

Clustering Analysis of Seismicity and Aftershock Identification

Ilya Zaliapin*

Department of Mathematics and Statistics, University of Nevada, Reno, Nevada 89557-0084, USA

Andrei Gabrielov†

Departments of Mathematics and Earth and Atmospheric Sciences, Purdue University, West Lafayette, Indiana 47907-1395, USA

Vladimir Keilis-Borok‡ and Henry Wong

Institute of Geophysics and Planetary Physics, and Department of Earth and Space Sciences, University of California Los Angeles, 3845 Slichter Hall, Los Angeles, California 90095-1567, USA

(Received 21 December 2007; published 30 June 2008)

We introduce a statistical methodology for clustering analysis of seismicity in the time-space-energy domain and use it to establish the existence of two statistically distinct populations of earthquakes: clustered and nonclustered. This result can be used, in particular, for nonparametric aftershock identification. The proposed approach expands the analysis of Baiesi and Paczuski [Phys. Rev. E **69**, 066106 (2004)] based on the space-time-magnitude nearest-neighbor distance η between earthquakes. We show that for a homogeneous Poisson marked point field with exponential marks, the distance η has the Weibull distribution, which bridges our results with classical correlation analysis for point fields. The joint 2D distribution of spatial and temporal components of η is used to identify the clustered part of a point field. The proposed technique is applied to several seismicity models and to the observed seismicity of southern California.

DOI: [10.1103/PhysRevLett.101.018501](https://doi.org/10.1103/PhysRevLett.101.018501)

PACS numbers: 91.30.Px, 02.50.-r, 91.30.Ab

Introduction.—Earthquake clustering is the most prominent feature of the observed seismicity. Well-studied types of clustering include aftershocks, foreshocks, pairs of large earthquakes, swarms, bursts of aftershocks, rise of seismic activity prior to a large regional earthquake, etc. Single clustering phenomena and their combination are essential elements of understanding the seismic stress redistribution and lithosphere dynamics [1], as well as constructing empirical earthquake prediction methods and evaluating regional seismic hazard [2].

Baiesi and Paczuski [3] have developed an elegant framework for earthquake clustering analysis by defining the pairwise earthquake distance η_{ij} as the expected number of events in the time-space domain bounded by events i and j . We expand here their approach to demonstrate the existence of two statistically distinct subpopulations in the observed seismicity of southern California: One corresponds to a “uniform” flow of events that is stationary in time but may be inhomogeneous in space, while the other corresponds to clustered events, which are much closer to each other in time and space than would be expected for the above “uniform” flow. The earthquakes from the clustered part, by and large, obey the conventional definitions of aftershocks [4]. Our analysis, therefore, provides an objective statistical foundation for developing a nonparametric aftershock identification procedure.

Our finding is supported by theoretical and numerical analyses of several seismicity models, including the epidemic type aftershock sequence (ETAS) model [5]. The main theoretical result is that for a stationary and homogeneous Poisson field with independent exponential magni-

tudes, the distance η has Weibull distribution, the same distribution as the Euclidean nearest-neighbor distance for a homogeneous point field. The proposed cluster detection technique is built upon the deviations of the observed nearest-neighbor distance η from this theoretical prediction. The key element of the applied analysis is the joint 2D distribution of spatial and temporal components of η ; this distribution clearly separates the clustered and nonclustered parts of a point field.

Distance between earthquakes.—Consider an earthquake catalog $\{t_i, \theta_i, \phi_i, m_i\}_{i=1, \dots, N}$, where each record i describes an individual earthquake with occurrence time t_i , position given by latitude θ_i and longitude ϕ_i , and magnitude m_i ; here, we do not consider the depth.

For any two earthquakes i and j we define the time-space-magnitude distance by

$$n_{ij} = \begin{cases} c\tau_{ij}r_{ij}^d 10^{-b(m_i - m_0)} & \tau_{ij} \geq 0 \\ \infty & \tau_{ij} < 0. \end{cases} \quad (1)$$

Here $\tau_{ij} = t_j - t_i$ is the earthquake interoccurrence time, r_{ij} the surface distance, d the (fractal) dimension of earthquake epicenters, and b the parameter of the Gutenberg-Richter relation (exponential fit to the magnitude distribution):

$$\mathbf{P}\{m > x\} = 10^{-b(x - m_0)} \quad \text{for } x \geq m_0. \quad (2)$$

By connecting each event with its nearest neighbor with respect to the distance n one obtains a time-oriented tree \mathbb{T} whose root is the first event in the catalog. Such trees were introduced and studied by Baiesi and Paczuski [3].

It is readily checked that the space-time volume of a ball of radius C , $B_C := \{(t, x, y, m) : n(t, r, m) < C\}$, is infinite due to heavy tails of the distance n in time when $d > 2$, in space when $d < 2$, and in both time and space for $d = 2$. Hence, any such ball almost surely contains an infinite number of events from N that prevents meaningful nearest-neighbor analysis. To avoid this, we introduce the truncated distance

$$\eta_{ij} = \begin{cases} n_{ij} & t_{ij} \leq t_0, r_{ij} \leq r_0 \\ \infty & \text{otherwise.} \end{cases} \quad (3)$$

Choosing t_0 and r_0 large enough will ensure that the measures η and n are equivalent within a bounded spatio-temporal area. The *nearest-neighbor* distance is defined as $\eta_j^* := \min_i \eta_{ij}$. We will drop the subindices ij or j unless it is important which pair of earthquakes is considered.

Main result: Poisson field.—Consider a spatiotemporal marked point field N with temporal component $t \in \mathbb{R}$, spatial component $\mathbf{x} \in \mathbb{R}^2$, and scalar marks m that represent the earthquake magnitude.

Assumption 1 (i) N is a stationary and homogeneous Poisson marked point field with intensity λ . (ii) Magnitude marks m_i are independent of the field (t_j, \mathbf{x}_j) and each other and have exponential distribution (2) with parameters \tilde{b}, m_0 . (iii) Let $f = b/\tilde{b}$, where b is the prior parameters of the Gutenberg-Richter law (2) used in (1).

Proposition 2 Under Assumption 1, the nearest-neighbor distance η^* has the following distribution, for large τ_0, r_0 :

$$P\{\eta^* < x\} = 1 - \exp\left[-\lambda\gamma\Psi\left(\frac{x}{\tau_0 r_0^d}\right)\right]. \quad (4)$$

Here γ is independent of x and we have

$$\Psi(w) \sim \begin{cases} w & d < 2, f < 1 \\ w \ln w & d = 2, f < 1 \\ w^{2/d} & d > 2, d > 2f \\ w^{2/d} \ln w & d > 2, d = 2f \\ w^{1/f} & d < 2f, f > 1 \\ w \ln w & d < 2, f = 1 \\ w(\ln w)^2 & d = 2, f = 1, \end{cases} \quad (5)$$

where $\Psi(w) \sim \psi(w)$ stays for $\lim_{w \rightarrow 0} \frac{\Psi(w)}{\psi(w)} = 1$.

The proof will be published elsewhere.

Proposition 2 implies that, for $b \neq \tilde{b}$, $d \neq 2$, and $d \neq 2f$, η^* has Weibull distribution. Furthermore, the distribution of η^* is independent of the magnitude threshold m_0 ; this facilitates analysis of data from different periods and regions that might have different m_0 .

We define the magnitude-normalized time and space components of η as

$$T_{ij} := \tau_{ij} 10^{-bm_i/2}, \quad R_{ij} := r_{ij}^d 10^{-bm_i/2}. \quad (6)$$

Obviously $\eta = TR$ (without loss of generality, we assumed here $c = 1$ and $m_0 = 0$) and Proposition 2 implies

that the distribution of the nearest-neighbor pair (T, R) is concentrated along the line $\log_{10} T + \log_{10} R = x_m$, where x_m is the mode of the distribution (4), while the level lines are of the form $\log_{10} T + \log_{10} R = \text{const}$. Figure 1(a) illustrates this by showing the empirical distribution of the pairs (T, X) for a Poisson homogeneous field with exponential magnitudes.

Modeled seismicity.—Here we analyze numerically the distribution of the nearest-neighbor distance η^* for three point field models: (i) homogeneous Poisson marked field, (ii) single self-excited aftershock series governed by Omori law, and (iii) ETAS model that combines the first two.

The ETAS model was introduced by Ogata [5]; it specifies a marked point process N by its conditional intensity at instant t and spatial location (x, y) :

$$\Lambda(t, x, y) = \Lambda_0 + \sum_{i: t_i < t} 10^{bm_i} \Lambda_T(\tau) \Lambda_R(r), \quad (7)$$

where $\Lambda_0 > 0$, $\tau = t - t_i$, $r^2 = (x - x_i)^2 + (y - y_i)^2$, and the temporal (Λ_T) and spatial (Λ_R) kernels are given by [5] $\Lambda_T(t) = (t + c)^{-1-\epsilon_T}$, $\Lambda_R(r) = (r + d)^{-1-\epsilon_R}$ with positive c, d, ϵ_T , and ϵ_R . Magnitudes are drawn independently from the exponential distribution (2).

A single aftershock series is a particular case of ETAS model with Λ_0 replaced by $\delta(0, 0, 0)$, which represents the main shock; its magnitude is a model parameter.

Figures 1 and 2 show the distributions of $\log_{10} \eta^*$ and corresponding pairs (T, R) . The Poisson model behaves as suggested by Proposition 2. For a single aftershock series, one observes almost symmetric (T, R) scatter, which suggests that time and space components of η are independent. Notably, the distribution of η^* in this case is close to log-normal (Fig. 2, middle panel). The ETAS distribution

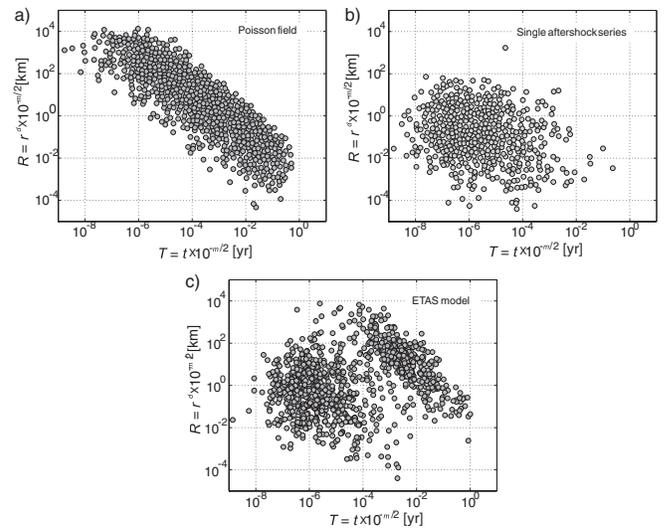


FIG. 1. Distribution of time and space components (T, R) , of the nearest-neighbor distance η^* for a 2D homogeneous Poisson field with exponential magnitudes (a), single aftershock series obeying Omori law (b), ETAS model (c). Here $b = 1$, $d = 2$.

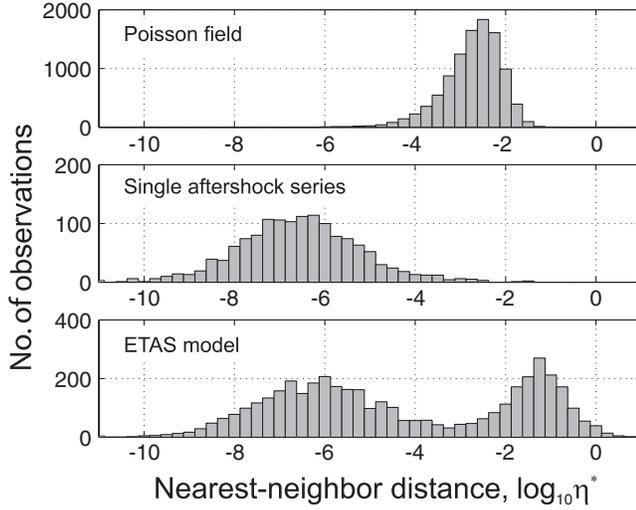


FIG. 2. Distribution of the logarithm of the nearest-neighbor distance $\log_{10}\eta^*$ for a 2D homogeneous Poisson field with exponential magnitudes (top), single aftershock series obeying Omori law (middle), ETAS model (bottom). Here, $b = 1$, $d = 2$.

has two prominent “modes”: A scatter along $TR = \text{const}$ in the upper right part of the plot [Fig. 1(c)] and an apparently independent scatter closer to the origin. Evidently, combining the homogeneous Poisson flow and aftershock clustering we have combined as well the corresponding modes of the (T, R) distributions.

Observed seismicity: Southern California.—We use a southern California earthquake catalog produced by the Advance National Seismic System [6], and consider earthquakes with magnitude $m \geq 2.0$ within a rectangular region bounded by 122°W , 114°W , 32°N , 37°N during January 1, 1984–December 31, 2004, the same as in [3].

The empirical distributions of the logarithm of the nearest-neighbor distance, $\log_{10}\eta^*$, and its components (T, R) are shown in Figs. 3 and 4. The distributions are prominently bimodal, revealing the existence of two statistically distinct earthquake populations. One of them corresponds to $\log_{10}T + \log_{10}R = \text{const}$; according to Proposition 2 it describes stationary (but possibly space-inhomogeneous) Poisson seismicity. The other population corresponds to $\log_{10}R \approx -2$; it corresponds mainly to the aftershock clustering.

To detect individual clustered events, we fix a threshold η_0 and remove all the links with $\eta_j^* > \eta_0$ from the tree \mathbb{T} . This will result in the forest (set of trees) $\mathbb{F}(\eta_0) = \{\mathbb{T}_i\}_{i=1}^{N(\eta_0)}$. Each tree \mathbb{T}_i in the forest corresponds to a single earthquake cluster; these clusters can be further analyzed in order to solve a particular applied problem. For example, aftershocks are often assumed to have smaller magnitude than the corresponding main shocks [4]. Possible earthquake clusters observed prior to the main shock are then called *foreshocks*. In this situation, it is natural to define the i th *main shock* as the largest earthquake within the tree \mathbb{T}_i ,

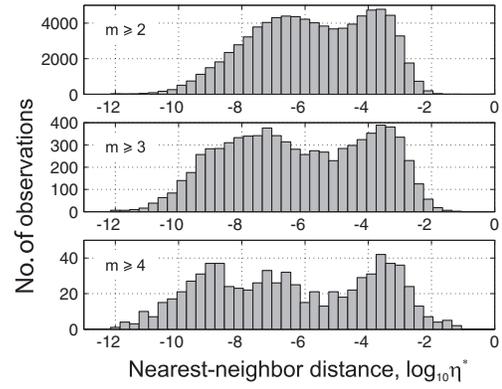


FIG. 3. Bimodal distribution of $\log_{10}\eta^*$ for the observed seismicity of southern California during 1984–2004; different panels correspond to different lower magnitude cutoffs. Here $b = 1$ and $d = 1.6$ as in [3]. Notice the bimodal structure with the two modes separated at $\eta \approx 10^{-5}$.

and *aftershocks (foreshocks)* as the events from \mathbb{T}_i that occurred later than (prior to) the main shock. The results of this aftershock-detection procedure in California are shown in Fig. 5; here we used $\eta_0 = 10^{-5}$ suggested by the distribution of η^* and (T, X) (Figs. 3 and 4). The figure focuses on the Landers earthquake, the largest one in California during the considered period. The three groups of earthquakes are identified as aftershocks: (a) the prominent earthquake cluster in the immediate vicinity of the

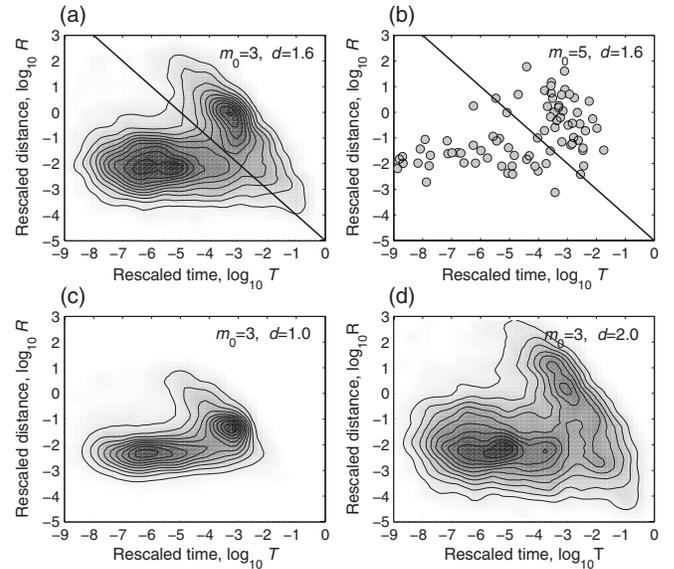


FIG. 4. Bimodal distribution of time and space components (T, R) of the nearest-neighbor distance η^* for the observed seismicity of southern California during 1984–2004. Solid line in (a) and (b) corresponds to $\log_{10}T + \log_{10}R = -5$. Different panels correspond to different values of m_0 and d : (a) $m_0 = 3.0$, $d = 1.6$, (b) $m_0 = 5.0$, $d = 1.6$, (c) $m_0 = 3.0$, $d = 1.0$, (d) $m_0 = 3.0$, $d = 2.0$. (a), (c), and (d) show smoothed density of points; panel (b) shows original points.

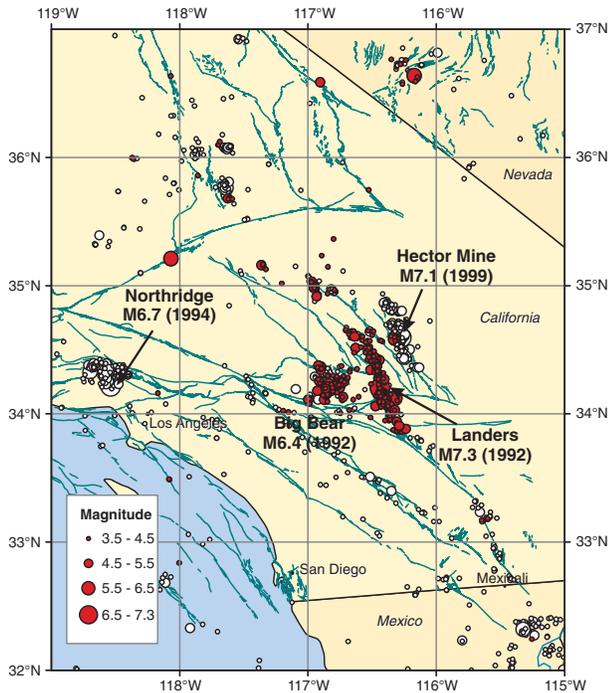


FIG. 5 (color online). Aftershock identification for Landers earthquake (June 28, 1992, $M=7.3$). The figure shows all earthquakes that occurred after the Landers earthquake. Shaded circles mark earthquakes identified as Landers' aftershocks; open circles mark the rest of the earthquakes.

Landers' epicenter, (b) the “secondary” aftershocks after the Big Bear earthquake, $M = 6.4$, which itself is the largest aftershock of Landers, and (c) several “distant aftershocks” that occurred immediately after Landers but at large distance from the latter. Both Northridge and Hector Mine aftershock clusters have not been associated with Landers. We emphasize though the existence of a distant Landers' aftershock close to the future epicenter of Hector Mine.

Conclusion and discussion.—We demonstrated the existence of statistically distinct clustered and nonclustered parts in the observed seismicity of southern California. The result, based on the bimodal joint distribution of the time and space components of the nearest-neighbor distance η , is robust with respect to the implied spatial dimension d of epicenters [Figs. 4(a), 4(c), and 4(d)] and the catalog's lower magnitude threshold m_0 [Figs. 4(a) and 4(b)]. This finding has important implications for various problems, aftershock detection being the most prominent one. The physical interpretation of the reported separation as well as its further applications will be considered in a forthcoming paper.

An analog of Proposition 2 is readily proven for any nearest-neighbor distance that depends multiplicatively on spatiotemporal point location and multidimensional mark \mathbf{m} : $\eta = \tau r^d f(\mathbf{m})$. It would be interesting to see how alternative definitions of η will alter the proposed clustering analysis. A noteworthy modification is to replace the term $r^d 10^{-bm}$ in the nearest-neighbor distance definition with the observed number of earthquakes of magnitude m in the respective spatial domain over a large time period; this will reflect the realistic regional fault geometry and possible spatial fluctuations of the magnitude distribution and may lead to better cluster detection.

This study was partly supported by NSF Grant No. ATM-0620838 and the Southern California Earthquake Center (SCEC). SCEC is funded by NSF Cooperative Agreement No. EAR-0106924 and USGS Cooperative Agreement No. 02HQAG0008. The SCEC contribution number for this Letter is 1137.

*Corresponding author.

zal@unr.edu

†agabriel@math.purdue.edu

‡vkb@ess.ucla.edu

- [1] B. Romanowicz, *Science* **260**, 1923 (1993); J. Dieterich, *J. Geophys. Res.* **99**, 2601 (1994); F. Press and C. Allen, *J. Geophys. Res.* **100**, 6421 (1995); K. R. Felzer and E. E. Brodsky, *Nature (London)* **441**, 735 (2006); G. C. P. King and D. D. Bowman, *J. Geophys. Res.* **108**, 2096 (2003).
- [2] *Geocomplexity and the Physics of Earthquakes*, edited by J. Rundle, D. Turcotte, and W. Klein (American Geophysical Union, Washington, DC, 2000); *Nonlinear Dynamics of the Lithosphere and Earthquake Prediction.*, edited by V. I. Keilis-Borok and A. A. Soloviev (Springer, Heidelberg, 2003); D. Sornette, *Critical Phenomena in Natural Sciences* (Springer-Verlag, Heidelberg, 2004), 2nd ed.
- [3] M. Baiesi and M. Paczuski, *Phys. Rev. E* **69**, 066106 (2004); M. Baiesi and M. Paczuski, *Nonlin. Proc. Geophys.* **12**, 1 (2005); M. Baiesi, *Physica (Amsterdam)* **360A**, 534 (2006).
- [4] L. Knopoff and J. K. Gardner, *Geophys. J. R. Astron. Soc.* **28**, 311313 (1972); V. I. Keilis-Borok, L. Knopoff, and I. M. Rotwain, *Nature (London)* **283**, 259 (1980); P. Reasenber, *J. Geophys. Res.* **90**, 5479 (1985); S. D. Davis and C. Frohlich, *Geophys. J. Int.* **104**, 289 (1991); G. M. Molchan and O. E. Dmitrieva, *ibid.* **109**, 501 (1992); J. Zhuang, Y. Ogata, and D. Vere-Jones, *J. Am. Stat. Assoc.* **97**, 369 (2002).
- [5] Y. Ogata, *Ann. Inst. Stat. Math.* **50**, 379 (1998).
- [6] Available at <http://quake.geo.berkeley.edu/anss/>