Earthquake clustering and localization of seismicity before large events

Ilya Zaliapin[1] and Yehuda Ben-Zion[2]

[1] Department of Mathematics and Statistics, University of Nevada, Reno
[2] Department of Earth Sciences and Southern California Earthquake Center, University of Southern California

Summary

We present recent results on assessing the degree of **regional clustering of earthquakes** and **progressive localization of seismicity** in relation to the earthquake cycle. First, we demonstrate that events included in the existing short-duration instrumental catalogs are concentrated within a very small fraction of the space-time volume, which is highly amplified by activity associated with the largest recorded events (**Fig. 1**). The earthquakes that are included in instrumental catalogs are unlikely to be fully representative of the long-term behavior of regional seismicity, creating a bias in a range of seismicity analyses. The results justify earthquake declustering and suggest a new robust metric for assessing quality of declustering independent of a particular declustering technique (**Figs. 2,3,4,5**). The second part focuses on progressive localization of seismicity, which corresponds to mechanical evolution of deformation from distributed failures in a rock volume to localized shear zones, culminating in generation of primary slip zones and large earthquakes. We present a methodology for estimation of localization using earthquake catalogs (**Figs. 6,7**) and showcase its applications to tracking preparation processes of large earthquakes (**Figs 8,9**). This analysis is performed with declustered catalogs. Methodologically, both discussed topics are based on the Receiver Operating Characteristic (ROC) analysis of seismicity (**Figs. 2,3,4,5**). We demonstrate how this unified framework is adopted for diverse tasks, including assessment of coupled space-time clustering after controlling for space and time marginal inhomogeneities of earthquake rates (**Figs. 3,4**), and tracking time-dependent transformations of a highly inhomogeneous earthquake space distribution (**Figs. 7,8,9**) . The examined data include crustal seismicity in California, Alaska and other regions, synthetic catalogs of the ETAS model, and acoustic emission data of laboratory fracturing experiments.

Earthquake Clustering: ROC Diagram

A high degree of earthquake clustering is commonly masked by a strong space-time concentration that is not obvious in a visual catalog inspection. **Figure 1** shows that at least half of events in the southern California catalog by *Hauksson et al.* (2012, extended) is in the form of extremely dense clusters not easily discernible by eye. The majority of earthquakes occur within the immediate spatio-temporal vicinity of other events, and do not evenly sample the active seismogenic volume.

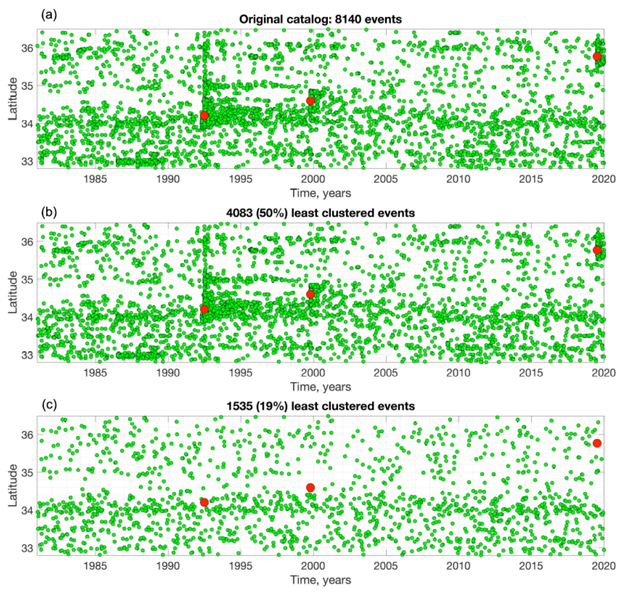
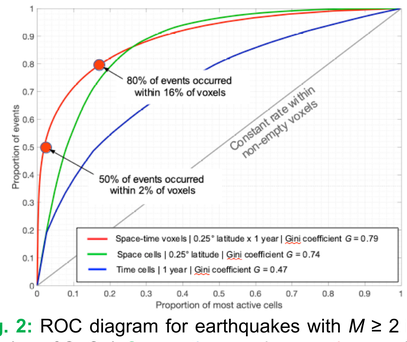


Fig. 1: Clustering of seismicity in SoCal. Time-latitude projection of earthquakes with magnitude $M \geq 3$ in the catalog of *Hauksson et al.* (2012, extended) during 1981-2020.



Receiver Operating Characteristic (ROC) diagram is quantifies the inhomogeneity of the space-time distribution of seismicity. We partition the examined space-time volume into voxels, count events in the voxels, and construct a diagram (**Fig. 2**) that shows the proportion of events in the most populated voxels (*y*-axis) vs. the proportion of such voxels (*x*-axis). We quantify the non-uniformity of the ROC diagram by the **Gini coefficient G** defined as twice the area between the diagram and the diagonal line. All realistic values of G are within the interval (0,1), where 0 corresponds to constant counts and 1 to an extreme concentration of all events within a single voxel.

Fig. 2: ROC diagram for earthquakes with $M \geq 2$ in *Hauksson et al.* (2012, extended) catalog of SoCal. Space, time, and space-time analyses.

Coupled Space-Time Clustering: Weighted ROC Diagram

To eliminate the effects of marginal space and time inhomogeneities on clustering and only quantify the coupled space-time variations, we introduce a **weighted ROC diagram**. Here, the *x*-axis is scaled in such a way that the product $J(x,t) = S(x)T(t)$ of the marginal space $S(x)$ and marginal time $T(t)$ rates of the estimated background seismicity corresponds to the diagonal. Formally, the *x*-axis in **Fig. 3** shows the proportion of the **factorized rate** $J(x,t)$ within the most active cells of the examined process, and the y-axis shows the respective proportion of events in the examined process. The marginal space and time inhomogeneities are reflected by $J(x,t)$ and are mapped onto the diagonal. Only **coupled space-time irregularities** cause deviation from the diagonal. The overall degree of the coupled space-time clustering is measured by the Gini coefficient G of this scaled ROC diagram. **Figure 4** illustrates a calculation of the factorized rate.

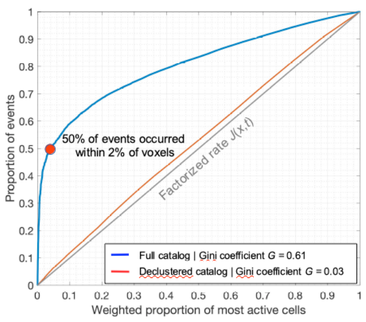
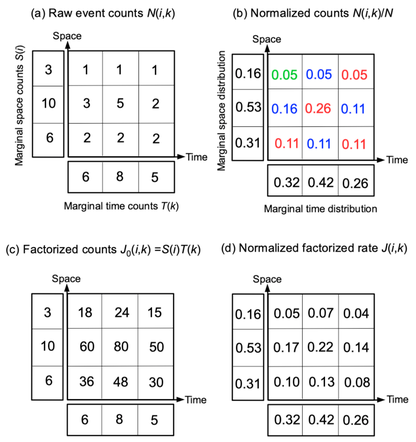


Fig. 3: Weighted ROC diagram for events with $M \geq 2$ in *Hauksson et al.* (2012, extended) catalog of southern California. Full catalog and declustered catalog.

Fig. 4: Factorized rate: Illustration



Coupled Clustering in Various Catalogs

Table 1: Summary of clustering analyses in Southern California, Western United States (WUS), Central and Eastern United States (CEUS), Alaska, Japan, global ISC-GEM catalog, and Epidemic Type Aftershock Sequence (ETAS) model.

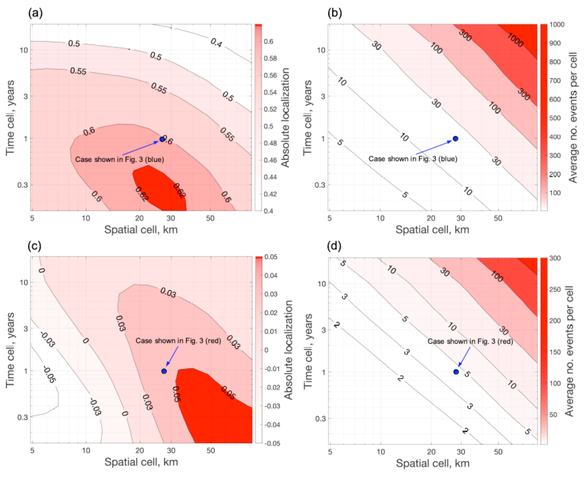
1	2	3	4	5	6	7	8	9	10	11
#	Catalog	Space range	Time range	Magnitude range	No. events	α_0*	No. background events†	Clustering w.r.t. constant rate‡, mean G (95% CI)	Clustering w.r.t. factorized rate $J(x,t)$‡, mean G (95% CI)	Background clustering w.r.t. factorized rate $J(x,t)$‡, mean G (95% CI)
1	HYS SoCal	119.5 – 115.5 W 32.8 – 36.5 N	1981 – 2020	2.0 – 7.3	89,341 (100%)	0	15,914 ± 110 (17.8 ± 0.1 %)	**0.77** (0.7, 0.82)	**0.55** (0.35, 0.63)	**0.031** (0, 0.06)
2	ComCat WUS	126 – 110 W 30 – 50 N	1981 – 2021	3.0 – 7.3	30,483 (100%)	0	8,494 ± 72 (27.9 ± 0.24 %)	**0.74** (0.59, 0.82)	**0.38** (0.2, 0.48)	**0.029** (0, 0.07)
3	ComCat CEUS	95 – 60 W 30 – 50 N	1981 – 2021	2.5 – 5.9	2,591 (100%)	0	1,172 ± 27 (45.2 ± 1.03 %)	**0.63** (0.51, 0.71)	**0.17** (0.07, 0.28)	**0.029** (0, 0.06)
4	ComCat Japan	138 – 146 W 32 – 44 N	1974 – 2021	4.0 – 9.1	17,823 (100%)	0	5,885 ± 59 (33.0 ± 0.33 %)	**0.67** (0.57, 0.72)	**0.39** (0.07, 0.51)	**0.044** (0, 0.09)
5	ComCat Alaska	180 – 120 E 45 – 75 N	1973 – 2020	4.0 – 8.0	15,008 (100%)	0	6,682 ± 56 (44.5 ± 0.37 %)	**0.71** (0.6, 0.81)	**0.18** (0.06, 0.29)	**0.002** (–0.02, 0.02)
6	ISC-GEM	World	1976 – 2017	5.8 – 9.3	7,066 (100%)	0	3,680 ± 41 (52.1 ± 0.58 %)	**0.64** (0.57, 0.72)	**0.11** (0.07, 0.16)	**0.006** (0, 0.03)
7	ETAS	600 × 600 km	20.9 years	2.5 – 7.3	26,986 (100%)	0	7,130 ± 73 (26.4 ± 0.27 %)	**0.66** (0.58, 0.71)	**0.05** (–0.09, 0.32)	**0.002** (–0.02, 0.03)

* Cluster threshold that controls the number of background events (*Zaliapin and Ben-Zion*, 2020).
† Mean and 95% confidence interval (CI) according to 10^4 realizations of stochastic declustering. The CI is estimated as ±1.96 sample standard deviation.
‡ Estimation is done for the space-time partitions with the average of at least 5 events per non-empty voxel and at least 300 voxels total. The CI limits are the 0.025 and 0.975 sample quantiles.

Notice that that the examined ETAS catalog has no substantial space-time coupling beyond that dictated by inhomogeneous background distribution. This is not the case in all examined observed catalogs.

Fig. 5: Stability of clustering measure G with respect to space and time resolutions of the ROC diagram.

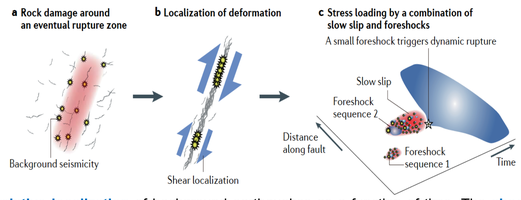
The catalog of *Hauksson et al.* (2012, extended) with magnitude $M \geq 2$ during 1981 – 2020. Gini coefficient *G* of the full catalog (panel a) and background earthquakes (panel c) in the ROC analysis with respect to the factorized rate $J(x,t)$ for different space (*x*-axis) and time (*y*-axis) resolutions. Average number of events per non-empty voxel in the experiment of panels (a,c) is shown in panels (b,d), respectively. Blue circles correspond to the cases illustrated in **Fig. 3**.



Localization Prior to Large Events

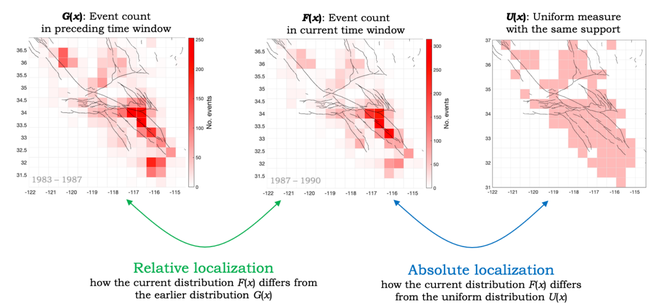
Progressive **localization of deformation** is a basic mechanical process that produces simultaneously reduced strength and increasing stress in a deforming rock volume (Kato & Ben-Zion, 2021). The localization framework describes the progressive evolution of deformation from distributed failures in a rock volume to localized shear zones, culminating in generation of primary slip zones and large earthquakes (**Fig. 6**).

Fig. 6: Localization concept



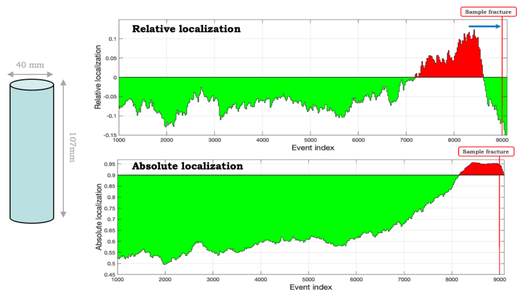
We quantify the **absolute and relative localization** of background earthquakes as a function of time. The **absolute localization** measures the "spikiness" of a measure, or the difference between a given spatial measure and the uniform measure with the same support. The **relative localization** of measure *P* with respect to measure *Q* reflects two simultaneous phenomena: (i) measure *P* is more localized in the absolute sense than *Q*, and (ii) both measures concentrate within the same spatial areas. This is done via ROC diagrams (Zaliapin and Ben-Zion, 2020).

Fig. 7: Absolute and relative localizations: Illustration



Localization in AE Data

The analysis is done in event window using cubic cells with linear size 10mm, windows $\omega_1 = \omega_2 = 500$ events, and threshold $P_0 = 1$. The sample has cylindrical shape with radius of 20mm and depth 107mm. Goebel et al., *GRL* (2013)



Positive values indicate that the current measure is a localized version of the earlier measure = it has the same support and co-located yet more prominent peaks

Negative values indicate that the current measure is less related to the earlier measure, e.g. has different support or peaks in different locations

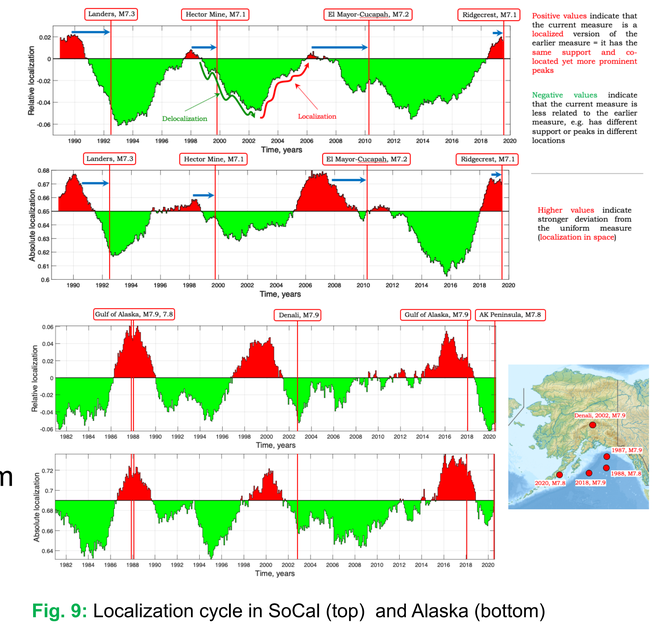
Higher values indicate stronger deviation from the uniform measure (localization in space)

Fig. 8: Localization in AE experiment of Goebel et al. (2013)

Localization in SoCal and Alaska

SoCal $M \geq 2$

Alaska $M \geq 4$ $z < 70$km



Positive values indicate that the current measure is a localized version of the earlier measure = it has the same support and co-located yet more prominent peaks

Negative values indicate that the current measure is less related to the earlier measure, e.g. has different support or peaks in different locations

Higher values indicate stronger deviation from the uniform measure (localization in space)

Fig. 9: Localization cycle in SoCal (top) and Alaska (bottom)

Acknowledgements

The study was supported by the National Science Foundation (grants EAR-1723033 and EAR-1722561) and the Southern California Earthquake Center project 21041.

References

Ben-Zion, Y. and Zaliapin, I. (2020) Localization and coalescence of seismicity before large earthquakes. GJI, 223(1), 561-583. https://doi.org/10.1093/gji/ggaa315
Goebel, T.H.W. et al. (2013) Acoustic emissions document stress changes over many seismic cycles in stick-slip experiments. GRL, 40(10), 2049–2054. https://doi.org/10.1002/grl.50507
Hauksson, E., Yang, W., and Shearer, P. M. (2012) Waveform relocated earthquake catalog for Southern California (1981 to June 2011). BSSA, 102(5), 2239–2244. https://doi.org/10.1785/0120120010
Kato, A. and Ben-Zion, Y. (2021) The generation of large earthquakes. Nature Reviews Earth & Environment, 2(1), 26-39. https://doi.org/10.1038/s43017-020-00108-w
Zaliapin, I. and Y. Ben-Zion (2020) Earthquake declustering using the nearest-neighbor approach in space-time-magnitude domain, JGR., 125, e2018JB017120. https://doi.org/10.1029/2018JB017120
Zaliapin, I. and Ben-Zion, Y. (2021) Perspectives on clustering and declustering of earthquakes, SRL, in press https://doi.org/10.1785/0220210127