Robust Quantification of Earthquake Clustering: Overcoming the Artifacts of Catalog Uncertainties

Report for SCEC Award #15120
Submitted March 10, 2016

Investigators: Ilya Zaliapin (UNR) and Yehuda Ben-Zion (USC)

I. Project Overview

A. Abstract

In the box below, describe the project objectives, methodology, and results obtained and their significance. If this work is a continuation of a multi-year SCEC-funded project, please include major research findings for all previous years in the abstract. (Maximum 250 words.)

The project aimed at improving the ability to use seismicity clusters to clarify physical processes associated with faulting in the crust. Specific science targets included (i) understanding the effects of (space-time varying) errors in earthquake catalogs and distinguishing such errors from genuine changes of seismicity, and (ii) finding cluster signatures that can distinguish natural from human-induced seismicity. The project builds on the PIs results from previous SCEC projects on quantitative characterization of earthquake clustering in space and time in relation to different event sizes and physical properties of the lithosphere.

First, we document and quantify effects of catalog uncertainties on results of statistical cluster analyses of seismicity in southern California. We present statistical evidence for three artifacts: (1) Increased distance between offspring and parents. (2) Underestimated clustering. (3) Overestimated background rates. We also find that short-term incompleteness leads to (4) Apparent magnitude dependence and temporal fluctuations of b-values. Next, we analyze statistical features of background and clustered subpopulations of earthquakes in different regions in an effort to distinguish between human-induced and natural seismicity. Induced seismicity is shown to have (i) higher rate of background events, (ii) faster temporal offspring decay, (iii) higher rate of repeating events, (iv) larger proportion of small clusters, and (v) larger spatial separation between parent and offspring.

The results can inform a range of studies focused on small-magnitude seismicity patterns in the presence of catalog uncertainties, as well as to improve seismic hazards assessment related to induced earthquakes.

B. SCEC Annual Science Highlights

Each year, the Science Planning Committee reviews and summarizes SCEC research accomplishments, and presents the results to the SCEC community and funding agencies. Rank (in order of preference) the sections in which you would like your project results to appear. Choose up to 3 working groups from below and re-order them according to your preference ranking.

Seismology
Earthquake Forecasting and Predictability (EFP)
Collaboratory for the Study of Earthquake Predictability (CSEP)

C. Exemplary Figure

Select one figure from your project report that best exemplifies the significance of the results. The figure may be used in the SCEC Annual Science Highlights and chosen for the cover of the Annual Meeting Proceedings Volume. In the box below, enter the figure number from the project report, figure caption and figure credits.

Figure 3: Change of clustering style in Coso and Salton Sea geothermal fields after beginning of active geothermal production. Figure shows the distribution of the rescaled time to parent T for the offspring within one parent rupture length from the parent. (a,c) Coso geothermal field; production began in 1987. (b,d) Salton Sea geothermal field; production began during 1988-1992. (a,b) Offspring before geothermal production. (c,d) Offspring during geothermal production. After *Zaliapin and Ben-Zion, BSSA* (2016).

D. SCEC Science Priorities

In the box below, please list (in rank order) the SCEC priorities this project has achieved. See *https://www.scec.org/research/priorities* for list of SCEC research priorities. *For example: 6a, 6b, 6c*

2b, 2f, 4e

E. Intellectual Merit

How does the project contribute to the overall intellectual merit of SCEC? *For example: How does the research contribute to advancing knowledge and understanding in the field and, more specifically, SCEC research objectives? To what extent has the activity developed creative and original concepts?*

The study combines novel approaches to earthquake cluster identification/classification and high quality earthquake catalogs from different environments toward improved understanding of seismicity in relation to human-induced earthquakes. An ability to track the evolving response of the crust to different loadings may be used to monitor the build up of stress in a region. The developed tools and results can have transformative impact on analysis of seismic hazard in active tectonic environments, oil and other production areas, and regions containing both, such as California.

F. Broader Impacts

How does the project contribute to the broader impacts of SCEC as a whole? *For example: How well has the activity promoted or supported teaching, training, and learning at your institution or across SCEC? If your project included a SCEC intern, what was his/her contribution? How has your project broadened the participation of underrepresented groups? To what extent has the project enhanced the infrastructure for research and education (e.g., facilities, instrumentation, networks, and partnerships)? What are some possible benefits of the activity to society?*

The addressed problems on natural/induced seismicity have critical societal and economic importance. The research can inform and impact significantly various related studies on earthquake physics. The developed cluster framework can be applicable to other processes that develop in space-time-energy domains (e.g., river/subsurface flows, aerosol dynamics, chemical reactions, and fires).

G. Project Publications

Papers

1. Zaliapin, I. and Y. Ben-Zion (2016) Discriminating characteristics of tectonic and human-induced seismicity. *Bull. Seismol. Soc. Am*., accepted.
2. Zaliapin, I. and Y. Ben-Zion (2015) Artifacts of earthquake location errors and short-term incompleteness on seismicity clusters in southern California. *Geophys. J. Intl.,* 202 (3): 1949-1968. doi: 10.1093/gji/ggv259.

Presentations

1. Zaliapin, I. and Y. Ben-Zion (2015) Discriminating characteristics of tectonic and human-induced seismicity. Abstract S13B-2828 (poster) presented at 2015 Fall Meeting of AGU, San Francisco, California, December 14-18, 2015.
2. Zaliapin, I. and Y. Ben-Zion (2015) Discriminating characteristics of tectonic and human-induced seismicity. *Proc. of Southern California Earthquake Center (SCEC) 2015 Annual Meeting, Palm Springs, CA, September 12-16, 2015,* Vol. XXV, p.197, poster 146.

3. Zaliapin, I. and Y. Ben-Zion (2015) Distinguishing artifacts of earthquake catalog errors from genuine seismicity patterns*. 26th General Assembly of International Union of Geodesy and Geophysics*, IUGG-2960 (oral), Prague, Czech Republic, June 22-July 2, 2015
4. Zaliapin, I. and Y. Ben-Zion (2015) Distinguishing artifacts of earthquake catalog errors from genuine seismicity patterns. Poster #101 presented at *2015 Annual Meeting of Seismological Society of America*, Pasadena, California, April 20-22, 2015.
5. Ruhl, C., R. Abercrombie, K. Smith, and I. Zaliapin (2015) Inside an Earthquake Swarm: Objective Identification and Analysis of Spatiotemporal Subclusters of the Mogul 2008 Earthquake Swarm in Reno, NV. Abstract S51A-2647 (poster) presented at 2015 Fall Meeting of AGU, San Francisco, California, December 14-18, 2015.

II. Technical Report

The project uses the seismic cluster techniques that were shown useful for identification and classification of statistically significant seismicity clusters in Southern California in relation to physical properties of the crust [*Zaliapin et al.*, 2008; *Zaliapin and Ben-Zion*, 2011, 2013a,b; *Gu et al.*, 2013]. The techniques are based on the nearest-neighbor analysis of earthquake distances briefly summarized below.

Earthquake distance: Each event *i* in an earthquake catalog is characterized by its occurrence time t_i, hypocenter, and magnitude m_i. The distance between earthquakes *i* and *j* is asymmetric in time and is defined as [*Baiesi and Paczuski*, 2004]:

$$\eta_{ij} = \begin{cases} t_{ij}(r_{ij})^d 10^{-bm_i}, & t_{ij} > 0; \\ \infty, & t_{ij} \le 0. \end{cases} \tag{1}$$

Here $t_{ij} = t_j - t_i$ is the intercurrence time, $r_{ij} \ge 0$ is the spatial distance between the hypocenters, and d is the (possibly fractal) dimension of the hypocenters.

Nearest-neighbor analysis: For each earthquake *j*, we find its nearest neighbor *i* and the corresponding nearest-neighbor distance η_{ij}. The nearest neighbor of an earthquake is called the parent. Each event has a single parent and can be the parent of multiple events that are referred to as its offspring. We also consider the space and time distances between the pairs of nearest neighbors normalized by the magnitude of the parent:

$$T_{ij} = t_{ij} 10^{-qbm_i}; R_{ij} = (r_{ij})^d 10^{-pbm_i}; q + p = 1. \tag{2}$$

Parameters: We perform all analyses working with event epicenters and using b = 1, d = 1.6, and p = 0.5. *Zaliapin and Ben-Zion* [2013a] have shown that the cluster analysis is stable with respect to these parameters. In particular, none of the qualitative findings of this study will change if any of the parameters will fluctuate within reasonable limits.

1. Artifacts of earthquake location errors and short-term incompleteness on seismicity clusters in southern California

We document and quantify effects of two types of catalog uncertainties – earthquake location errors and short-term incompleteness – on results of statistical cluster analyses of seismicity in southern California. In the main part of the study we analyze 117,076 events with m ≥ 2 in southern California during 1981-2013 from the waveform-relocated catalog of *Hauksson et al.* [2013]. We present statistical evidence for three artifacts caused by the absolute and relative location errors: (1) Increased distance between offspring and parents. (2) Underestimated clustering, quantified by the number of offspring per event, the total number of clustered events, and some other statistics. (3) Overestimated background rates. We also find that short-term incompleteness leads to (4) Apparent magnitude dependence and temporal fluctuations of b-values. The reported artifacts are robustly observed in three additional catalogs of southern California: the relocated catalog of Richards-Dinger and Shearer [2000] during 1975-1998, and the two sub-catalogs – 1961-1981 and 1981-2013 – of the Advances National Seismic System. This implies that the reported artifacts are not specific to a particular (re)location method. The comparative quality of the four examined catalogs is reflected in the magnitude of the artifacts. The location errors in the examined catalogs mostly affect events with m<3.5, while for larger magnitudes the location error effects are negligible. This is explained by comparing the location error and rupture lengths of events and their parents. Finally, our analysis suggests that selected aggregated cluster statistics (e.g., proportion of singles) are less prone to location artifacts than individual statistics like the distance to parent or parent-offspring assignment. The results can inform a range of studies focused on small-magnitude seismicity patterns in the presence of catalog uncertainties. **Figure 1** shows the distribution of the distance *r* to parent (in km) for the parent-offspring pairs with m < 3.5 in the San Jacinto fault zone for the four catalogs. The distribution uniformly shifts from lower to larger distances as we go from HYS catalog to Dinger-Shearer to ANSS-2 to ANSS-1. The increase of the distribution median between each pair of catalogs is about one third of the magnitude order; accordingly, the difference between the median of the best quality catalog (HYS) and the worst quality catalog (ANSS-1) is about an order of magnitude – from 0.2km to 2km.

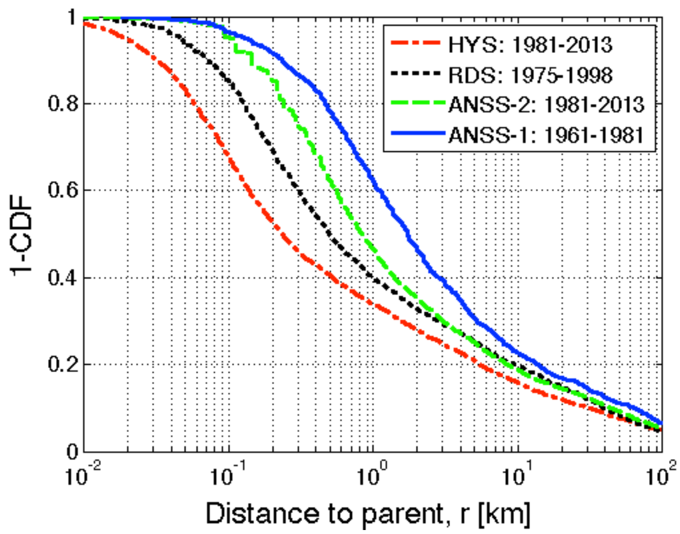

Figure 1: Comparative analysis of the distance to parent within the San Jacinto region in the four catalogs of southern California (HYS, DS, ANSS-1, ANSS-2). Here we only consider parent-offspring pairs with magnitudes below 3.5. The distance uniformly increases with the decreasing event location quality in the following catalog order: HYS (best quality, shortest distance), DS, ANSS-2, ANSS-1 (worst quality, longest distance).

2. Discriminating characteristics of tectonic and human-induced seismicity

We analyze statistical features of background and clustered subpopulations of earthquakes in different regions in an effort to distinguish between human-induced and natural seismicity. Analysis of "end-member" areas known to be dominated by human-induced earthquakes (the Geyser geothermal field in northern California and TauTona gold mine in South Africa) and regular tectonic activity (the San Jacinto fault zone in southern California and Coso region excluding the Coso geothermal field in eastern central California) reveals several distinguishing characteristics. Induced seismicity is shown to have (i) higher rate of background events (both absolute and relative to the total rate), (ii) faster temporal offspring decay, (iii) higher rate of repeating events, (iv) larger proportion of small clusters, and (v) larger spatial separation between parent and offspring, compared to regular tectonic activity. The temporal decay of the offspring in the six examined regions is compared in **Figure 2**; the figure shows the estimated density of close offspring as a function of time after the parent. This analysis suggests that the offspring in areas of induced seismicity tend to decay much faster than in tectonic areas. The temporal decay of the offspring intensity $\Lambda(t)$ in all examined regions is closely approximated by a power law $\Lambda(t) \propto t^{-h}$,. The power exponent changes from $h \approx 2$ in induced areas to $h \approx 1.5$ in the mixed regions and to $h \approx 1$ in tectonic regions. The reported differences also successfully discriminate seismicity within the Coso and Salton Sea geothermal fields in California before and after the expansion of geothermal production during the 1980s. The transition from tectonic to human-induced earthquakes in the Coso and Salton Sea geothermal fields leads to an increase in both the proportion of the background events and the absolute intensity of the background events, as well as to more rapid temporal offspring decay. These observations are further confirmed by **Figure 3** that compares the distribution of the rescaled time T to parent in the Coso and Salton Sea geothermal fields prior to and after the expansion of the geothermal production. Both regions exhibit a clear transition from a unimodal distribution of T with slow temporal decay in clusters and undeveloped background mode to a bimodal distribution with a clear separation between clustered and background (having many repeaters) modes and a fast temporal offspring decay in clusters. The same transition is reported for tectonic vs. induced regions. **Figure 4** illustrates and compares cluster style in all six examined regions.

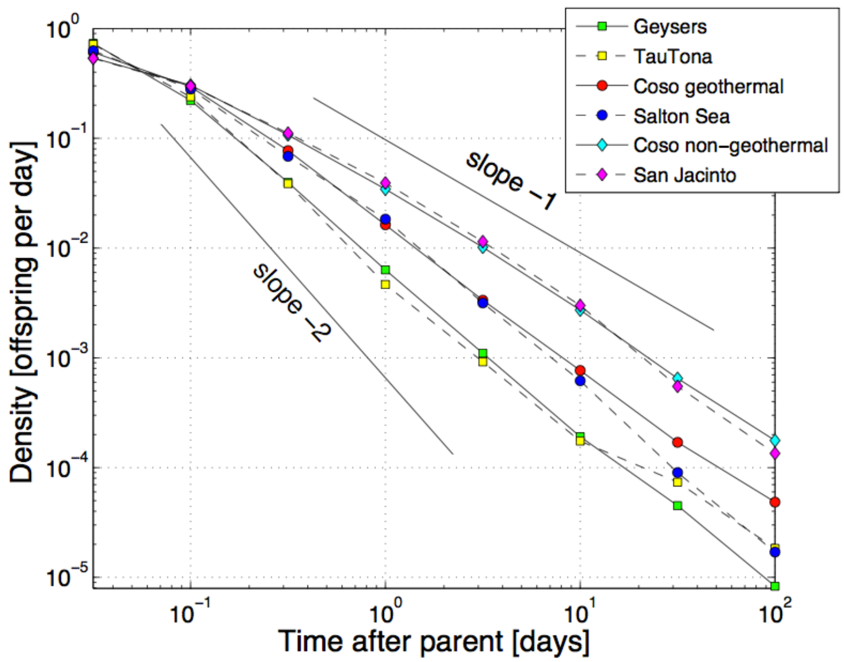

Figure 2: Temporal decay of the offspring intensity. Normalized intensity (average number of offspring per parent per day scaled to integrate to unity) as a function of time t after the parent. Analysis uses all close offspring (η < η0).

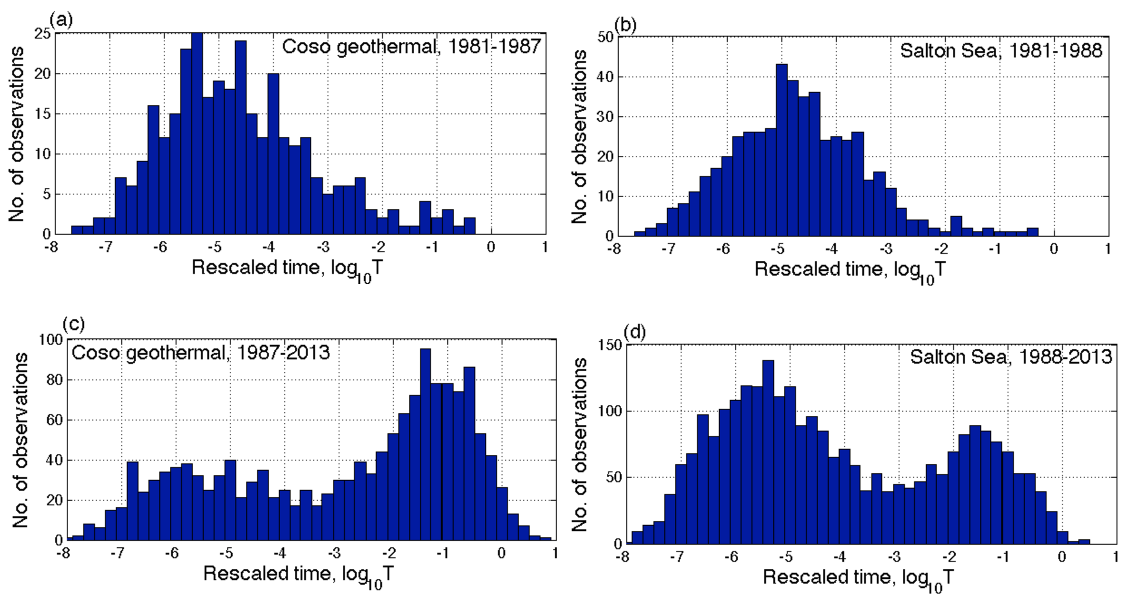

Figure 3: Change of clustering style in Coso and Salton Sea geothermal fields after beginning of active geothermal production. Figure shows the distribution of the rescaled time to parent T for the offspring within one parent rupture length from the parent. (a,c) Coso geothermal field; production began in 1987. (b,d) Salton Sea geothermal field; production began during 1988-1992. (a,b) Offspring before geothermal production. (c,d) Offspring during geothermal production.

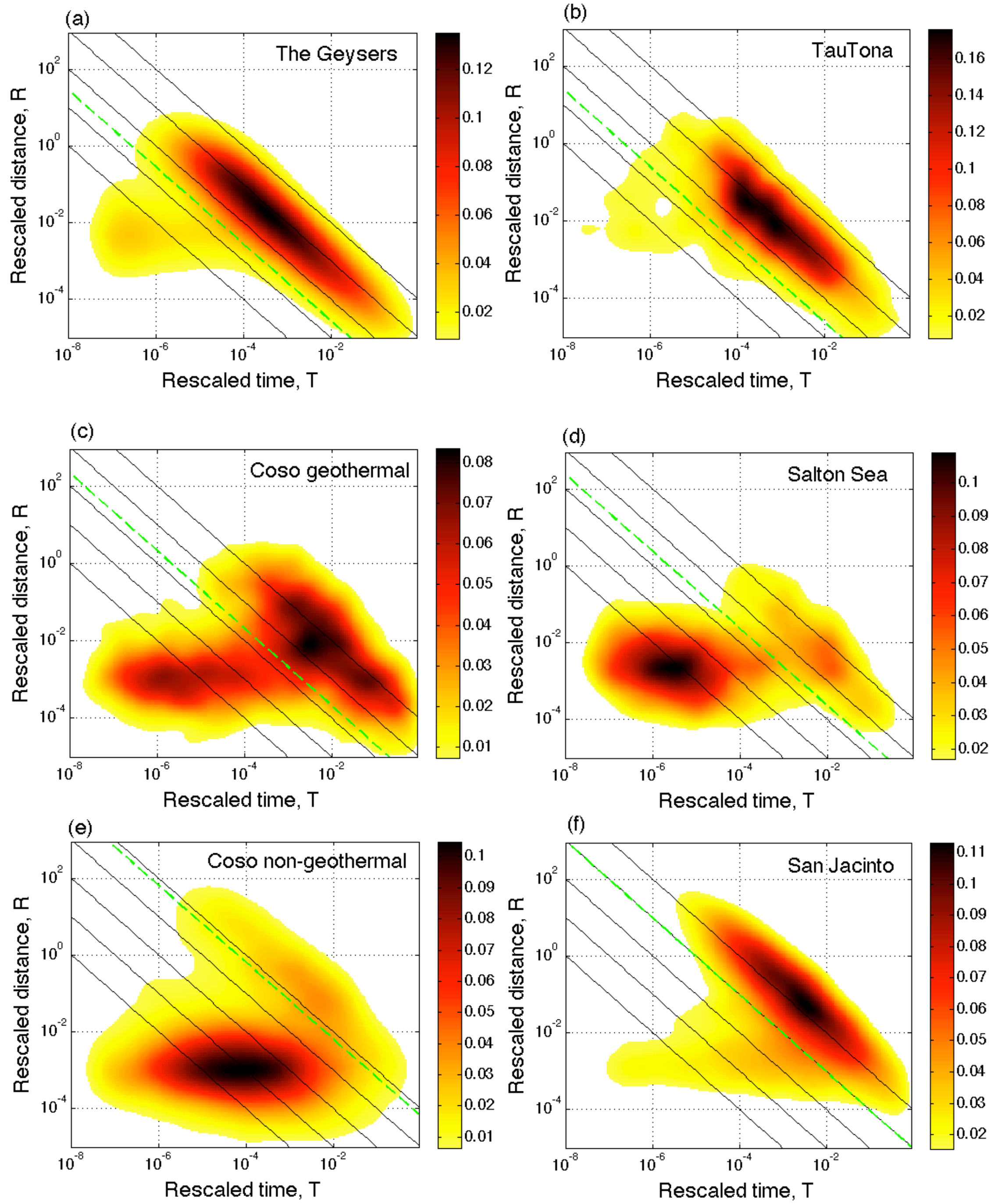

Figure 4: Clustering style of seismicity in the six examined regions. Each panel shows the joint 2-D distribution of the rescaled time T and distance R to the parent in a selected region. In TauTona we only show events that happened during midnight – 1PM, when the mining activity is minimal. The solid diagonal lines are the same in all panels and correspond (from bottom to top) to: $\eta = 10^{-8}$, 10^{-7}, 10^{-6}, 10^{-5}, and 10^{-4}. The dashed diagonal line depicts the mode separation threshold η_0. The sidebar indicates the density values. For visual convenience we cut the lower 5% of each distribution (transparent background).

B. References

Baiesi, M and M. Paczuski, 2004, Scale-free networks of earthquakes and aftershocks. *Phys. Rev. E*, 69, 066106.

Gu, C., A. Y. Schumann, M. Baiesi, and J. Davidsen (2013). Triggering cascades and statistical properties of aftershocks. *J. Geophys. Res.: Solid Earth*, 118(8), 4278-4295.

Hauksson, E., W. Yang, and P. M. Shearer (2012), Waveform Relocated Earthquake Catalog for Southern California (1981 to June 2011); *Bull. Seismol. Soc. Am.,* 102, 5, 2239-2244. doi: 10.1785/0120120010.

Hauksson, E., W. Yang, and P. M. Shearer (2013), Waveform Relocated Earthquake Catalog for Southern California, 1981-2013; available from http://www.data.scec.org/research-tools/downloads.html

Zaliapin, I., A. Gabrielov, H. Wong, and V. Keilis-Borok (2008). Clustering analysis of seismicity and aftershock identification, *Phys. Rev. Lett*., 101.

Zaliapin, I. and Y. Ben-Zion (2011). Asymmetric distribution of early aftershocks on large faults in California, *Geophys. J. Intl*., **185**, 1288-1304, doi: 10.1111/j.1365-246X.2011.04995.x.

Zaliapin, I., and Ben-Zion, Y. (2013a). Earthquake clusters in southern California I: Identification and stability. Journal of Geophysical Research: Solid Earth, 118(6), 2847-2864.

Zaliapin, I. and Y. Ben-Zion (2013b). Earthquake clusters in southern California II: Classification and relation to physical properties of the crust, *J. Geophys. Res*., 118(6), 2865-2877.