Report for a 2012 SCEC project

"Towards a unifying statistical framework for identification and analysis of earthquake clusters"

Principal Investigators:

Ilya Zaliapin
Department of Mathematics and Statistics
University of Nevada, Reno NV 89557
Phone: (775) 784-6077, Fax: (775) 784-6378
E-mail: zal@unr.edu

Yehuda Ben-Zion
Department of Earth Sciences
University of Southern California
Los Angeles, CA 90089
Phone: (213) 740-6734, Fax: (213) 740-8801
E-mail: benzion@usc.edu

Project period: February 1, 2012 – January 31, 2013

SUMMARY

- The project has developed and tested a robust method for comprehensive detection and analysis of earthquake clusters.
- The accuracy and stability of the method was demonstrated using the Epidemic Type Aftershock Sequence (ETAS) model with known cluster structure. Analysis of the ETAS model demonstrates that the cluster detection results are *accurate* and *stable* with respect to (i) three numerical parameters of the method, (ii) variations of the minimal reported magnitude, (iii) catalog incompleteness, and (iv) location errors.
- The method was applied to a 1981-2011 relocated seismicity catalog of southern California having 111,981 events with magnitudes $m \geq 2$, and corresponding synthetic catalogs produced by the Epidemic Type Aftershock Sequence (ETAS) model. Application of the method to the observed catalog separates the 111,981 examined earthquakes into 41,393 statistically significant *clusters* comprised of *foreshocks, mainshocks* and *aftershocks*.
- Systematic analysis allows us to detect several new features of seismicity that include (i) existence of a significant population of *single-event clusters*; (ii) existence of foreshock activity in natural seismicity that exceeds expectation based on the ETAS model; and (iii) dependence of all cluster properties, except area, on the *magnitude difference* of events from mainshocks but not on their absolute values.
- Furthermore, the project results (i) demonstrate that the clustering associated with the largest earthquakes, $m > 7$, is statistically different from that of small-to-medium earthquakes; (ii) establish the existence of two dominant types of small-to-medium magnitude earthquake families – *burst-like* and *swarm-like sequences* – and a variety of intermediate cluster forms obtained as a mixture of the two dominant types; (iii) suggest a simple new quantitative measure for identifying the cluster type based on its topological structure; (iv) demonstrate systematic spatial variability of the cluster characteristics on a scale of tens of kilometers in relation to heat flow and other properties governing the effective viscosity of a region; and (v) establish correlation between the family topological structure and a dozen of metric properties traditionally considered in the literature (number of aftershocks, duration, spatial properties, *b*-value, parameters of Omori-Utsu and Båth law, *etc*.). The burst-like clusters likely reflect highly-brittle failures in relatively cold regions, while the swarm-like clusters are likely associated with mixed brittle-ductile failures in regions with relatively high temperature and/or fluid content.
- The results of this project may be used to develop improved region-specific hazard estimates and earthquake forecasts.

TECHNICAL DESCRIPTION

INTRODUCTION

Earthquake clustering is an essential aspect of seismicity with signatures in space, time and size (e.g. magnitude, potency/moment, energy) domains that provide key information on earthquake dynamics. Clustering is the most prominent form of the existing variety of structures and patterns of seismicity, understood in the broadest sense as various deviations from a time-Stationary space-Inhomogeneous marked Poisson (SIP) process. Despite the overall agreement about the existence of multiple types of repeatedly observed seismic clusters, reflected by a well-developed cluster terminology, a formal definition of seismic clusters is lacking. This limits the ability of performing systematic cluster analysis. Even the most prominent type of earthquake clusters – aftershocks – does not have a commonly accepted definition. Accordingly, the existing cluster studies rely on various *ad-hoc* assumptions, which are well suited for addressing particular focused questions yet typically insufficient for general use. For the same reason, the majority of aftershock studies are associated with the largest earthquakes in a region. These events are characterized by extremely high intensity of aftershock series, at least in the mainshock vicinity, which allows one to accurately identify most aftershocks by a simple window approach and ensures that alternative methods lead to similar results. The behavior of aftershock sequences of small-to-medium magnitude events is largely unsettled.

This project developed a technique for detection of statistically significant clusters of arbitrary sizes and magnitudes. The employed cluster technique is characterized by (i) soft parameterization that uses only 3 easily-estimated parameters (the *b*-value of the magnitude distribution, the spatial dimension of epicenters, and the threshold that separates “close” and “distant” events), (ii) ability to uniformly analyze clusters associated with mainshocks of greatly different magnitude, (iii) demonstrated high stability of the cluster detection with respect to the employed parameters, minimal reported magnitude, catalog incompleteness, and location errors, and (iv) absence of underlying assumptions or governing models for the expected earthquake cluster structure. The combination of these properties distinguishes our technique from other existing algorithms [e.g., *Gardner and Knopoff*, 1974; *Reasenberg*, 1985; *Molchan and Dmitrieva*, 1992; *Zhuang et al.*, 2002; *Dzwinel et al. 2005; Marsan and Lengline*, 2008].

METHODOLOGY FOR CLUSTER DETECTION

The proposed methodology for cluster detection is described below, following *Zaliapin and Ben-Zion* [2013a].

Detection of clusters: The network of earthquakes

The main tool of the analysis is a *spanning network of earthquakes*, which is constructed by the nearest-neighbor approach using a particular earthquake distance in time-space-magnitude coordinates. Consider an earthquake catalog, where each event i is characterized by its occurrence time t_i, hypocenter (ϕ_i, λ_i, d_i), and magnitude m_i. The distance between earthquakes i and j is asymmetric in time and is defined as:

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

Here, $t_{ij} = t_j - t_i$ is the intercurrence time, $r_{ij} \geq 0$ is the spatial distance between the earthquake hypocenters, and d is the (possibly fractal) dimension of the space containing the earthquake hypocenters. For each earthquake j, we identify its nearest neighbor i and the corresponding nearest-neighbor distance η_{ij}. As a result, each earthquake has a single parent (the nearest earlier neighbor) and can be the parent for multiple children events. Hence, all earthquakes are connected in the nearest-neighbor time-oriented spanning tree.

Consider the space and time distances between the nearest neighbors normalized by the magnitude of the earlier event:

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

It is readily seen that $\log \eta_{ij} = \log T_{ij} + \log R_{ij}$. *Zaliapin et al.* [2008] and *Zaliapin and Ben-Zion* [2012a] demonstrated that a time-homogeneous, space-inhomogeneous Poisson process with exponential magnitudes corresponds to a unimodal joint distribution of (log T, log R) that is concentrated along a line log T + log R = const. Significantly, catalogs of the observed seismicity exhibit a prominent bimodal joint distribution of (log T, log R): One mode corresponds to background events (similar to the Poisson process), while the other consists of a large subpopulation of events located considerably closer in time and space to their parents than expected in a Poisson process with no clustering (see **Figure 1**).

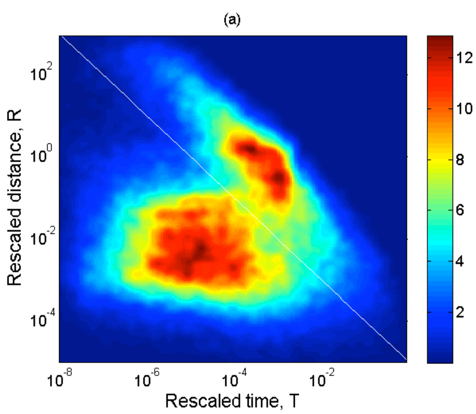
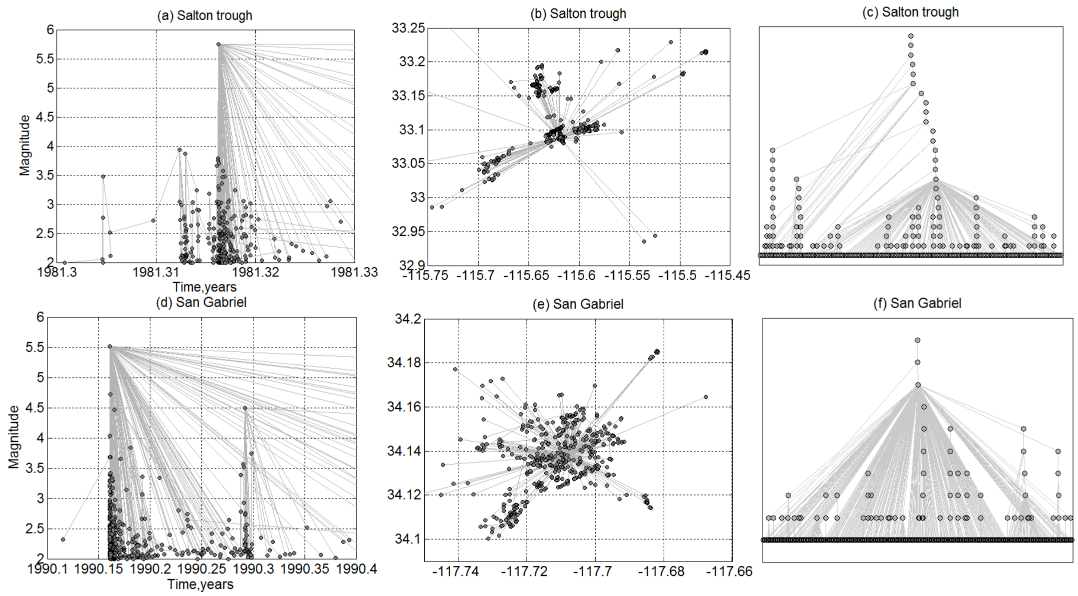


Figure 1. The joint distribution of (log T, log R) for earthquakes of $m \geq 2$ from the relocated catalog of *Hauksson et al.* [2012]. The white line visually depicts the separation between two seismic modes. The mode in the upper right part is expected in a Poisson process; it largely corresponds to background events. The mode in the left bottom part corresponds to events that occur closer to their parents than expected for a Poisson process; it largely corresponds to clusters, mainly to aftershocks and foreshocks.

The identification of individual clusters is done as follows: Each link in the spanning tree of earthquakes is assigned strength equal to the corresponding nearest-neighbor distance η. If one removes the "weak" links that correspond to events (and their parent links) belonging to the upper right background mode of the joint distribution shown in **Figure 1,** the spanning tree is decomposed into a set of trees. Each of these trees corresponds to a statistically significant cluster of earthquakes – all events within a cluster are now connected to their parents by "strong" links.

Stability of cluster detection: *Zaliapin and Ben-Zion* [2013a] used the observed seismicity and synthetic catalogs produced by the ETAS model to demonstrate that the detected clusters are stable with respect to (i) incompleteness of catalog, (ii) minimal reported magnitude, (iii) earthquake location errors, and (iv) three numerical parameter of the algorithm.



<u>Figure 2:</u> Two types of seismic clusters: an example. Circles correspond to earthquakes, lines to parent links. (a,b,c) A swarm-like family in Salton trough area – a region with decreased effective viscosity. (d,e,f) A burst-like family in San Gabriel area – a region with increased effective viscosity. (a,d) Magnitude vs. time. (b,e) Space map. (c,f) Topologic tree. The two families have largest event with similar magnitude but very different topologic structures (compare panels c and f).

Existence of clusters of distinct types: *Zaliapin and Ben-Zion* [2013a] demonstrated that the seismicity clusters in the southern California earthquake catalog of *Hauksson et al.* [2012] can be represented as a combination of clusters of three main types: (i) Singles – clusters consisting of one event; a significant part of singles cannot be explained by catalog artifacts (incompleteness, existence of the lowest reported magnitude, etc.) and may reflect certain crustal processes and conditions; (ii) Burst-like clusters characterized by small topological depth (small number of offspring generations), decreased foreshock and aftershock productivity, and smaller spatio-temporal extent, and (iii) Swarm-like clusters characterized by high topological depth (large number of offspring generations), increased foreshock and aftershock productivity, and higher spatio-temporal extent. Examples of bust-like and swarm-like clusters of comparable magnitudes are shown in **Figure 2**. There is no sharp boundary between clusters of different types, especially for the clusters of small size. The cluster type can be quantified by a simple scalar measure – average topological depth $<d>$ of the leaves in the tree that represents the cluster in the time-space-energy domain. Notably, the ETAS model can only reproduce well the burst-like clusters; the number of singles in the observed catalog is statistically higher than in the ETAS model with parameters estimated for southern California.

Association between cluster type and physical properties of a region: The clusters of different types have distinct preferred geographic locations. Burst-like clusters occur predominantly within areas with decreased levels of heat flow and fluid content and thicker seismogenic zone. Overall, these regions can be characterized as predominantly brittle with high effective viscosity. Swarm-like clusters occur predominantly in the areas with increased levels of heat flow and fluid content, high geothermal activity, and thinner seismogenic zone.

Overall, these regions are likely associated with mixed brittle-ductile rheology having decreased level of effective viscosity. Our observational findings are consistent with basic theoretical results of *Ben-Zion and Lyakhovsky* [2006] with a visco-elastic continuum damage rheology on the interplay between brittle and ductile failure mechanisms in the crust. An example of persistent spatial occurrence is illustrated in **Figure 3** that shows spatial distribution of the average leaf depth – the main parameter used for the definition of the cluster type. *Zaliapin and Ben-Zion* [2013b] have shown that the average levels of over a dozen of cluster statistics, new and traditionally considered in cluster studies, have statistically significant difference between the test regions with properties corresponding to relatively high (Ventura, San Bernardino, San Gabriel, Mojave) and relatively low (Coso and Salton trough) effective viscosity of the crust (see **Figure 3** for region definition).

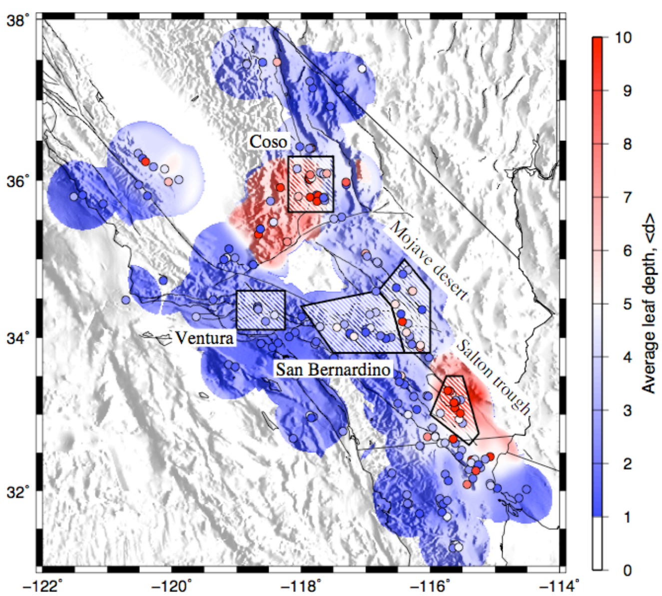


Figure 3: Spatial distribution of clusters of different types in southern California (magnitude $m > 4$, size $N > 10$). The cluster type is quantified using a scalar measure *average topologic leaf depth* <d>. Large depth corresponds to swarm-like clusters, small depth – to burst-like clusters. The figure shows five special study regions of *Zaliapin and Ben-Zion* [2013b]. Regions with decreased level of effective viscosity (Coso and Salton trough) are characterized by predominantly swarm-like clusters. Regions with increased level of effective viscosity (Ventura, San Bernardino, San Gabriel, Mojave) – have predominantly burst-like clusters.

Correlation between the cluster type and traditional cluster statistics: The cluster type, measured by the average leaf depth <d> of the respective topological tree was shown to have statistically significant association with over a dozen of metric cluster statistics traditionally considered in the aftershock studies [*Zaliapin and Ben-Zion*, 2013a,b]: number of fore/aftershocks, relative moment of fore/aftershocks compared to the moment of the mainshock, parameters of Omori and Bath laws, area of fore/aftershocks, *etc.* At the same time, it has the best spatial concentration (see **Figure 3**) and hence has been chosen as the master parameter in quantifying the cluster type.

Project publications:

Zaliapin, I. and Y. Ben-Zion (2013a) Earthquake clusters in southern California, I: Identification and stability. In review.

Zaliapin, I. and Y. Ben-Zion (2013b) Earthquake clusters in southern California, II: Classification and relation to physical properties of lithosphere. In review.

References

Ben-Zion, Y. (2008) Collective behavior of earthquakes and faults: continuum-discrete transitions, progressive evolutionary changes and different dynamic regimes, Rev. Geophys., 46, RG4006, doi:10.1029/2008RG000260.

Ben-Zion, Y. and V. Lyakhovsky (2006) Analysis of aftershocks in a lithospheric model with seismogenic zone governed by damage rheology, Geophys. J. Intl., 165, 197-210.

Dzwinel, W., D. A. Yuen, K. Boryczko, Y. Ben-Zion, S. Yoshioka and T. Ito (2005). Cluster Analysis, Data-Mining, Multi-dimensional Visualization of Earthquakes over Space, Time and Feature Space, *Nonlinear Proc. in Geophys.*, 12, 117-128.

Enescu, B., S. Hainzl and Y. Ben-Zion (2009) Correlations of Seismicity Patterns in Southern California with Surface Heat Flow Data, Bull. Seism, Soc. Am., **99**, doi: 10.1785/0120080038.

Gardner, J. K. and L. Knopoff (1974) Is the sequence of earthquakes in Southern California, with aftershocks removed, Possionian?, *Bull. Seis. Soc. Am.,* 64(5) 1363-1367.

Hauksson, E. (2011) Crustal geophysics and seismicity in southern California. *Geophysical Journal International,* 186: 82–98. doi: 10.1111/j.1365-246X.2011.05042.x

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.

Holtkamp, S. G., M. E. Pritchard and R. B. Lohman (2011) Earthquake swarms in South America. Geophys. J. Intl, doi: 10.1111/j.1365-246X.2011.05137.x

Marsan, D. and O. Lengline (2008) Extending Earthquakes' Reach Through Cascading, *Science*, 319(5866), 1076-1079, doi:10.1126/science.1148783.

Molchan, G. and O. Dmitrieva (1992) Aftershock identification – Methods and new approaches. *Geophys. J. Intl.,* 109 (3), 501-516.

Reasenberg, P. (1985), Second-order moment of central California seismicity, 1969-82, *J. Geophys. Res.*, 90, 5479-5495.

Shearer, P. M. (2012) Self-similar earthquake triggering, Bath's law, and foreshock/aftershock magnitudes: Simulations, theory, and results for southern California. J. Geophys. Res., 117, B06310

Vidale, J.E., Boyle, K.L. & Shearer, P.M. (2006) Crustal earthquake bursts in California and Japan: their patterns and relation to volcanoes, Geophys. Res. Lett., 33(20), L20313, doi:10.1029/2006GL027723.

Vidale, J.E. & Shearer, P.M. (2006) A survey of 71 earthquake bursts across southern California: Exploring the role of pore fluid pressure fluctuations and aseismic slip as drivers, J. Geophys. Res., 111(B5), B05312, doi:10.1029/2005JB004034.

Zaliapin, I., A. Gabrielov, V. Keilis-Borok, and H. Wong (2008) Clustering analysis of seismicity and aftershock identification. Phys. Rev. Lett., 101, 018501. doi: 10.1103/PhysRevLett.101.018501

Zaliapin, I. and Y. Ben-Zion (2013a) Earthquake clusters in southern California, I: Identification and stability. In review.

Zaliapin, I. and Y. Ben-Zion (2013b) Earthquake clusters in southern California, II: Classification and relation to physical properties of lithosphere. In review.

Zhuang, J.,Y. Ogata and D. Vere-Jones (2002). Stochastic declustering of space-time earthquake occurrences. *J. Amer. Stat. Assoc.*, 97: 369-380.