

Informative Modes of Seismicity in Nearest-Neighbor Earthquake Proximities

 Yu-Fang Hsu¹ , Ilya Zaliapin^{2†}, and Yehuda Ben-Zion^{1,3} 

¹Department of Earth Sciences, University of Southern California, Los Angeles, CA, USA, ²Department of Mathematics and Statistics, University of Nevada, Reno, NV, USA, ³Statewide California Earthquake Center, University of Southern California, Los Angeles, CA, USA

†Deceased

Key Points:

- Nearest-neighbor diagrams provide an efficient way to identify different types of earthquake patterns
- Repeating events, large versus small-moderate mainshocks, and induced seismicity have characteristic signatures
- Nearest-neighbor diagrams help to evaluate features in catalogs constructed by different methods

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

Y.-F. Hsu,
yufanghs@usc.edu

Citation:

Hsu, Y.-F., Zaliapin, I., & Ben-Zion, Y. (2024). Informative modes of seismicity in nearest-neighbor earthquake proximities. *Journal of Geophysical Research: Solid Earth*, 129, e2023JB027826. <https://doi.org/10.1029/2023JB027826>

Received 9 SEP 2023
Accepted 8 MAR 2024

Author Contributions:

Formal analysis: Yu-Fang Hsu,

Ilya Zaliapin, Yehuda Ben-Zion

Investigation: Yu-Fang Hsu,

Ilya Zaliapin, Yehuda Ben-Zion

Methodology: Yu-Fang Hsu,

Ilya Zaliapin, Yehuda Ben-Zion

Project administration: Yehuda Ben-Zion

Resources: Ilya Zaliapin, Yehuda Ben-Zion

Software: Ilya Zaliapin

Supervision: Ilya Zaliapin, Yehuda Ben-Zion

Writing – original draft: Yu-Fang Hsu, Yehuda Ben-Zion

Abstract We analyze nearest-neighbor proximities of earthquakes in California based on the joint distribution (T, R) of rescaled time T and rescaled distance R between pairs of earthquakes (Zaliapin & Ben-Zion, 2013a, <https://doi.org/10.1002/jgrb.50179>), using seismic catalogs from several regions and several catalogs for the San Jacinto Fault Zone (SJFZ). The study aims to identify informative modes in nearest-neighbor diagrams beyond the general background and clustered modes, and to assess seismic catalogs derived by different methods. The results show that earthquake clusters with large and small-to-medium mainshocks have approximately diagonal and horizontal (T, R) distributions of the clustered mode, respectively, reflecting different triggering distances of mainshocks. Earthquakes in the creeping section of San Andreas Fault have a distinct “repeaters mode” characterized by very large rescaled times T and very small rescaled distances R , due to nearly identical locations of repeating events. Induced seismicity in the Geysers and Coso geothermal fields follow mostly the background mode, but with larger rescaled times T and smaller rescaled distances R compared to tectonic background seismicity. We also document differences in (T, R) distributions of catalogs constructed by different techniques (analyst-picks, template-matching and deep-learning) for the SJFZ, and detect a mode with very large R and small T in the template-matching and deep-learning based catalogs. This mode may reflect dynamic triggering by passing waves and/or catalog artifacts.

Plain Language Summary Earthquake patterns in space and time provide important information about the properties of active faults, the surrounding crust, and different loading mechanisms. Nearest-neighbor diagrams of earthquake proximities in a combined space-time-magnitude domain can be used to detect different types of seismicity patterns. Using such diagrams to analyze several earthquake catalogs in California, we examine characteristics of different sections of the San Andreas and San Jacinto faults along with geothermal areas and the entire southern California. The results highlight features of seismicity beyond the well-known background and clustered modes of earthquakes, including a mode of repeating events in the creeping section of the San Andreas fault and differences between events in regular tectonic areas and geothermal fields. Analyses of different catalogs for the region around the San Jacinto fault also reveal differences related to the different techniques used to construct the catalogs. The results show that nearest-neighbor diagrams can provide a simple and efficient way to identify different types of seismicity and assess the quality of earthquake catalogs.

1. Introduction

Earthquakes are fundamentally clustered in space and time (e.g., Omori, 1895; Richter, 1958; Utsu, 2002; Zaliapin & Ben-Zion, 2022). The root cause for seismic clustering is that earthquakes tend to trigger additional events with an intensity that increases with the earthquake size and decays with distance and time from the event (e.g., Felzer et al., 2004; Helmstetter et al., 2005; Ogata, 1999). Earthquake clusters have different forms including bursts, swarms, and other types of event sequences with internal properties that depend on various conditions (e.g., Ben-Zion, 2008; Shearer, 2012). These include the temperature and fluid content, which control together with the rock type the effective viscosity of the deforming medium, and the stress field (e.g., Ben-Zion & Lyakhovskiy, 2006; Enescu et al., 2009; Kisslinger & Jones, 1991; McGuire et al., 2005).

Zaliapin and Ben-Zion (2013a, 2013b) used the nearest-neighbor analysis described in Section 3.1 to identify and classify earthquake clusters in southern California. The nearest-neighbor analysis estimates proximities between pairs of earthquakes in a combined space-time-magnitude domain, and identifies a “parent” to each earthquake as the nearest-neighbor earlier event. The joint distribution of rescaled time (T) and rescaled space (R) of the nearest-

© 2024. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

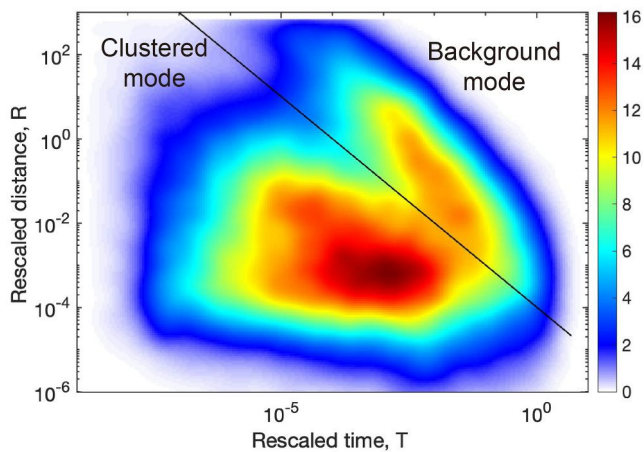


Figure 1. The joint distribution of rescaled time (T) and space (R) components of the nearest-neighbor proximity (Equation 2) of $M \geq 1.7$ events from 1981 to 2022 in the HYS catalog in southern California. The black line is $\log T + \log R = -3.99$.

neighbor proximities (Equations 1 and 2 in Section 3.1) follow a bimodal distribution with two main modes corresponding to background and clustered events. The background seismicity follows the time-stationary space-inhomogeneous Poisson process, while earthquake clusters have shorter (rescaled) space-time distances (Baiesi & Paczuski, 2004; Zaliapin et al., 2008). This is illustrated in Figure 1 using seismicity from 1981 to 2022 in the southern California earthquake catalog of Hauksson et al. (2012, extended to later years).

In this study, we use nearest-neighbor earthquake proximities in California to identify additional informative modes of seismicity in the joint distribution (T, R), beyond the main background and clustered modes. Toward this goal, we analyze earthquake catalogs covering regions dominated by different processes. These include the Geysers and Coso geothermal fields, the creeping section of the San Andreas Fault (SAF), and the San Jacinto Fault Zone (SJFZ). Another goal of this study is to perform a comparative analysis of earthquake catalogs produced by different methods to examine the impact of the different methods on the clustering properties of the events. This is done by analyzing nearest-neighbor earthquake proximities in three different catalogs for the SJFZ: the Hauksson et al. (2012), the related template-based

Ross et al. (2019), and the machine-learning-based White et al. (2021a, 2021b) catalogs. The nearest-neighbor results provide useful information on statistical features of events in, and the quality of, the catalogs. We note that the total number of events in a catalog is not a sufficient criterion for the catalog quality. This is because multiple environmental and anthropogenic sources can lead to many false small events (e.g., Johnson et al., 2019; Sheng et al., 2023), and template-based detections of smaller earthquakes increase the number of events but may bias the resolved structure of the seismicity.

In the following sections, we first describe the different regions and seismic catalogs examined in this study. We then describe the nearest-neighbor methodology and the obtained results. The analysis of data in the creeping section of the SAF reveals a distinct mode of seismicity indicative of repeating earthquakes. The results also highlight differences between clusters of earthquakes with different mainshock sizes, between seismicity in regions dominated by induced versus regular tectonic events, and between different earthquake catalogs for the SJFZ. The results are discussed and summarized in the final section of the paper.

2. Regions and Data

2.1. Research Areas and Catalogs

We examine five areas in California with different types of seismicity (Figure 2), focusing on details of the nearest-neighbor analysis results in the different regions. Starting at the north, the Geysers Geothermal field (GyGF) is dominated by induced seismicity. The creeping section of the SAF, including the north transition zone, the fast-creeping central zone, the region around Parkfield, and the south transition zone, is characterized by localized seismicity with repeating earthquakes. The Coso Geothermal field (CoGF) is dominated by induced seismicity. Southern California and the SJFZ have complex patterns of tectonic earthquakes.

The analysis employs high-quality seismic catalogs for each area. For the SJFZ we use three alternative high-quality catalogs produced by complementary techniques. The availability of several high-quality catalogs for the SJFZ allows us to perform comparative analysis not only between different regions but also between different catalogs for the same region. Table 1 summarizes the key attributes of the examined catalogs. In all cases, only earthquakes with M above the magnitude of completeness M_c are analyzed. The M_c values are taken from published results when available. For the creeping section of the SAF, the GyGF, and southern California, the values of M_c are estimated by the best combination of maximum curvature, 90% probability of M_c , and 95% probability of M_c (Wiemer & Wyss, 2000).

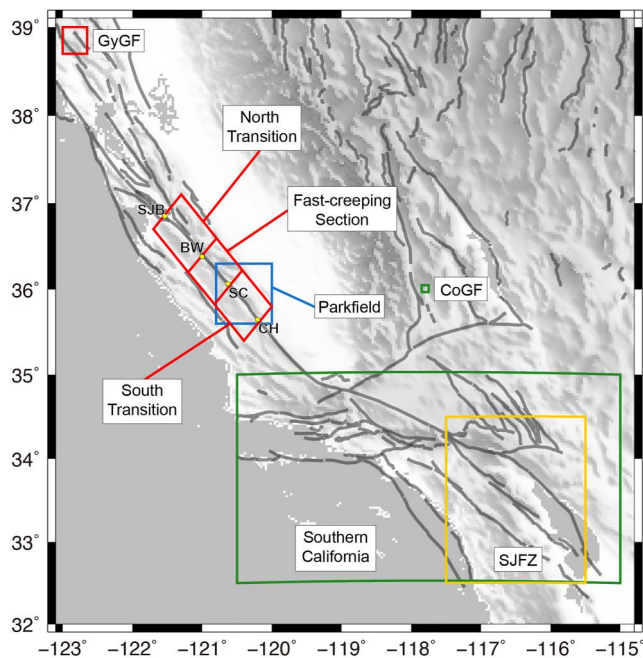


Figure 2. A map of different regions examined in this study. The analyses in the red, blue, and green boxes are based on the Waldhauser and Schaff (2008) catalog (WS), the Schaff and Waldhauser (2005) catalog (SW), and the Hauksson et al. (2012, extended) catalog (HYS), respectively. The analyses in the yellow region around the San Jacinto Fault Zone use the HYS catalog along with the Ross et al. (2019) catalog (RTHS), and the White et al. (2021a, 2021b) catalog (WBV). The gray lines mark major faults. SJB, BW, SC, and CH denote San Juan Bautista, Bitterwater, Slack Canyon, and Cholame, respectively. Labels on white background refer to names used in the text.

2.1.1. The Geysers Geothermal Field (GyGF)

The GyGF in northern California was constructed in the 1960s to produce geothermal energy and is dominated by human-induced seismicity due to production activities (Majer & Peterson, 2007; Martínez-Garzón et al., 2014). Induced seismicity prominently developed in the area after 1979 (Eberhart-Phillips & Oppenheimer, 1984). In 1997 and 2003, the GyGF conducted two major wastewater injection projects (Gunasekera et al., 2003; Majer & Peterson, 2007). The increasing fluid injection rate resulted in a significantly increased level of earthquake activities in the northwest part of the GyGF (Kwiatek et al., 2015; Martínez-Garzón et al., 2014). Most induced seismicity occurs in the northwest of the GyGF, and the earthquake magnitudes can be as high as 3.4, possibly connected to the large injection volume in the reservoir (Kwiatek et al., 2015).

For the GyGF, we analyze data from the Waldhauser and Schaff (2008) relocated catalog (WS) because this catalog has the lowest location errors in northern California. The WS catalog uses arrival time detections by analysts from the Northern California Seismic Network since 1984. The resolution in hypocenter locations is improved with waveform cross-correlation (CC) and double-difference (DD) methods (Waldhauser, 2001; Waldhauser & Ellsworth, 2000). The uncertainties of the location errors at the 95% confidence level are 0.7 km horizontally and 2 km vertically. The WS catalog in the GyGF has $M_c = 1$ (Zaliapin & Ben-Zion, 2016a).

2.1.2. The Creeping Section of the San Andreas Fault

The creeping section of the SAF includes remnants of subduction zone rocks with active dehydration (Fagereng & Diener, 2011; Irwin & Barnes, 1975; Kirby et al., 2014). It extends generally from northwest of San Juan Bautista to Cholame (Burford & Harsh, 1980). Accompanied by aseismic creeping motion, the seismicity in this region includes repeating earthquakes (repeaters) that occur in nearly identical locations and have nearly identical

waveforms and magnitudes (Nadeau & Johnson, 1998; Nadeau et al., 1995). The time intervals between pairs of repeating earthquakes are correlated with the seismic moment of the earlier earthquake in the pair (Chen et al., 2007; Nadeau & Johnson, 1998; Vidale et al., 1994). As shown in Figure 2, we analyze the entire creeping section as well as several subregions.

The north transition zone extends from northwest of San Juan Bautista to Bitterwater and has a transition from a locked fault to creeping motion. The central zone extends from Bitterwater to Slack Canyon and has the highest

Table 1
Earthquake Catalogs Used in the Study

Regions	Catalogs	Methods	Time span	Magnitude range	No. events
GyGF	WS (Waldhauser & Schaff, 2008)	Analyst-picking	1984–2011	1.0–4.5	77,421
North Transition	WS (Waldhauser & Schaff, 2008)	Analyst-picking	1984–2011	1.0–5.5	31,060
Fast-creeping section	WS (Waldhauser & Schaff, 2008)	Analyst-picking	1984–2011	1.0–4.5	5,333
South Transition	WS (Waldhauser & Schaff, 2008)	Analyst-picking	1984–2011	1.0–6.0	7,900
Parkfield	SW (Schaff & Waldhauser, 2005)	Analyst-picking	1984–2005	1.0–6.0	6,441
CoGF	HYS (Hauksson et al., 2012, extend)	Analyst-picking	1981–2022	1.0–4.41	5,300
Southern California	HYS (Hauksson et al., 2012, extend)	Analyst-picking	1981–2022	1.7–7.3	145,531
SJFZ	HYS (Hauksson et al., 2012, extend)	Analyst-picking	2008–2017	1.0–5.71	63,756
	RTHS (Ross et al., 2019)	Template matching		1.0–5.72	113,942
	WBV (White et al., 2021a, 2021b)	Deep learning		1.0–5.54	54,980

creep rate (Burford & Harsh, 1980) and the highest non-clustered seismicity rate in the creeping section (Liu et al., 2022). The south transition extends from Slack Canyon to Cholame, including Parkfield, and has a transition from creeping motion to a locked fault. For the creeping section, other than the Parkfield area, we use the WS catalog for analysis with an estimated $M_c = 1$.

The Parkfield area is characterized by a large proportion of repeating earthquakes (Nadeau et al., 1995), many of which occur along lines reflecting transitions from creeping fault patches to locked regions (Nadeau & Johnson, 1998; Nadeau & McEvilly, 1999, 2004). This area also had quasi-periodic moderate earthquakes of magnitude $M \sim 6.0$ over the last 100 years or so, the most recent of which was in 2004. For analysis in the Parkfield area, we use the Schaff and Waldhauser (2005) relocated catalog (SW), created with a similar methodology to the WS catalog, because it has higher resolution event locations. More than 40% of the earthquakes in the SW catalog are repeating events with a CC value >0.9 , and the analyzed earthquakes have $M_c = 1$.

2.1.3. The Coso Geothermal Field (CoGF)

The CoGF has both natural and human-induced seismicity (Feng & Lees, 1998; Hauksson & Unruh, 2007). The tectonic loading and magmatic source cause natural earthquakes, and geothermal power production triggers induced seismicity (Feng & Lees, 1998; Manley & Bacon, 2000; Schoenball et al., 2015). Geothermal power production in the CoGF started in 1977 and reached its peak in the early 1990s (Monastero, 2002; Schoenball et al., 2015). The significant increase in injection rate strongly increased the seismicity rate in the 1990s and changed the earthquake spatial distribution (e.g., Zaliapin & Ben-Zion, 2016a). After the 1990s, injection rates decreased with time and the seismicity rate slightly declined (Schoenball et al., 2015). The northwest CoGF has the highest seismicity rate and the events caused by injection in CoGF can have magnitudes as high as $M = 4.1$ (Schoenball et al., 2015).

In the CoGF, we use the Hauksson et al. (2012, extended) relocated catalog (HYS) for analysis. The HYS catalog uses arrival time picks by analysts from the Southern California Seismic Network (SCSN) (Caltech, 1926) since 1981. The resolution in hypocenter locations is improved with waveform CC and DD methods based on 1D and 3D velocity models. The catalog construction includes application of the GrowClust method (Trugman & Shearer, 2017) to improve locations of nearby events using differential travel times from waveform cross-correlations. Events that are successfully relocated have at least 5 differential times measured for one or more event pairs. The magnitude of completeness in the CoGF has M_c values that fluctuate around 1 (Zaliapin & Ben-Zion, 2016a, and Figure S1 in Supporting Information S1).

2.1.4. Southern California

The region between latitude and longitude degrees 32.5° to 35° and -120.5° to -115° is referred to here as southern California. The fault system and seismicity patterns in this region are considerably broader and more complex than in central and northern California (Ross et al., 2022). The earthquakes are dominated by tectonic events associated primarily with the subparallel strike-slip faults of the San Andreas system, dipping faults in the transverse range, and strike-slip faults in the Eastern California Shear Zone (Ben-Zion & Zaliapin, 2020; Hauksson et al., 2012). The HYS catalog is used for analysis in southern California and the magnitude of completeness is about $M_c = 1.7$.

2.1.5. San Jacinto Fault Zone

The SJFZ is the most seismically active branch of the San Andreas system in southern California (Ross et al., 2017) and it accommodates 10–15 mm/yr of the plate motion (Blisniuk et al., 2013; Fialko, 2006). The SJFZ includes several highly active fault strands that produce complex patterns of seismicity (Sanders & Magistrale, 1997) and it had over 15 $M > 7$ earthquakes over the past 4,000 years (Rockwell et al., 2015). The largest magnitude earthquake in the SJFZ region for the examined period is 5.71. For analysis in the SJFZ, we use the HYS catalog, the Ross et al. (2019) catalog (RTHS) based on template-matching, and the White et al. (2021a, 2021b) catalog (WBV) produced by deep-learning. The magnitude of completeness of the HYS catalog in the SJFZ is about $M_c = 1$ (Zaliapin & Ben-Zion, 2016a). To compare these three catalogs, we use $M \geq 1$ events for analysis.

The RTHS catalog uses data of previously detected earthquakes in the HYS catalog as templates and detects new earthquakes based on the similarity between templates and 24-hr continuous waveforms. The catalog includes new events with median absolute deviation (MAD) of cross-correlated waveforms from all stations, channels, and phases above different thresholds. The RTHS catalog also relocates the newly detected events with the Growclust DD method to better constrain the hypocenters. In this study, we use events from the RTHS catalog with MAD of cross-correlated waveforms above 9.5. The magnitude of completeness of such events in the RTHS catalog is $M_c = 0.3$ (Ross et al., 2019).

The WBV catalog is generated using deep-learning-based phase picking, association, and relocation (White et al., 2021a, 2021b). Specifically, the EQTransformer algorithm (Mousavi et al., 2020) is first used to detect P- and S-wave earthquake arrival times from the SCSN and additional stations around the SJFZ during 2008–2017. Second, the Rapid Earthquake Association and Location algorithm (REAL; Zhang et al., 2019) is used to associate the detected arrival times with earthquake sources. The travel times are traced back using 1-D and 3-D velocity models to estimate the origin arrival time and earthquake location. Last, the Growclust method is used to relocate the earthquakes and the MCMC sampling procedure is used to estimate the uncertainties of the earthquake locations. Events that cannot be relocated by the Growclust DD method are subject to a quality control test. As in White et al. (2019), an event needs to have a location error of less than 3 km, more than three S-wave arrival detections, and at least 25% S-to-P phase pick ratio to remain in the catalog. The magnitude of completeness of the WBV catalog is $M_c = 0.93$ (White et al., 2021a, 2021b).

3. Methods

3.1. Nearest-Neighbor Analysis

The nearest-neighbor technique has been applied to multiple problems of statistical seismology including detection and classification of earthquake clusters, catalog declustering, and characterization of induced seismicity (e.g., Gentili et al., 2019; Karimi & Davidsen, 2023; Martínez-Garzón et al., 2018; Ruhl et al., 2016; Schoenball et al., 2015; Zaliapin & Ben-Zion, 2013a, 2013b, 2016b). The core of the method is a statistical identification of an event (the nearest-neighbor) that has the largest probability of triggering (closest proximity) to a given event in the examined catalog. Formally, to identify the nearest neighbor of event j , we calculate the earthquake proximity η_{ij} between event j and every earlier event i using

$$\eta_{ij} = \begin{cases} t_{ij} r_{ij}^d 10^{-bm_i} & \text{if } t_{ij} > 0 \\ \infty & \text{if } t_{ij} \leq 0 \end{cases} \quad (1)$$

where $t_{ij} = t_j - t_i$ is the temporal separation [in years] between event j and event i , r_{ij} is the (2D or 3D) space distance [in km] between event j and event i , d is the fractal dimension of epicenters or hypocenters, b is the b -value of the Gutenberg-Richter statistics in the examined region, and m_i is the magnitude of event i . In this study, we perform 3D analysis with hypocenters using $d = 2.6$ and $b = 1$ (Sammis et al., 1987; Zaliapin & Ben-Zion, 2013a). For each event j (offspring), its nearest-neighbor i (parent) is the earthquake that minimizes the proximity measure given by Equation 1.

The nearest neighbor proximity η_{ij} can be represented as a product of its rescaled space (R_{ij}) and time (T_{ij}) components:

$$\begin{aligned} \eta_{ij} &= T_{ij} R_{ij} \\ T_{ij} &= t_{ij} 10^{-pbm_i} ; R_{ij} = r_{ij}^d 10^{-qbm_i} \\ p + q &= 1 \end{aligned} \quad (2)$$

Following Zaliapin and Ben-Zion (2013a, 2013b, 2016b), we use generally $p = 0.5$ and $q = 0.5$. With these choices, the R_{ij} and T_{ij} are the spatial distance and time separation between event pair (i, j) normalized by a factor proportional to the rupture length of the parent i . The joint distribution of (T, R) is calculated using logarithmic binning, and in regions dominated by tectonic seismicity, it is generally bimodal, consisting of background and clustered modes. This also holds with some modifications discussed further below for induced earthquakes.

As illustrated in Figure 1, the mode above and following the diagonal line $\log T + \log R = \text{const.}$ corresponds to background events with nearest-neighbor as in a stationary homogeneous Poisson process. The other mode below and to the left of the diagonal line represents different types of clustered events having shorter proximities than expectations based on random occurrence. Connecting each event to its parent produces a time-oriented spanning tree (Zaliapin & Ben-Zion, 2013a). To separate background seismicity from earthquake clusters, we remove the connections between parents and offspring with $\eta_{ij} > \eta_0$. The threshold η_0 is determined by a Gaussian mixture model with the 1-D $\log \eta$ -distribution of background and clustered modes (Zaliapin & Ben-Zion, 2016b). We use the expectation–maximization algorithm to determine the maximum likelihood boundary η_0 between these two modes (Figure S2 in Supporting Information S1). This partitions the spanning tree into a spanning forest (a collection of trees), where each tree corresponds to an individual earthquake cluster. We note that the presence of repeating events or other unusual seismicity features in the examined data might require some modifications to the separation of background and clustered events, but for consistency with previous applications, we use the same procedure as in Zaliapin and Ben-Zion (2016b).

In each earthquake cluster, we further define the event with the largest magnitude as a mainshock. The events occurring before and after the mainshock are called foreshocks and aftershocks, respectively. The event type identification (background, clustered) depends on a specific application. One can identify the mainshocks as background events, and all other events as clustered. Alternatively, one can identify the first events in each cluster as background, and all other events as clustered (Gu et al., 2013; Zaliapin & Ben-Zion, 2020).

4. Results

4.1. Main Modes

As mentioned in Section 1, the joint distribution of rescaled time and rescaled space components (T , R) of the nearest-neighbor proximity in southern California (Figure 1) has two clear modes referred to as background and clustered seismicity. These two modes exist generally in other catalogs representing large space-time domains and in synthetic seismicity based on the Epidemic Type Aftershock Sequence model (e.g., Gu et al., 2013; Kossobokov & Nekrasova, 2017; Moradpour et al., 2014; Peresan & Gentili, 2018; Reverso et al., 2015; Zaliapin & Ben-Zion, 2013a, 2016b). The background mode has a diagonal distribution along and above the line $\log T + \log R = \text{const.}$, which represents events with rates following a Poisson process, random locations, and magnitudes taken from the Gutenberg–Richter distribution (Zaliapin & Ben-Zion, 2013a; Zaliapin et al., 2008). The clustered mode is below the diagonal line with smaller rescaled time T and rescaled distance R , which implies clustering in time and space. The clustered mode in Figure 1 is quasi-parallel to the rescaled time axis, indicating that the spatial clustering around the main faults in southern California is stronger than the temporal clustering (Zaliapin & Ben-Zion, 2022). In the following sections, we only consider clusters with at least two events and omit singles (clusters with one event).

4.2. Clusters With Different Mainshock Sizes in Southern California

The HYS catalog for southern California includes 11 earthquake clusters with $M > 5.5$ mainshocks and 9,037 clusters with $M \leq 5.5$ mainshocks. There are 55,261 and 56,624 events in the clusters with $M > 5.5$ and $M \leq 5.5$ mainshocks. Generally, the logarithm of the number of events in the clusters (productivity) is positively correlated with the mainshock magnitudes (Figure S3a in Supporting Information S1) as found in previous studies (e.g., Helmstetter et al., 2005; Trugman & Ben-Zion, 2023). A few clusters with relatively small mainshock magnitudes (between M_2 and M_3) have higher productivity compared to other clusters with similar mainshock magnitudes, which may be related to elevated surface heat flow, high fault misalignment, and/or other variables (e.g., Trugman & Ben-Zion, 2023). Each cluster has a single mainshock (the largest magnitude event), while the number of foreshocks and aftershocks varies between clusters. For the clusters with $M > 5.5$ mainshocks, 10.11% of the earthquakes are classified as foreshocks, 0.02% as mainshocks, and the remaining 89.87% as aftershocks. For the clusters with $M \leq 5.5$ mainshocks, 27.88% of the earthquakes are identified as foreshocks, 15.96% as mainshocks, and 56.16% as aftershocks.

The earthquake clusters with mainshock sizes of $M > 5.5$ and $M \leq 5.5$ in southern California occupy different parts in the joint distribution (T , R) of the nearest-neighbor proximity (Figure 3). The clusters with $M > 5.5$ mainshocks (Figure 3a) have essentially no background mode, given the negligible number of the first events in these 11 earthquake clusters, and a clustered mode following a diagonal shape reflecting the fact that larger

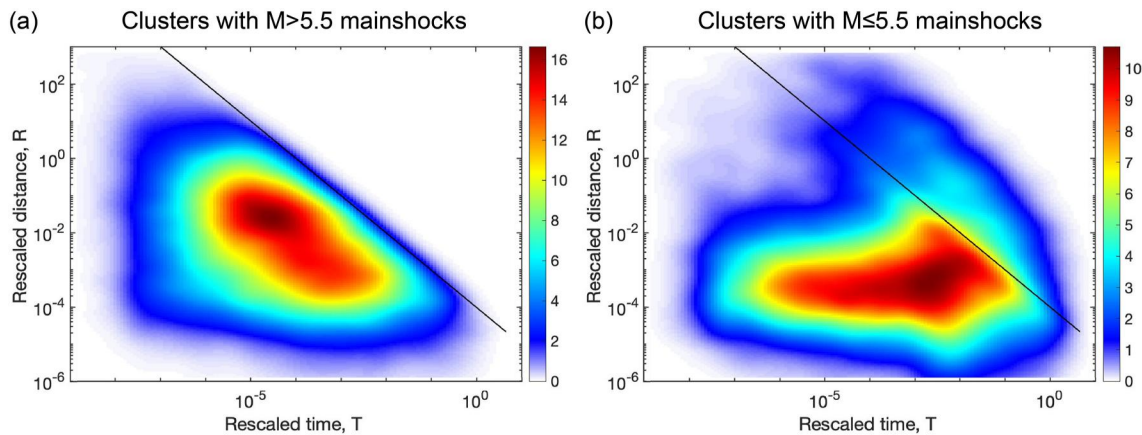


Figure 3. (a) The joint distribution of rescaled time (T) and space (R) components of the nearest-neighbor proximity (Equation 2) of earthquake clusters with $M > 5.5$ mainshocks in the HYS catalog in southern California (see Figure 2). (b) Same as (a) for earthquake clusters with $M \leq 5.5$ mainshocks in southern California. The black lines are $\log T + \log R = -3.9$.

mainshocks trigger events at larger distances. The range of R values in Figure 3a is relatively large and extends from 10^{-5} to 10^0 . On the other hand, earthquake clusters with $M \leq 5.5$ mainshocks (Figure 3b) have a visible background mode since there are 9,037 first events in these clusters, and a clustered mode following a horizontal distribution with R values below 10^{-2} . Adding the clustered modes in Figures 3a and 3b produces a combined clustered mode that is quasi-parallel to the T axis (Figure 1).

Location errors and short-term incompleteness after moderate-large earthquakes are known to affect nearest-neighbor and other statistical seismology results associated with small events (e.g., de Arcangelis et al., 2018; Mizrahi et al., 2021; Zaliapin & Ben-Zion, 2015). To check that the patterns of Figure 3 are not produced by small event artifacts, we repeat the analysis using only earthquakes with magnitudes larger than 2 and 3 (Figure S4 in Supporting Information S1). The discussed differences between clusters with $M > 5.5$ and $M \leq 5.5$ mainshocks remain, although they are displayed in more patchy forms as the minimum event size increases and the amount of data decreases.

The physical (non-rescaled) space distances between aftershocks and their parents in the clusters with $M > 5.5$ mainshocks has a distribution with two broad peaks around $10^{-0.8}$ (0.15) km and $10^{0.8}$ (6.5) km (Figure S5a in Supporting Information S1). The first peak is consistent with a characteristic length scale (half-width) of core damage zones of faults and near-fault seismicity (e.g., Ben-Zion & Sammis, 2003; Powers & Jordan, 2010), while the second peak is consistent with a typical length scale (mid-depth) of the seismogenic zone in Southern California (e.g., Hauksson et al., 2012; White et al., 2019). It is interesting to note that Moradpour et al. (2014) used the nearest-neighbor clustering method to analyze the decay of aftershocks density with distance from mainshocks. They found changes at length scales of a few hundred meters and about 10 km, somewhat larger than the peaks in Figure S5 in Supporting Information S1, and interpreted the first as reflecting finite rupture size effects on estimated distances and the second as reflecting the thickness of the brittle crust.

The duration of earthquake clusters generally increases with mainshock magnitudes with scatter reflecting different cluster types and productivity (Figure S3b in Supporting Information S1). Some clusters with relatively small mainshock magnitudes and long duration correspond to swarms that occur gradually and have many more foreshocks (Zaliapin & Ben-Zion, 2013b). The values of the rescaled time component T in the earthquake clusters with $M > 5.5$ and $M \leq 5.5$ mainshocks vary from 10^{-7} to 10^0 (Figure 3). The physical time separations between events and their parents in the earthquake clusters vary from 10^{-5} to 10^4 days (Figures S4b and S4d in Supporting Information S1) and the temporal distributions depend on the mainshock magnitude. In earthquake clusters with $M > 5.5$ mainshocks, most events are aftershocks and the time separations between foreshocks and their parents range from 10^{-5} (0.00001) to $10^{1.8}$ (63) days (Figure S5b in Supporting Information S1). In contrast, for clusters with $M \leq 5.5$ mainshocks, there are more foreshocks with time separations from parents that can be up to 10^4 days (Figure S5d in Supporting Information S1). This is associated with the fact that the clusters with $M > 5.5$ mainshocks are burst-type with relatively short duration and small number of foreshocks, while the clusters with

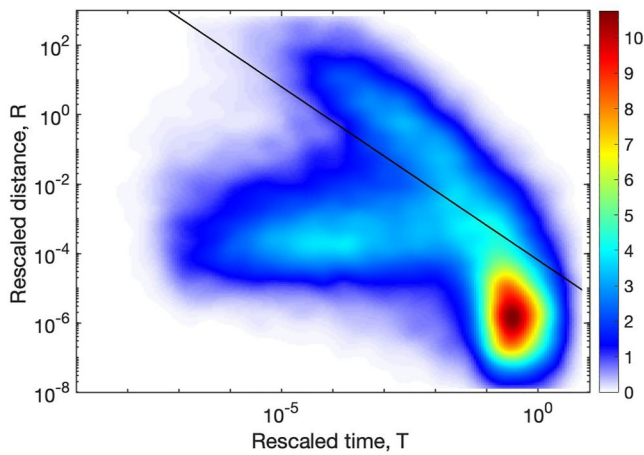


Figure 4. The joint distribution of rescaled time (T) and space (R) components of the nearest-neighbor proximity (Equation 2) of $M \geq 1$ events in the WS catalog in the creeping section of the San Andreas Fault. The black line is $\log T + \log R = -4.19$.

$M \leq 5.5$ mainshocks include swarms with relatively long durations and more foreshocks (Zaliapin & Ben-Zion, 2013b). Moreover, only earthquake clusters with $M \leq 5.5$ mainshocks include an appreciable background mode that enlarges the ranges of values in the joint distribution of (T , R). The distribution of the physical time and space distances between the events in the background mode and their parents is generally larger compared to that between events in the clustered mode (i.e., foreshocks and aftershocks) and their parents (Figures S5c and S5d in Supporting Information S1).

The small time separations between events in the examined catalogs are limited by short-term incompleteness (e.g., de Arcangelis et al., 2018) but this is not likely to significantly affect the results highlighted in this work (Zaliapin & Ben-Zion, 2015). The varying relative foreshock number with different mainshock magnitudes is consistent with the difference in the productivity coefficient for foreshocks and aftershocks in southern California (Petrillo & Lippiello, 2020).

4.3. Repeating Earthquakes

Figure 4 shows the joint distribution (T , R) of the nearest-neighbor proximities in the entire creeping zone. The results include the 2 main modes of the joint distribution associated with background and clustered seismicity, along with a strong third mode in the lower-right corner reflecting repeating earthquakes. The clustered mode in Figure 4 is horizontal because the mainshock magnitudes in this region are generally small. The lower-right mode of repeaters is characterized by a vertical distribution with large rescaled time T and small rescaled distances R . This mode is also shown clearly for earthquakes in the Parkfield region discussed in more detail below as well as for events in the north and south transition zones and fast creeping section (Figure S6 in Supporting Information S1).

Figure 5 presents the joint distribution (T , R) for all $M > 1$ events in the Parkfield region (Figure 5a), along with results not including the cluster associated with the 2004 M6 Parkfield event (Figure 5b). The clustered model is considerably weaker in Figure 5b without the cluster of the M6 earthquake, but the vertical mode reflecting repeaters remains strong. The R values in this vertically distributed mode are concentrated below 10^{-5} , and the physical distances between parents and offspring in this mode (Figure S7 in Supporting Information S1) are very short (< 0.03 km). The rescaled time T of this vertical mode covers a narrow range (10^{-1} – 10^0) of values, but the physical time between parents and offspring can be up to 18 years. Since the vertical mode of the repeaters does not follow a diagonal line, the occurrence times between parents and offspring in this mode are not Poissonian. The magnitude differences between parents and offspring in the vertical mode are mostly less than 0.5 (Figure S8

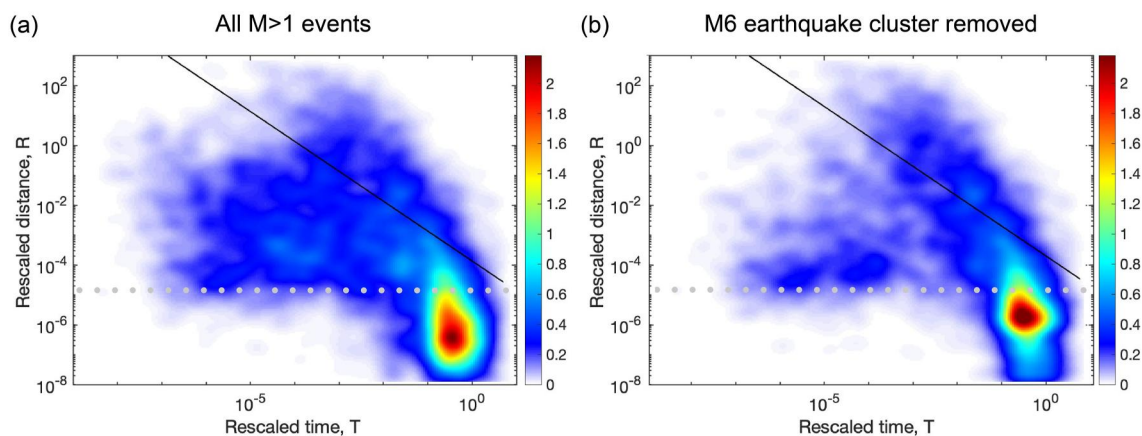


Figure 5. The joint distribution of rescaled time (T) and space (R) components of the nearest-neighbor proximity (Equation 2) of $M \geq 1$ events in the SW catalog for the Parkfield area. The horizontal line $R = 10^{-5}$ separate between the clustered and repeating earthquake modes. The most intensive mode (red colors) corresponds to repeating earthquakes. (a) The whole SW catalog in the Parkfield area (see Figure 2). The black lines are $\log T + \log R = -3.9$. (b) Same as (a) with the 2004 M6 Parkfield aftershock sequence removed. The black lines are $\log T + \log R = -3.7$.

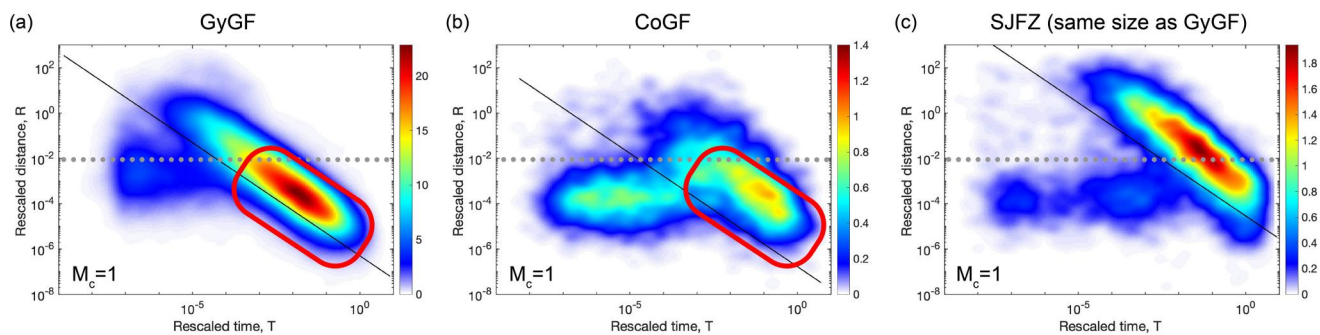


Figure 6. (a) The joint distribution of rescaled time (T) and space (R) components of the nearest-neighbor proximity (Equation 2) of $M \geq 1$ events in the WS catalog for the Geysers Geothermal field (GyGF) (see Figure 2). The black line is $\log T + \log R = -6.3$. (b) Same as (a) in the HYS catalog for the Coso Geothermal field (see Figure 2). The black line is $\log T + \log R = -6.8$. (c) Same as (b) for the San Jacinto Fault Zone (SJFZ) with the same size of the spatial extent of the GyGF (Note that the region around the SJFZ in panel (c) is smaller than the SJFZ region used for the later results shown in Figures 7–9.). The black line is $\log T + \log R = -4.5$. The horizontal line $R = 10^{-2}$ highlights the difference between the background seismicity in the red circles in the geothermal fields (a) and (b) and the background seismicity in a SJFZ region of comparable size to the GyGF (c). The magnitude of completeness is indicated in the lower left of each panel.

in Supporting Information S1). The highly close locations and similar magnitudes justify the association of this mode with repeating earthquakes.

4.4. Induced Seismicity

The joint distribution (T, R) for the GyGF (Figure 6a) and CoGF (Figure 6b) exhibit a background mode of seismicity following the diagonal line, but with smaller R and less events with small T compared to the background seismicity in the SJFZ (Figure 6c) with a similar spatial extent as the GyGF. The values of T and R are concentrated above 10^{-4} and below 10^{-2} , respectively. As in other cases with small mainshock sizes, the clustered mode in the GyGF and CoGF is horizontal. The ratio of event intensity in the clustered versus the background mode in the GyGF (Figure 6a) is lower than in the CoGF (Figure 6b). This may stem from the fact that the CoGF has a mixture of induced and tectonic events. For the events encircled in the red in Figure 6, the physical space separations between parents and offspring are below 1.5 km while the physical time separations can be above 25 years (Figures S9a and S9b in Supporting Information S1). In contrast, the physical distances between parents and offspring in the SJFZ can be up to 35 km (Figure S9c in Supporting Information S1). The relatively small distances between the events in the GyGF and CoGF imply smaller regions with high stresses (more concentrated loadings) in the geothermal fields relative to the tectonically active SJFZ.

4.5. Comparisons of Catalogs in the SJFZ

The joint distributions of (T, R) in the HYS and RTHS catalogs for the SJFZ have more intensive clustered mode than the background mode (Figures 7a and 7b). The RTHS catalog, which adds template-based detections to the HYS catalog, exhibits a stronger/weaker clustered/background mode than the HYS catalog. The WBV catalog has a considerably larger ratio of background/clustered modes than the HYS and RTHS catalogs (Figure 7c). These differences may result from the larger number of small, non-template-based, events in the WBV catalog, and possibly also less accurate event locations and other differences in the catalog construction procedures.

The RTHS and WBV catalogs also have a distinct mode in the upper-left corner of the (T, R) distributions marked by red boxes in Figures 7b and 7c, with very small rescaled time T and very large rescaled space R . In the HYS catalog, there are only 260 events (0.4%) in the location of the upper-left mode. In the RTHS and WBV catalogs, there are 3,563 events (3.1%) and 1,210 events (2.2%) in the upper-left mode. The upper-left mode remains similar for ranges of p and q values of Equation 2 between 0.3 and 0.7 (Figure S10 in Supporting Information S1). The results in the upper-left mode may reflect artifacts produced by the catalog construction process. Alternatively, this mode of (T, R) values may reflect triggering of events (e.g., by seismic waves from remote earthquakes) with much larger distance within a very short time interval compared to the clustered events.

In the RTHS catalog, most of the events in the upper-left mode occur around the time of the 2010 $M7.2$ El Mayor earthquake (Figure 8a). The spatial distribution of the events in the upper-left mode is generally consistent with the main faults in the study area (Figure 8b). The physical space distance and time separation between these events

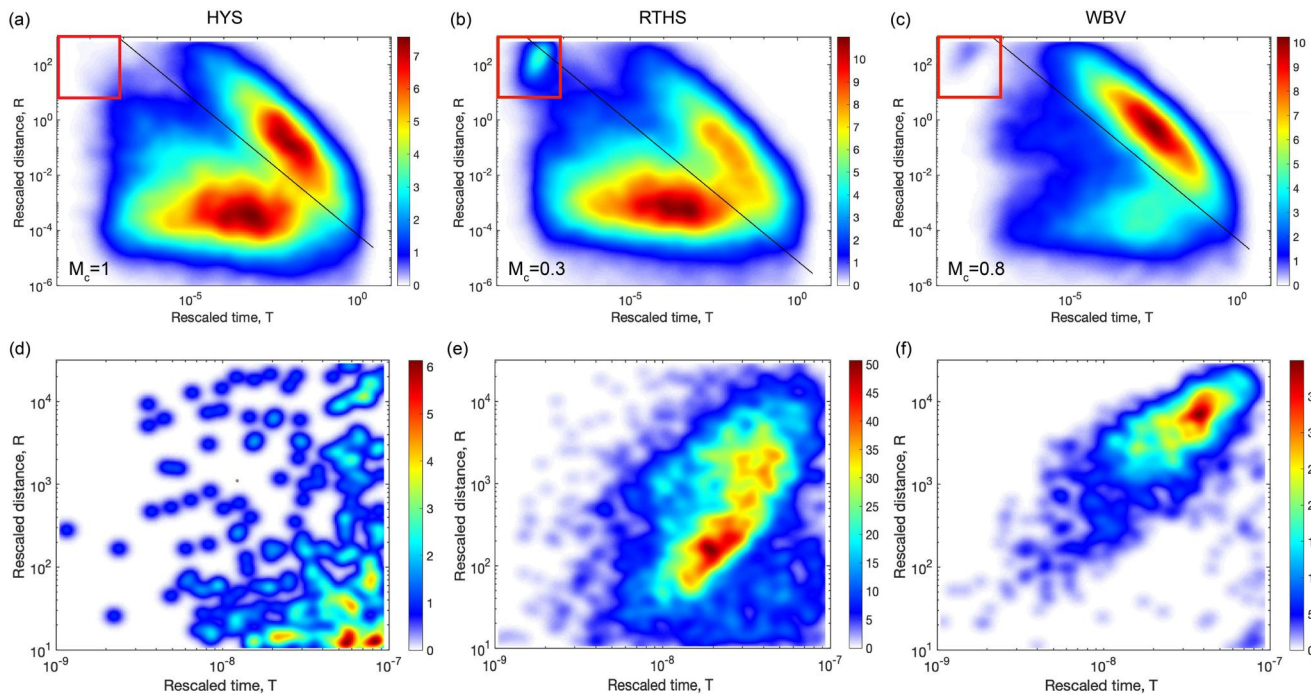


Figure 7. The joint distribution of rescaled time (T) and space (R) components of the nearest-neighbor proximity of $M \geq 1$ events in the San Jacinto Fault Zone (see Figure 2). (a) The HYS catalog with a black line following $\log T + \log R = -4.2$. (b) The RTHS catalog with a black line following $\log T + \log R = -5.1$. (c) The WBV catalog with a black line following $\log T + \log R = -4.3$. The red boxes in panels (b) and (c) denote the upper-left modes discussed in the text. (d) The density distribution of rescaled time (T) and space (R) components of the nearest-neighbor proximity of $M \geq 1$ events in the red box in panel (a). (e) Same as (d) in the red box in panel (b). (f) Same as (d) in the red box in (c). Note that the density distributions in panels (d)–(f) are not normalized.

and their parents can be up to 150 km within 40 s. The space-time distribution of many events in the upper-left mode follows a line with a slope of 3–4 km/sec (Figure 8c), which corresponds generally to the average S-wave velocity in the area (Allam & Ben-Zion, 2012). This suggests that the events in the upper-left mode in the RTHS catalog may be related to propagation of S-waves. Moreover, as indicated by the survival function $S(x) = 1 - F(x) = \text{Prob}(\text{offspring number of parents} > x)$, the parents of events in the upper-left mode in the RTHS catalog possess many more offspring than other parents with the same magnitude (Figure 8d). This is consistent with dynamic triggering of many events by passing seismic waves (e.g., Brodsky & van der Elst, 2014).

In the WBV catalog, most events in the upper-left mode occurred around 2010, 2012, and 2016, which coincide with the times of the 2010 $M7.2$ El Mayor earthquake and $M > 5$ local earthquakes in the SJFZ (Figure 9a). The spatial distribution of the events in this mode are scattered circularly around the SJFZ (Figure 9b). The space-time separations between these events and their parents can be up to 150 km and around 25 s. Most events in this mode follow two lines with slopes of 3–4 km/sec and 6–7 km/sec (Figure 9c), which correspond to P- and S-wave velocities, respectively (Allam & Ben-Zion, 2012). The parents of the events in this mode have similar offspring numbers to other parents with the same magnitude (Figure 9d), in contrast to what is observed in the RTHS catalog.

5. Discussion and Conclusions

We identify several types of seismicity patterns based on statistical characterization of nearest-neighbor proximities of earthquakes (e.g., Zaliapin & Ben-Zion, 2013a, 2013b, 2016a, 2016b). In addition to the general modes of background and clustered events discussed in previous studies, the analysis reveals variations between seismicity recorded at different regions and catalogs constructed by different methods. The main observed features include (Figure 10) different distributions (T, R) of rescaled time T and rescaled distance R for earthquake clusters with large versus small-to-medium mainshocks, a distinct mode of repeating earthquakes, and different types of background seismicity for induced versus regular tectonic earthquakes. We also observe significant differences in the relative proportion of background to clustered earthquakes in three catalogs for the SJFZ, and detect an

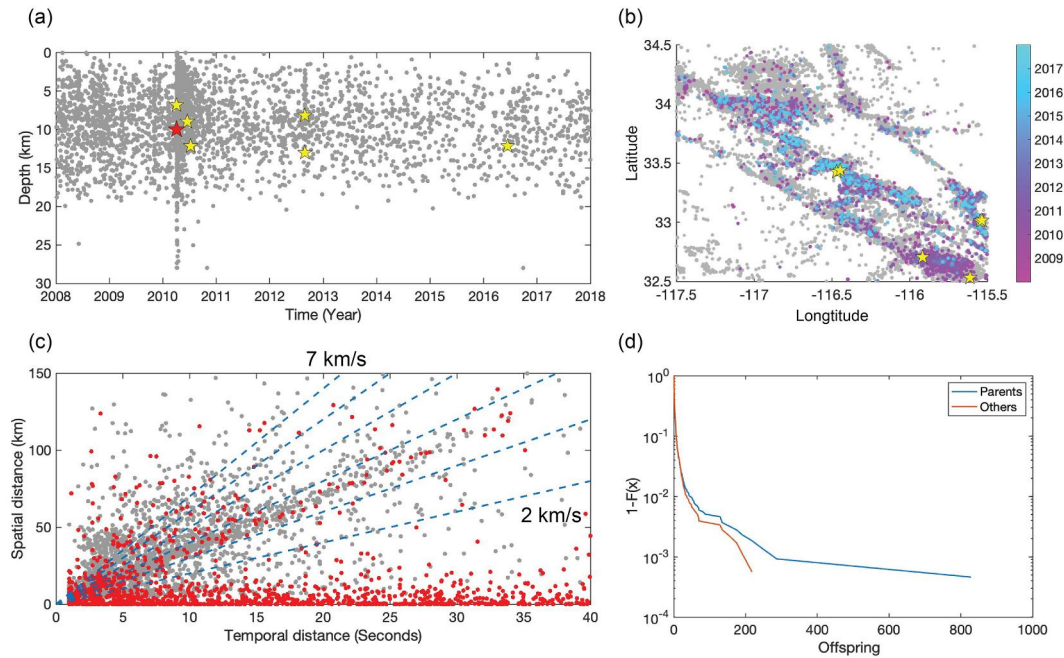


Figure 8. (a) The time–depth distribution of events in the red box of Figure 7b based on the RTHS catalog. The red star denotes the 2010 M7.2 El Mayor earthquake and the yellow stars mark $M > 5$ events in the San Jacinto Fault Zone (SJFZ). (b) A mapview of the events in the red box of Figure 7b with a colorbar denoting occurrence times. The yellow stars mark $M > 5$ events. (c) The physical space and time distance from events in the red box of Figure 7b to their parents in the RTHS catalog for the SJFZ. The red dots mark events occurring 1 month after the 2010 M7.2 El Mayor earthquake. The blue lines are possible migration velocities from 2 to 7 km/s. (d) The survival functions of the offspring number of parents of events in the red box of Figure 7b (blue curve) and the other nearest events with the same magnitude of these parents (orange curve).

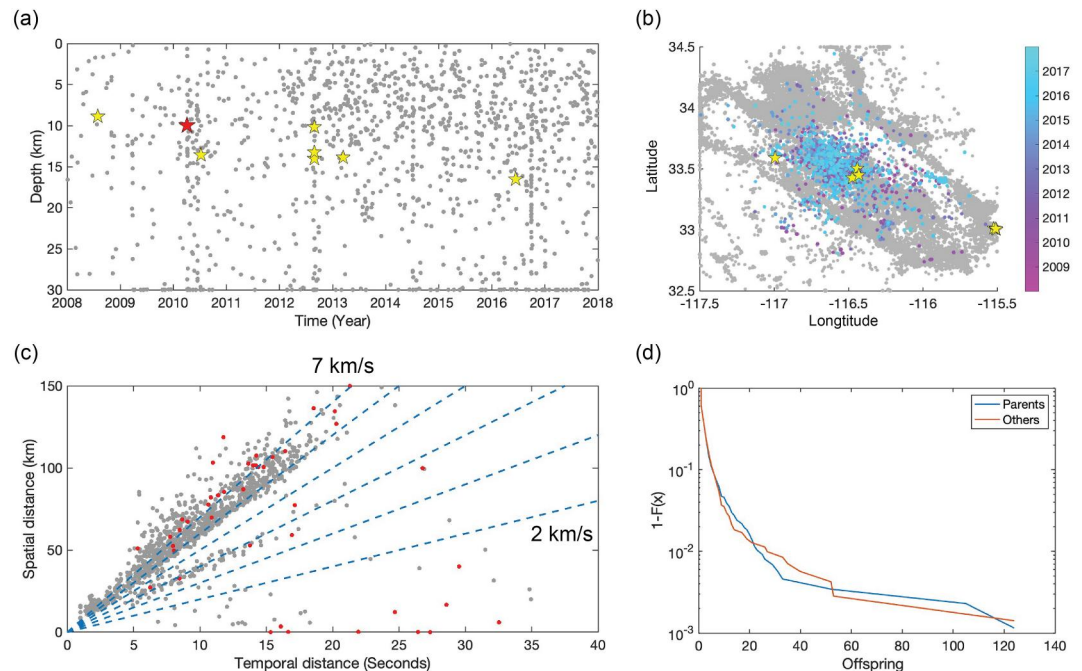


Figure 9. (a)–(d) Similar to Figures 8a–8d for events in the red box of Figure 7c based on the WBV catalog. The red dots in panel (c) mark events occurring 1 month after the 2010 M7.2 El Mayor earthquake.

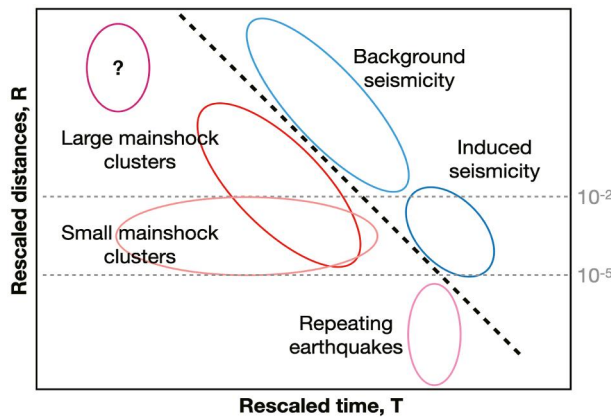


Figure 10. A schematic diagram of different seismicity patterns in the joint distribution of rescaled time (T) and space (R) components of the nearest-neighbor proximity (Equation 2).

anomalous mode in the nearest-neighbor diagrams of the SJFZ catalogs generated with template-matching and deep-learning. The results indicate that nearest-neighbor diagrams of earthquakes provide an efficient way of identifying different types of seismicity, and comparing statistical properties of earthquakes in catalogs generated by different techniques, thus providing information on the quality of the catalogs.

Earthquake clusters with $M > 5.5$ mainshocks have approximately diagonal (T, R) distribution of the clustered mode in the nearest-neighbor diagram, while the distribution for clusters with $M \leq 5.5$ mainshocks are approximately parallel to the horizontal (rescaled T) axis (Figure 3). Figure S11 in Supporting Information S1 demonstrates variations in the nearest-neighbor diagrams of earthquake clusters with different mainshock ranges. The clustered modes vary progressively from horizontal to diagonal shape with increasing mainshock magnitude, which becomes pronounced for mainshocks with $M > 5.5$. These results on different clustering/triggering patterns are at odds with a common assumption that small and large events produce the same stress change at distances normalized by rupture length (e.g., Helmstetter

et al., 2005). However, the assumed scale-free framework holds only for an infinite fault in an infinite solid. For natural faults with finite seismogenic zones limited to the brittle upper crust, there is a mechanical distinction between events that do or do not break most or the entire seismogenic zone (e.g., $M \sim 6$ in California). Seismogenic-zone-size events release the stress stored in the brittle crust around that fault section during the interseismic period, while transferring it to the surrounding regions. In contrast, smaller events transfer stress internally in the crust and modify the stress heterogeneities but do not change the average stress over the seismogenic zone (e.g., Ben-Zion, 1996). It is thus reasonable to expect a change in the dynamics of seismicity for mainshocks with rupture areas approaching the size of the seismogenic zone. We note that a length scale corresponding to the thickness of the brittle crust was also found to produce a strong change in the decay of aftershocks density with distance from mainshocks (Moradpour et al., 2014).

The mode of repeating earthquakes, observed in the creeping section of the SAF, is characterized by a vertical (T, R) distribution with very large rescaled time T and very small rescaled space R , compared to the clustered mode of regular earthquakes. This is consistent with the highly similar locations of repeating earthquakes and provides a rapid and simple way of identifying repeating events. The clustered and repeating earthquake modes can be roughly separated by the line $R = 10^{-5}$. The vertical (T, R) distribution also implies that the values of rescaled time T between event pairs in this mode is nearly identical. Since the rescaled time involves a division of the time by the seismic moment of the parent event (to some power), this is consistent with the observed correlation between the logarithm of the recurrence intervals of repeating earthquakes in the SAF and $\log(M_0)$ of previous events (Chen et al., 2007).

The catalogs for the GyGF and CoGF exhibit a background mode with larger rescaled time T and smaller rescaled space R compared to tectonic background seismicity. This observation is consistent with previous studies (e.g., Schoenball et al., 2015; Zaliapin & Ben-Zion, 2016a). Most of the seismicity in the examined geothermal fields behaves as background events with small triggering distances due to the concentrated loadings of fluid injection. We also observed that the background seismicity in regions dominated by induced seismicity are mostly distributed below the line $R = 10^{-2}$.

The comparisons of the joint distributions in the HYS, the WBV, and the RTHS catalogs for the SJFZ provide useful information on general features of the catalogs. Since there is no way of validating earthquake catalogs, comparing statistical properties of background and clustered seismicity contributes (along with estimates of M_c , the total number of events, and visual inspections) to assessing the quality of catalogs constructed differently for overlapping space-time domains. This is an issue of increasing importance with the growing number of techniques including multiple template- and AI-based methods for developing catalogs.

The RTHS catalog has a higher proportion of clustered to background events compared to the HYS catalog. This is because template-based detections are spatially close to the templates and hence likely to be clustered and enhance the clustered mode. The WBV catalog has a much lower proportion of clustered to background events than the HYS and RTHS catalogs. The WBV catalog differs from these two other catalogs in that (a) it uses data

from many more stations around the SJFZ, (b) it is based on deep-learning, and (c) it may include a larger number of events that are not relocated but are validated by a quality-control test (White et al., 2019). The WBV catalog retains nearly 29% of events that are not relocated by the Growclust algorithm, in contrast to the HYS catalog that only retains 0.1% of events that are not relocated (Hauksson et al., 2012). The non-relocated events tend to have more scattered locations compared to relocated events and are more likely to be identified as background seismicity in the nearest-neighbor analysis.

The proportion of clustered events is an important indicator of characteristics of the examined region because it reflects key properties of the major faults and ambient conditions such as the temperature and fluid content (e.g., Zaliapin & Ben-Zion, 2013b, 2016a, 2022). However, the comparisons between the three catalogs for the SJFZ show that this proportion can depend strongly on the method used to construct the catalog. This implies that establishing connections between characteristics of seismicity and the crust can only be done in a relative sense by performing comparative analysis of a large region (e.g., southern California) using a single regional catalog.

The upper-left mode(s) observed in the joint (T , R) distribution in the WBV and RTHS catalogs may reflect dynamic triggering during the passage of surface waves from remote earthquakes (Brodsky & van der Elst, 2014; Hill et al., 1993) and/or artifacts in the catalog construction (e.g., multiple locations of some events or ghost events at some time windows). Most of the events in the upper-left mode in the joint distribution of (T , R) in the RTHS catalog occur around the time of the 2010 $M7.2$ El Mayor earthquake (Figure 8a). This is consistent with previous studies on dynamic triggering of local seismicity in the SJFZ (e.g., Meng & Peng, 2014). The spatial distribution of events in this mode mostly follows the main faults in the SJFZ (Figure 8b). The space-time distribution between the events in the upper-left mode and their parents generally follows the average S-wave velocity in the SJFZ (Figure 8c). The parents of events in this mode also have more offspring compared to other parents (Figure 8d). However, previous studies also indicated that the thresholds used to detect events with template matching and multiple environmental and anthropogenic sources may result in false detections (Johnson et al., 2019; Meng et al., 2013; Sheng et al., 2023), which might also lead to the upper-left mode in the joint distribution (T , R).

Corresponding results for the WBV catalog show that most events in this mode occur around the time of the 2010 El Mayor earthquake and the local $M > 5$ earthquakes in 2012 and 2016 (Figure 9a). The events in the upper-left mode in the WBV catalog have a scattered circular distribution in the SJFZ (Figure 9b) and they mostly follow a high velocity line in the space-time distribution that may correspond to a P wave (Figure 9c), which is less likely to produce dynamic triggering but may lead to spurious automatic detections. The similar offspring number for parents of the events in the upper-left mode and other parents (Figure 9d) also suggests that dynamic triggering is an unlikely cause of these events. The WBV catalog uses automatic picking of P- and S-wave arrival time based on travel time prediction of the used deep-learning algorithm. The events in the upper-left mode may thus reflect false detection generated by the estimated P- and S-wave arrival time.

Zaliapin and Ben-Zion (2015) examined in detail effects of location errors and short-term incompleteness on properties of nearest-neighbor clustering analysis. Location errors that are larger than or are comparable to the rupture length of the analyzed events tend to result in increased spatial separation between parents and offspring, overestimated background seismicity rates, and underestimated rates of clustered events. For the examined relocated catalogs, these issues are relevant primarily to $M < 2$ events (Ben-Zion, 2008; Zaliapin & Ben-Zion, 2015). To ensure that the highlighted results are not strongly affected by location errors, we repeated the analysis leading to Figure 3 using larger minimum event size ($M \geq 2$ and $M \geq 3$) and the general patterns for clusters with different mainshock sizes remain similar (Figure S4 in Supporting Information S1). We also performed nearest-neighbor analysis in multiple subregions likely to have different errors (because of different densities of stations) and obtained consistent results (Figure S12 in Supporting Information S1). Short-term incompleteness was found by Zaliapin and Ben-Zion (2015) to modify primarily the estimated b -values and (related to this) the estimated average magnitude in a cluster that is inversely proportional to the b -value (e.g., Aki, 1965; Eneva & Ben-Zion, 1997). While these issues can affect some aspects of our analyses, they are unlikely to modify strongly the general results highlighted in the paper.

In conclusion, repeating earthquakes, induced seismicity, and earthquake clusters with large versus small-to-medium mainshocks occupy distinguished positions in the joint distribution (T , R) of nearest-neighbor earthquake proximities. Our results demonstrate that nearest-neighbor proximity diagrams provide a simple and effective way to identify these types of earthquake patterns using existing earthquake catalogs. The results may be

followed by additional analyses using other techniques, which are more elaborate and typically involve additional model assumptions and user choices (e.g., Gao et al., 2023). Nearest-neighbor diagrams also provide a useful tool to assess the quality of earthquake catalogs and compare key features of seismicity in catalogs generated by different techniques.

Data Availability Statement

Data products for this study were accessed through (NCEDC, 2014; SCEDC, 2013). The Schaff and Waldhauser (2005) and Waldhauser and Schaff (2008) catalogs used in the study are from (NCEDC, 2014). The employed Hauksson et al. (2012, extended) and Ross et al. (2019) catalogs are from the SCEDC and are available via (SCEDC, 2013). The White et al. (2021a, 2021b) catalog is available via (White et al., 2021a).

Acknowledgments

We dedicate this paper to our close friend and colleague Ilya Zaliapin who passed away tragically in May 2023. The manuscript benefited from constructive comments by referee Joern Davidsen, an anonymous referee, an anonymous associate editor and editor Rachel Abrecrombie. The study was funded by the National Science Foundation (Grants EAR-1723033 and EAR-1722561) and the Southern California Earthquake Center (based on NSF Cooperative Agreement EAR-1600087 and USGS Cooperative Agreement G17 AC00047).

References

- Aki, K. (1965). Maximum likelihood estimate of b in the formula $\log N = a - bM$ and its confidence limits. *Bulletin of the Earthquake Research Institute*, 43, 237–239.
- Allam, A. 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. <https://doi.org/10.1111/j.1365-246X.2012.05544.x>
- Baiesi, M., & Paczuski, M. (2004). Scale-free networks of earthquakes and aftershocks. *Physical Review E*, 69(6), 066106. <https://doi.org/10.1103/PhysRevE.69.066106>
- Ben-Zion, Y. (1996). Stress, slip and earthquakes in models of complex single-fault systems incorporating brittle and creep deformations. *Journal of Geophysical Research*, 101, 5677–5706.
- Ben-Zion, Y. (2008). Collective behavior of earthquakes and faults: Continuum-discrete transitions, progressive evolutionary changes, and different dynamic regimes. *Reviews of Geophysics*, 46(4). <https://doi.org/10.1029/2008RG000260>
- Ben-Zion, Y., & Lyakhovskiy, 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>
- Ben-Zion, Y., & Sammis, C. G. (2003). Characterization of fault zones. *Pure and Applied Geophysics*, 160, 677–715. https://doi.org/10.1007/978-3-0348-8010-7_11
- Ben-Zion, Y., & Zaliapin, I. (2020). Localization and coalescence of seismicity before large earthquakes. *Geophysical Journal International*, 223(1), 561–583. <https://doi.org/10.1093/gji/ggaa315>
- Blisniuk, K., Oskin, M., Mériaux, A. S., Rockwell, T., Finkel, R. C., & Ryerson, F. J. (2013). Stable, rapid rate of slip since inception of the San Jacinto fault, California. *Geophysical Research Letters*, 40(16), 4209–4213. <https://doi.org/10.1002/grl.50819>
- Brodsky, E. E., & van der Elst, N. J. (2014). The uses of dynamic earthquake triggering. *Annual Review of Earth and Planetary Sciences*, 42(1), 317–339. <https://doi.org/10.1146/annurev-earth-060313-054648>
- Burford, R. O., & Harsh, P. W. (1980). Slip on the San Andreas fault in central California from alignment array surveys. *Bulletin of the Seismological Society of America*, 70(4), 1233–1261. <https://doi.org/10.1785/BSSA0700041233>
- California Institute of Technology (Caltech) (1926). Southern California Seismic Network. International federation of digital seismograph networks. Other/Seismic Network. <https://doi.org/10.7914/SN/CI>
- Chen, K. H., Nadeau, R. M., & Rau, R. J. (2007). Towards a universal rule on the recurrence interval scaling of repeating earthquakes? *Geophysical Research Letters*, 34(16). <https://doi.org/10.1029/2007GL030554>
- de Arcangelis, L., Godano, C., & Lippiello, E. (2018). The overlap of aftershock coda waves and short-term postseismic forecasting. *Journal of Geophysical Research*, 123, 5661–5674. <https://doi.org/10.1029/2018JB015518>
- Eberhart-Phillips, D., & Oppenheimer, D. H. (1984). Induced seismicity in The Geysers geothermal area, California. *Journal of Geophysical Research*, 89(B2), 1191–1207. <https://doi.org/10.1029/JB089iB02p01191>
- Enescu, B., Hainzl, S., & Ben-Zion, Y. (2009). Correlations of seismicity patterns in Southern California with surface heat flow data. *Bulletin of the Seismological Society of America*, 99(6), 3114–3123. <https://doi.org/10.1785/0120080038>
- Eneva, M., & Ben-Zion, Y. (1997). Techniques and parameters to analyze seismicity patterns associated with large earthquakes. *Journal of Geophysical Research*, 102(B8), 17785–17795. <https://doi.org/10.1029/97jb00994>
- Fagereng, Å., & Diener, J. F. (2011). San Andreas Fault tremor and retrograde metamorphism. *Geophysical Research Letters*, 38(23). <https://doi.org/10.1029/2011GL049550>
- Felzer, K. R., Abercrombie, R. E., & Ekstrom, G. (2004). A common origin for aftershocks, foreshocks, and multiplets. *Bulletin of the Seismological Society of America*, 94(1), 88–98. <https://doi.org/10.1785/0120030069>
- Feng, Q., & Lees, J. M. (1998). Microseismicity, stress, and fracture in the Coso geothermal field, California. *Tectonophysics*, 289(1–3), 221–238. [https://doi.org/10.1016/S0040-1951\(97\)00317-X](https://doi.org/10.1016/S0040-1951(97)00317-X)
- Fialko, Y. (2006). Interseismic strain accumulation and the earthquake potential on the southern San Andreas fault system. *Nature*, 441(7096), 968–971. <https://doi.org/10.1038/nature04797>
- Gao, D., Kao, H., & Liu, J. (2023). Identification of repeating earthquakes: Controversy rectification. *Seismological Research Letters*, 1–11. <https://doi.org/10.1785/0220230124>
- Gentili, S., Peresan, A., Talebi, M., Zare, M., & Di Giovambattista, R. (2019). A seismic quiescence before the 2017 Mw 7.3 Sarpol Zahab (Iran) earthquake: Detection and analysis by improved RTL method. *Physics of the Earth and Planetary Interiors*, 290, 10–19. <https://doi.org/10.1016/j.pepi.2019.02.010>
- Gu, C., Schumann, A. Y., Baiesi, M., & Davidsen, J. (2013). Triggering cascades and statistical properties of aftershocks. *Journal of Geophysical Research: Solid Earth*, 118(8), 4278–4295. <https://doi.org/10.1002/jgrb.50306>
- Gunasekera, R. C., Foulger, G. R., & Julian, B. R. (2003). Reservoir depletion at The Geysers geothermal area, California, shown by four-dimensional seismic tomography. *Journal of Geophysical Research*, 108(B3). <https://doi.org/10.1029/2001JB000638>
- Hauksson, E., & Unruh, J. (2007). Regional tectonics of the Coso geothermal area along the intracontinental plate boundary in central eastern California: Three-dimensional Vp and Vp/Vs models, spatial-temporal seismicity patterns, and seismogenic deformation. *Journal of Geophysical Research*, 112(B6). <https://doi.org/10.1029/2006JB004721>

- 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. <https://doi.org/10.1785/0120120010>
- Helmstetter, A., Kagan, Y. Y., & Jackson, D. D. (2005). Importance of small earthquakes for stress transfers and earthquake triggering. *Journal of Geophysical Research*, 110(B5). <https://doi.org/10.1029/2004JB003286>
- Hill, D. P., Reasenber, P. A., Michael, A., Arabaz, W. J., Beroza, G., Brumbaugh, D., et al. (1993). Seismicity remotely triggered by the magnitude 7.3 Landers, California, earthquake. *Science*, 260(5114), 1617–1623. <https://doi.org/10.1126/science.260.5114.1617>
- Irwin, W. P., & Barnes, I. (1975). Effect of geologic structure and metamorphic fluids on seismic behavior of the San Andreas fault system in central and northern California. *Geology*, 3(12), 713–716. [https://doi.org/10.1130/0091-7613\(1975\)3<713:EOGSAM>2.0.CO;2](https://doi.org/10.1130/0091-7613(1975)3<713:EOGSAM>2.0.CO;2)
- Johnson, C. W., Meng, H., Vernon, F., & Ben-Zion, Y. (2019). Characteristics of ground motion generated by wind interaction with trees, structures, and other surface obstacles. *Journal of Geophysical Research: Solid Earth*, 124(8), 8519–8539. <https://doi.org/10.1029/2018JB017151>
- Karimi, K., & Davidsen, J. (2023). Separating primary and secondary mechanisms for seismicity induced by hydraulic fracturing. *Bulletin of the Seismological Society of America*, 113(5), 1–10. <https://doi.org/10.1785/0120220200>
- Kirby, S. H., Wang, K., & Brocher, T. M. (2014). A large mantle water source for the northern San Andreas fault system: A ghost of subduction past. *Earth Planets and Space*, 66(1), 1–18. <https://doi.org/10.1186/1880-5981-66-67>
- Kisslinger, C., & Jones, L. M. (1991). Properties of aftershock sequences in southern California. *Journal of Geophysical Research*, 96(B7), 11947–11958. <https://doi.org/10.1029/91JB01200>
- Kossobokov, V. G., & Nekrasova, A. K. (2017). Characterizing aftershock sequences of the recent strong earthquakes in Central Italy. *Pure and Applied Geophysics*, 174(10), 3713–3723. <https://doi.org/10.1007/s00024-017-1624-9>
- Kwiatak, G., Martínez-Garzón, P., Dresen, G., Bohnhoff, M., Sone, H., & Hartline, C. (2015). Effects of long-term fluid injection on induced seismicity parameters and maximum magnitude in northwestern part of The Geysers geothermal field. *Journal of Geophysical Research: Solid Earth*, 120(10), 7085–7101. <https://doi.org/10.1002/2015JB012362>
- Liu, Y. K., Ross, Z. E., Cochran, E. S., & Lapusta, N. (2022). A unified perspective of seismicity and fault coupling along the San Andreas Fault. *Science Advances*, 8(8), eabk1167. <https://doi.org/10.1126/sciadv.abk1167>
- Majer, E. L., & Peterson, J. E. (2007). The impact of injection on seismicity at The Geysers, California Geothermal Field. *International Journal of Rock Mechanics and Mining Sciences*, 44(8), 1079–1090. <https://doi.org/10.1016/j.ijrmm.2007.07.023>
- Manley, C. R., & Bacon, C. R. (2000). Rhyolite thermobarometry and the shallowing of the magma reservoir, Coso volcanic field, California. *Journal of Petrology*, 41(1), 149–174. <https://doi.org/10.1093/ptrology/41.1.149>
- Martínez-Garzón, P., Kwiatak, G., Sone, H., Bohnhoff, M., Dresen, G., & Hartline, C. (2014). Spatiotemporal changes, faulting regimes, and source parameters of induced seismicity: A case study from The Geysers geothermal field. *Journal of Geophysical Research: Solid Earth*, 119(11), 8378–8396. <https://doi.org/10.1002/2014JB011385>
- Martínez-Garzón, P., Zaliapin, I., Ben-Zion, Y., Kwiatak, G., & Bohnhoff, M. (2018). Comparative study of earthquake clustering in relation to hydraulic activities at geothermal fields in California. *Journal of Geophysical Research: Solid Earth*, 123(5), 4041–4062. <https://doi.org/10.1029/2017JB014972>
- McGuire, J. J., Boettcher, M. S., & Jordan, T. H. (2005). Foreshock sequences and short-term earthquake predictability on East Pacific Rise transform faults. *Nature*, 434(7032), 457–461. <https://doi.org/10.1038/nature03377>
- Meng, X., & Peng, Z. (2014). Seismicity rate changes in the Salton Sea geothermal field and the San Jacinto fault zone after the 2010 M W 7.2 El Mayor-Cucapah earthquake. *Geophysical Journal International*, 197(3), 1750–1762. <https://doi.org/10.1093/gji/ggu085>
- Meng, X., Peng, Z., & Hardebeck, J. L. (2013). Seismicity around Parkfield correlates with static shear stress changes following the 2003 Mw6.5 San Simeon earthquake. *Journal of Geophysical Research: Solid Earth*, 118(7), 3576–3591. <https://doi.org/10.1002/jgrb.50271>
- Mizrahi, L., Nandan, S., & Wiemer, S. (2021). Embracing data incompleteness for better earthquake forecasting. *Journal of Geophysical Research: Solid Earth*, 126(12), e2021JB022379. <https://doi.org/10.1029/2021JB022379>
- Monastero, F. C. (2002). An overview of industry-military cooperation in the development of power operations at the coso geothermal field in Southern California. *Geothermal Resources Council - Bulletin*, 31, 188–195.
- Moradpour, J., Hainzl, S., & Davidsen, J. (2014). Nontrivial decay of aftershock density with distance in Southern California. *Journal of Geophysical Research: Solid Earth*, 119(7), 5518–5535. <https://doi.org/10.1002/2014JB010940>
- Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C. (2020). Earthquake transformer—An attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, 11(1), 1–12. <https://doi.org/10.1038/s41467-020-17591-w>
- Nadeau, R. M., Foxall, W., & McEvelly, T. V. (1995). Clustering and periodic recurrence of microearthquakes on the San Andreas fault at Parkfield, California. *Science*, 267(5197), 503–507. <https://doi.org/10.1126/science.267.5197.503>
- Nadeau, R. M., & Johnson, L. R. (1998). Seismological studies at Parkfield VI: Moment release rates and estimates of source parameters for small repeating earthquakes. *Bulletin of the Seismological Society of America*, 88(3), 790–814. <https://doi.org/10.1785/BSSA0880030790>
- Nadeau, R. M., & McEvelly, T. V. (1999). Fault slip rates at depth from recurrence intervals of repeating microearthquakes. *Science*, 285(5428), 718–721. <https://doi.org/10.1126/science.285.5428.718>
- Nadeau, R. M., & McEvelly, T. V. (2004). Periodic pulsing of characteristic microearthquakes on the San Andreas fault. *Science*, 303(5655), 220–222. <https://doi.org/10.1126/science.1090353>
- NCEDC. (2014). Northern California earthquake data center. UC Berkeley Seismological Laboratory. [Dataset]. <https://doi.org/10.7932/NCEDC>
- Ogata, Y. (1999). Seismicity analysis through point-process modeling: A review. *Pure and Applied Geophysics*, 155, 471–507. https://doi.org/10.1007/978-3-0348-8677-2_14
- Omori, F. (1895). *On the after-shocks of earthquakes* (Doctoral dissertation). The University of Tokyo.
- Peresan, A., & Gentili, S. (2018). Seismic clusters analysis in Northeastern Italy by the nearest-neighbor approach. *Physics of the Earth and Planetary Interiors*, 274, 87–104. <https://doi.org/10.1016/j.pepi.2017.11.007>
- Petrillo, G., & Lippiello, E. (2020). Testing of the foreshock hypothesis within an epidemic like description of seismicity. *Geophysical Journal International*, 225(2), 1236–1257. <https://doi.org/10.1093/gji/ggaa611>
- Powers, P. M., & Jordan, T. H. (2010). Distribution of seismicity across strike-slip faults in California. *Journal of Geophysical Research*, 115(B5). <https://doi.org/10.1029/2008JB006234>
- Reverso, T., Marsan, D., & Helmstetter, A. (2015). Detection and characterization of transient forcing episodes affecting earthquake activity in the Aleutian Arc system. *Earth and Planetary Science Letters*, 412, 25–34. <https://doi.org/10.1016/j.epsl.2014.12.012>
- Richter, C. F. (1958). *Elementary seismology* (p. 768). W.H. Freeman.
- 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. <https://doi.org/10.1007/s00024-014-0955-z>

- Ross, Z. E., Ben-Zion, Y., & Zaliapin, I. (2022). Geometrical properties of seismicity in California. *Geophysical Journal International*, 231(1), 493–504. <https://doi.org/10.1093/gji/ggac189>
- 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. <https://doi.org/10.1126/sciadv.1601946>
- Ross, Z. E., Trugman, D. T., Hauksson, E., & Shearer, P. M. (2019). Searching for hidden earthquakes in Southern California. *Science*, 364(6442), 767–771. <https://doi.org/10.1126/science.aaw688>
- Ruhl, C. J., Abercrombie, R. E., Smith, K. D., & Zaliapin, I. (2016). Complex spatiotemporal evolution of the 2008 Mw 4.9 Mogul earthquake swarm (Reno, Nevada): Interplay of fluid and faulting. *Journal of Geophysical Research: Solid Earth*, 121(11), 8196–8216. <https://doi.org/10.1002/2016JB013399>
- Sammis, C., King, G., & Biegel, R. (1987). The kinematics of gouge deformation. *Pure and Applied Geophysics*, 125(5), 777–812. <https://doi.org/10.1007/BF00878033>
- Sanders, C., & Magistrale, H. (1997). Segmentation of the northern San Jacinto fault zone, southern California. *Journal of Geophysical Research*, 102(B12), 27453–27467. <https://doi.org/10.1029/97JB02377>
- SCEDC. (2013). Southern California earthquake center. Caltech [Dataset]. <https://doi.org/10.7909/C3WD3xH1>
- Schaff, D. P., & Waldhauser, F. (2005). Waveform cross-correlation-based differential travel-time measurements at the Northern California Seismic Network. *Bulletin of the Seismological Society of America*, 95(6), 2446–2461. <https://doi.org/10.1785/0120040221>
- Schoenball, M., Davatzes, N. C., & Glen, J. M. (2015). Differentiating induced and natural seismicity using space-time-magnitude statistics applied to the Coso Geothermal field. *Geophysical Research Letters*, 42(15), 6221–6228. <https://doi.org/10.1002/2015GL064772>
- Shearer, P. M. (2012). Self-similar earthquake triggering, Båth's law, and foreshock/aftershock magnitudes: Simulations, theory, and results for southern California. *Journal of Geophysical Research*, 117(B6). <https://doi.org/10.1029/2011JB008957>
- Sheng, Y., Mordret, A., Brenguier, F., Boué, P., Vernon, F., Takeda, T., et al. (2023). Seeking repeating anthropogenic seismic sources: Implications for seismic velocity monitoring at fault zones. *Journal of Geophysical Research: Solid Earth*, 128(1), e2022JB024725. <https://doi.org/10.1029/2022JB024725>
- Trugman, D. T., & Ben-Zion, Y. (2023). Coherent spatial variations in the productivity of earthquake sequences in California and Nevada. *The Seismic Record*, 3(4), 322–331. <https://doi.org/10.1785/0320230039>
- 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. <https://doi.org/10.1785/0220160188>
- Utsu, T. (2002). Statistical features of seismicity. *International Geophysics Series*, 81(A), 719–732. [https://doi.org/10.1016/S0074-6142\(02\)80246-7](https://doi.org/10.1016/S0074-6142(02)80246-7)
- Vidale, J. E., Ellsworth, W. L., Cole, A., & Marone, C. (1994). Variations in rupture process with recurrence interval in a repeated small earthquake. *Nature*, 368(6472), 624–626. <https://doi.org/10.1038/368624a0>
- Waldhauser, F. (2001). HypoDD: A computer program to compute double-difference earthquake locations. <https://doi.org/10.7916/D8SN072H>
- Waldhauser, 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. <https://doi.org/10.1785/0120000006>
- Waldhauser, F., & Schaff, D. P. (2008). Large-scale relocation of two decades of Northern California seismicity using cross-correlation and double-difference methods. *Journal of Geophysical Research*, 113(B8). <https://doi.org/10.1029/2007JB005479>
- White, M. C., 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(7), 6908–6930. <https://doi.org/10.1029/2019JB017641>
- White, M. C., Ben-Zion, Y., & Vernon, F. L. (2021a). Detailed earthquake catalog for the San Jacinto Fault Zone in southern California (2008–2020), Version 1 [dataset]. *Mendeley Data*. <https://doi.org/10.17632/7ywkdx7c62.1>
- White, M. C., Ben-Zion, Y., & Vernon, F. L. (2021b). A detailed earthquake catalog for the San Jacinto Fault-zone region in Southern California and the period 2008–2020. In *Poster presentation at 2021 SCEC annual meeting. SCEC contribution 11485*.
- Wiemer, S., & Wyss, M. (2000). Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the western United States, and Japan. *Bulletin of the Seismological Society of America*, 90(4), 859–869. <https://doi.org/10.1785/0119990114>
- Zaliapin, I., & Ben-Zion, Y. (2013a). Earthquake clusters in southern California I: Identification and stability. *Journal of Geophysical Research: Solid Earth*, 118(6), 2847–2864. <https://doi.org/10.1002/jgrb.50179>
- Zaliapin, I., & Ben-Zion, Y. (2013b). Earthquake clusters in southern California II: Classification and relation to physical properties of the crust. *Journal of Geophysical Research: Solid Earth*, 118(6), 2865–2877. <https://doi.org/10.1002/jgrb.50178>
- Zaliapin, I., & Ben-Zion, Y. (2015). Artifacts of earthquake location errors and short-term incompleteness on seismicity clusters in southern California. *Geophysical Journal International*, 202(3), 1949–1968. <https://doi.org/10.1093/gji/ggv259>
- Zaliapin, I., & Ben-Zion, Y. (2016a). Discriminating characteristics of tectonic and human-induced seismicity. *Bulletin of the Seismological Society of America*, 106(3), 846–859. <https://doi.org/10.1785/0120150211>
- Zaliapin, I., & Ben-Zion, Y. (2016b). A global classification and characterization of earthquake clusters. *Geophysical Journal International*, 207(1), 608–634. <https://doi.org/10.1093/gji/ggw300>
- Zaliapin, I., & Ben-Zion, Y. (2020). Earthquake declustering using the nearest-neighbor approach in space-time-magnitude domain. *Journal of Geophysical Research: Solid Earth*, 125(4), e2018JB017120. <https://doi.org/10.1029/2018jb017120>
- Zaliapin, I., & Ben-Zion, Y. (2022). Perspectives on clustering and declustering of earthquakes. *Seismological Research Letters*, 93(1), 386–401. <https://doi.org/10.1785/0220210127>
- Zaliapin, I., Gabrielov, A., Keilis-Borok, V., & Wong, H. (2008). Clustering analysis of seismicity and aftershock identification. *Physical Review Letters*, 101(1), 018501. <https://doi.org/10.1103/PhysRevLett.101.018501>
- Zhang, M., Ellsworth, W. L., & Beroza, G. C. (2019). Rapid earthquake association and location. *Seismological Research Letters*, 90(6), 2276–2284. <https://doi.org/10.1785/0220190052>