlation, and cluster analysis to locate southern California seismicity from 1981 to 2005. *Journal of Geophysical Research*, 112, B12309. <https://doi.org/10.1029/2007JB004986>
- Lomax, A., Michelini, A., & Curtis, A. (2009). *Earthquake location, direct, global-search methods* Edited by R. A. Meyers. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Magistrale, H., & Zhou, H.-W. (1996). Lithologic control of the depth of earthquakes in Southern California. *Science*, 273(5275), 639–642.
- McGuire, J., & Ben-Zion, Y. (2005). High-resolution imaging of the Bear Valley section of the San Andreas fault at seismogenic depths with fault-zone head waves and relocated seismicity. *Geophysical Journal International*, 163(1), 152–164. <https://doi.org/10.1111/j.1365-246X.2005.02703.x>
- Meier, M. A., Ross, Z. E., Ramachandran, A., Balakrishna, A., Nair, S., Kundzicz, P., et al. (2018). Reliable real-time seismic signal/noise discrimination with machine learning. *Journal of Geophysical Research: Solid Earth*, 124, 788–800. <https://doi.org/10.1029/2018JB016661>
- Nadeau, R., Antolik, M., Johnson, P. A., Foxall, W., & McEvilly, T. V. (1994). Seismological studies at Parkfield III: Microearthquake clusters in the study of fault-zone dynamics. *Bulletin of the Seismological Society of America*, 84(2), 247–263.
- Ogata, Y., & Katsura, K. (1993). Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogues. *Geophysical Journal International*, 113(3), 727–738. <https://doi.org/10.1111/j.1365-246X.1993.tb04663.x>
- Ozakin, Y., & Ben-Zion, Y. (2015). Systematic receiver function analysis of the Moho geometry in the Southern California plate-boundary region. *Pure and Applied Geophysics*, 172(5), 1167–1184.
- Peng, Z., & Ben-Zion, Y. (2005). Spatiotemporal variations of crustal anisotropy from similar events in aftershocks of the 1999 M7.4 Izmit and M7.1 Duzce, Turkey, earthquake sequences. *Geophysical Journal International*, 160(3), 1027–1043. <https://doi.org/10.1111/j.1365-246X.2005.02569.x>
- Podvin, P., & Lecomte, I. (1991). Finite difference computation of travel times in very contrasted velocity models: A massively parallel approach and its associated tools. *Geophysical Journal International*, 105(1), 271–284.
- Qin, L., Ben-Zion, Y., Qiu, H., Share, P.-E., Ross, Z. E., & Vernon, F. L. (2018). Internal structure of the San Jacinto fault zone in the trifurcation area southeast of Anza, California, from data of dense seismic arrays. *Geophysical Journal International*, 213(1), 98–114.
- Qiu, H., Ben-Zion, Y., Ross, Z. E., Share, P.-E., & Vernon, F. L. (2017). Internal structure of the San Jacinto fault zone at Jackass Flat from data recorded by a dense linear array. *Geophysical Journal International*, 209(3), 1369–1388.
- Rice, J. R. (1993). Spatio-temporal complexity of slip on a fault rate-dependent friction. *Journal of Geophysical Research*, 98(B6), 9885–9907.
- Richter, C. F. (1935). An instrumental earthquake magnitude scale. *Bulletin of the Seismological Society of America*, 25, 1–32.
- Rockwell, T. K., Dawson, T. E., Young Ben-Horin, J., & Seitz, G. (2015). A 21-event, 4,000-year history of surface ruptures in the Anza seismic gap, San Jacinto Fault, and implications for long-term earthquake production on a major plate boundary fault. *Pure and Applied Geophysics*, 172(5), 1143–1165.
- Rolandone, F., Bürgmann, R., & Nadeau, R. M. (2004). The evolution of the seismic-aseismic transition during the earthquake cycle: Constraints from the time-dependent depth distribution of aftershocks. *Geophysical Research Letters*, 31, L23610. <https://doi.org/10.1029/2004GL021379>
- Ross, Z. E., & Ben-Zion, Y. (2014). Automatic picking of direct P, S seismic phases and fault zone head waves. *Geophysical Journal International*, 199(1), 368–381. <https://doi.org/10.1093/gji/ggu267>
- Ross, Z. E., Ben-Zion, Y., White, M. C., & Vernon, F. L. (2016). Analysis of earthquake body wave spectra for potency and magnitude values: implications for magnitude scaling relations. *Geophysical Journal International*, 207(2), 1158–1164.
- Ross, Z. E., Hauksson, E., & Ben-Zion, Y. (2017). Abundant off-fault seismicity and orthogonal structures in the San Jacinto fault zone. *Science Advances*, 3(3), e1601946.
- Ross, Z. E., Trugman, D. T., Hauksson, E., & Shearer, P. M. (2019). Searching for hidden earthquakes in Southern California. *Science*, 364, 767–771.
- Ross, Z. E., White, M. C., Vernon, F. L., & Ben-Zion, Y. (2016). An improved algorithm for real-time S-wave picking with application to the (augmented) ANZA network in Southern California. *Bulletin of the Seismological Society of America*, 106(5), 2013–2022. <https://doi.org/10.1785/0120150230>
- Rubin, A. M., & Gillard, D. (2004). Aftershock asymmetry/rupture directivity among central San Andreas fault microearthquakes. *Journal of Geophysical Research*, 105(B8), 19,095–19,109. <https://doi.org/10.1029/2000jb900129>
- Sanders, C., & Kanamori, H. (1984). A seismotectonic analysis of the Anza seismic gap, San Jacinto fault zone, southern California. *Journal of Geophysical Research*, 89(B7), 5873–5890.
- Schaff, D. P., Bokelmann, G. H. R., Beroza, G. C., Waldhauser, F., & Ellsworth, W. L. (2002). High-resolution image of Calaveras Fault seismicity. *Journal of Geophysical Research*, 107(B9), 2186. <https://doi.org/10.1029/2001jb000633>
- Scholz, C. H. (2002). *The mechanics of earthquakes and faulting* (2nd ed.), pp. 471. New York: Cambridge University Press.
- Share, P.-E., Ben-Zion, Y., Ross, Z. E., Qiu, H., & Vernon, F. L. (2017). Internal structure of the San Jacinto fault zone at Blackburn Saddle from seismic data of a linear array. *Geophysical Journal International*, 210(2), 819–832.
- Share, P.-E., Guo, H., Thurber, C. H., Zhang, H., & Ben-Zion, Y. (2019). Seismic imaging of the Southern California plate boundary around the south-central transverse ranges using double-difference tomography. *Pure and Applied Geophysics*, 176, 1117–1144.
- Sharp, R. V. (1967). San Jacinto fault zone in the Peninsular Ranges of Southern California. *Geological Society of America*, 78(6), 705–730.
- Shearer, P., Hauksson, E., & Lin, G. (2005). Southern California hypocenter relocation with waveform cross-correlation, part 2: Results using source-specific station terms and cluster analysis. *Bulletin of the Seismological Society of America*, 95(3), 904–915.
- Shelly, D. R., Ellsworth, W. L., & Hill, D. P. (2016). Fluid-faulting evolution in high definition: Connecting fault structure and frequency-magnitude variations during the 2014 Long Valley Caldera, California, earthquake swarm. *Journal of Geophysical Research: Solid Earth*, 121, 1776–1795. <https://doi.org/10.1002/2015JB012719>
- Sibson, R. H. (1982). Fault zone models, heat flow, and the depth distribution of earthquakes in the continental crust of the United States. *Bulletin of the Seismological Society of America*, 72(1), 151–163.
- Sibson, R. H. (1986). Earthquakes and rock deformation in crustal fault zones. *Annual Review of Earth and Planetary Sciences*, 14, 149–175.
- Small, P., Gill, D., Maechling, P. J., Taborda, R., Callaghan, S., Jordan, T. H., et al. (2017). The SCEC Unified Community Velocity Model Software Framework. *Seismological Research Letters*, 88(6), 1539–1552.
- Southern California Earthquake Center (2013). SCECD. Southern California Earthquake Data Center.
- Trugman, D. T., & Shearer, P. M. (2017). GrowClust: A hierarchical clustering algorithm for relative earthquake relocation, with application to the Spanish Springs and Sheldon, Nevada, earthquake sequences. *Seismological Research Letters*, 88(2A), 379–391.
- Vernon, F. L. (1982). ANZA Regional Network. San Diego: International Federation of Digital Seismograph Networks.

- Vernon, F. L., & Ben-Zion, Y. (2010). San Jacinto Fault Zone Experiment Network. International Federation of Digital Seismograph Networks.
- Vidale, J. E., & Shearer, P. M. (2006). A survey of 71 earthquake bursts across southern California: Exploring the role of pore fluid pressure fluctuations and aseismic slip as drivers. *Journal of Geophysical Research*, 111, B05312, <https://doi.org/10.1029/2005JB004034>
- Waldbauer, F., & Ellsworth, W. L. (2000). A Double-difference earthquake location algorithm: Method and application to the Northern Hayward Fault, California. *Bulletin of the Seismological Society of America*, 90(6), 1353–1368.
- Wdowinski, S. (2009). Deep creep as a cause for the excess seismicity along the San Jacinto Fault. *Nature Geoscience*, 2(12), 882–885.
- Zhu, L., & Kanamori, H. (2000). Moho depth variation in southern California from teleseismic receiver functions. *Journal of Geophysical Research*, 105(B2), 2969–2980.