Aftershock identification

Ilya Zaliapin*
Department of Mathematics and Statistics, University of Nevada, Reno, NV 89557-0084, Corresponding author.

Andrei Gabrielov†
Departments of Mathematics and Earth and Atmospheric Sciences, Purdue University, West Lafayette, IN, 47907-1395

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, CA 90095-1567

(Dated: December 20, 2007)

Earthquake aftershock identification is closely related to the question "Are aftershocks different from the rest of earthquakes?" We give a positive answer to this question and introduce a general statistical procedure for clustering analysis of seismicity that can be used, in particular, for aftershock detection. The proposed approach expands the analysis of Baiesi and Paczuski [PRE, 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 Weibull distribution, which bridges our results with classical correlation analysis for unmarked point fields. We introduce a 2D distribution of spatial and temporal components of η, which allows us to identify the clustered part of a point field. The proposed technique is applied to several synthetic seismicity models and to the observed seismicity of Southern California.

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

INTRODUCTION

Earthquake clustering is the most prominent feature of the observed seismicity. The centennial world-wide observations have revealed a wide variety of clustering phenomena that unfold in the time-space-magnitude domain (magnitude being the logarithmic measure of earthquake energy) and provide the most reliable and useful information about the essential properties of earthquake flow. 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, switching of the global seismic activity between different parts of the Earth, *etc.* Single clustering phenomena and their combination are an essential element 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 studying earthquake clustering by defining the pairwise earthquake distance η_{ij} via the expected number of events in a particular time-space-magnitude domain bounded by events i and j. These authors used the distance η_{ij} to develop a tree-based statistical technique for earthquake cluster analysis and established several scaling laws for the observed earthquake clusters.

We expand here the approach of Baiesi and Paczuski [3] to demonstrate the existence of two statistically distinct subpopulations in the observed seismicity of Southern California: One corresponds to a uniform, absolutely random flow of events while another to earthquake clustering. The earthquakes from the clustering part, by and large, obey the conventional definitions of aftershocks [4]. Our analysis, therefore, provides an objective statistical foundation for aftershock identification that requires no prior clustering parameters like space-time windows traditionally used for aftershock detection [4].

Our finding is supported by theoretical and numerical analyses of several seismicity models, including ETAS [5]. The main theoretical result is that for a homogeneous spatio-temporal Poisson field with independent exponential magnitudes, 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 build upon the deviations of the observed nearest-neighbor distance η from this theoretical prediction. The key element of the applied analysis is the 2D distribution of spatial and temporal components of η; this distribution clearly separates the clustered and non-clustered parts of a point field.

DISTANCE BETWEEN EARTHQUAKES

Consider an earthquake catalog $\{t_i,\, \theta_i,\, \phi_i, m_i\}_{i=1,\dots,N}$. 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

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

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

$$\mathsf{P}\{m > x\} = 10^{-b\,(x-m_0)} I_{\{x>m_0\}}. \qquad (2)$$

Connecting each event with its nearest neighbor with respect to the distance n one obtains a time-oriented tree $\mathcal{T}$ 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 in metric n, $B_C := \{(t,x,y,m) : n(t,r,m) < C\}$, 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

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

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 spatio-temporal 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. Below we formulate our main result, starting with essential assumptions about the field N.

Assumption 1 (i) N is a 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 $\tilde{b}$, $\tilde{m}_0$. (iii) Let $f = b/\tilde{b}$ and $\mu_0 = 10^{\tilde{b}(\tilde{m}_0 - m_0)}$ where b and m_0 are the prior parameters of the Gutenberg-Richter law (2) used in (1).

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

$$\mathsf{P}\{\eta_j^* < x\} = 1 - \exp\left(-\lambda\gamma\,\Psi\left(\frac{x}{\tau_0\, r_0^d\, \mu_0^f}\right)\right). \qquad (4)$$

Here γ is independent of x and we have

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

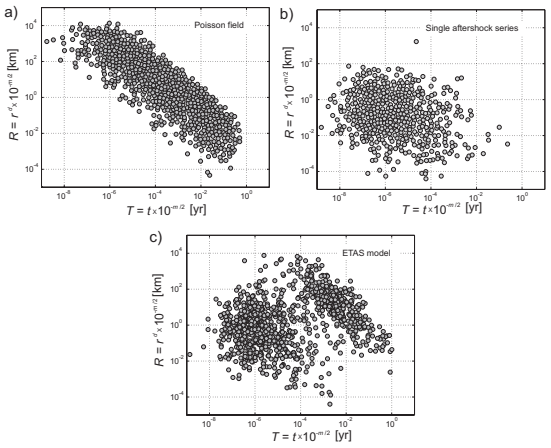

FIG. 1: Distribution of time and space components, (T, R), of the nearest-neighbor distance η^* for homogeneous Poisson field with exponential magnitudes (a), single aftershock series obeying Omori law (b), ETAS model (c).

where $\Psi(w) \sim \psi(w)$ stays for $\lim_{w\to\infty} \dfrac{\Psi(w)}{\psi(w)} = 1$.

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, when the latter is known (which is obviously the case in practice). This facilitates analysis of data from different periods and regions that might have different m_0.

Let earthquake i be the nearest neighbor for earthquake j, that is $\eta_j^* = \eta_{ij}$. We define, for arbitrary $0 \le q \le 1$,

$$T_{ij} := \tau_{ij}\, 10^{-b\, m_i\, q}, \quad R_{ij} := r_{ij}^d\, 10^{-b\, m_i\, (1-q)}. \qquad (6)$$

Obviously $\eta_j^* = T\,R$ (without loss of generality, we assumed here $C = 1$ and $m_0 = 0$) and Proposition 2 implies that the distribution of the pair (T, R) is concentrated along the line $\log_{10} T + \log_{10} R = x_{\rm m}$, where $x_{\rm m}$ is the mode of the distribution (4), while the level lines are of the form $\log_{10} T + \log_{10} R$ =const. Figure 1a 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 nearest-neighbor distances η_j^* 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 Epidemic Type Aftershock Sequence (ETAS) model was introduced by Y. Ogata [5]; it specifies a marked point process N by its conditional intensity at instant t and spa-

tial location (x, y):

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

where $\Lambda_0 > 0, \tau = t - t_i, r^2 = (x - x_i)^2 + (y - y_i)^2$, and the temporal ($\Lambda_T$) and spatial ($\Lambda_R$) 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.

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

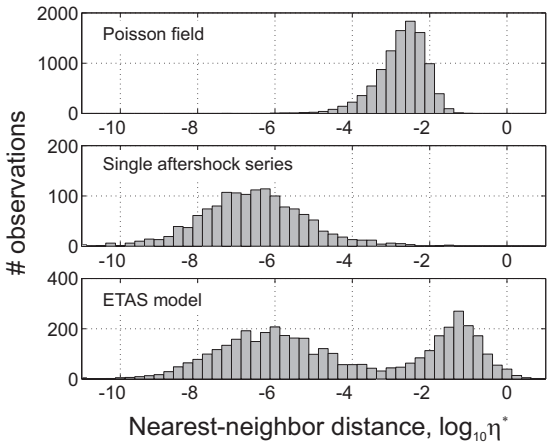

FIG. 2: Distribution of the logarithm of the nearest-neighbor distance, $\log_{10}\eta^*$, for homogeneous Poisson field with exponential magnitudes (top), single aftershock series obeying Omori law (middle), ETAS model (bottom).

Figures 1 and 2 show the distributions of $\log_{10}\eta^*$ and corresponding pairs (T, R). The Poisson model behaves as suggested by the Proposition 2. For a single aftershock series, one observes almost symmetric scatter, which suggests that T and R are independent. This is the most important difference from the Poisson model. The ETAS distribution has two prominent "modes": A scatter along TR =const. in the upper right part of the plot 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 (ANSS) [6], and consider earthquakes with magnitude $m \geq 2.0$ that fall within the square region bounded by $122^\circ W$, $114^\circ W$, $32^\circ N$, $37^\circ N$ during January 1, 1984 - December 31, 2004.

The empirical distributions of the logarithm of the nearest-neighbor distance, $\log_{10}\eta^*$, and its components

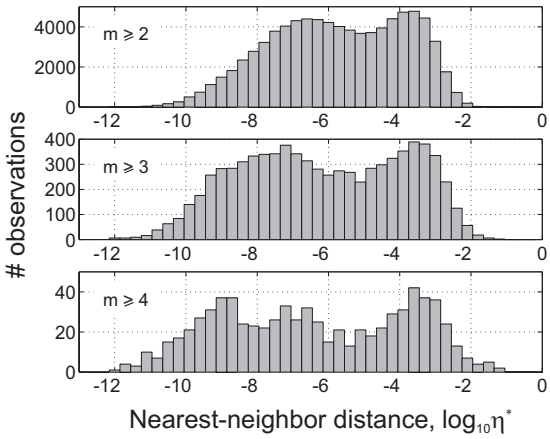

FIG. 3: Distribution of $\log_{10}\eta^*$ for the observed seismicity of Southern California during 1984-2004; different panels correspond to different lower magnitude cutoffs. Here $d = 1.6$ as in [3], $p = q = 1/2$. Notice the bimodal structure with the same boundary between modes at $\eta \approx 10^{-5}$.

$(T,\ R)$ are shown in Figs. 3,4. Both distributions are prominently bimodal reminiscent of that observed for ETAS model; they reveal existence of two statistically distinct earthquake populations. One of them corresponds to $\log_{10} T + \log_{10} R \approx 10^{-3}$; according to the Proposition 2 it describes homogeneous (Poisson) seismicity. The other population corresponds to $\log_{10} R \approx 10^{-2}$; it corresponds to the aftershock clustering.

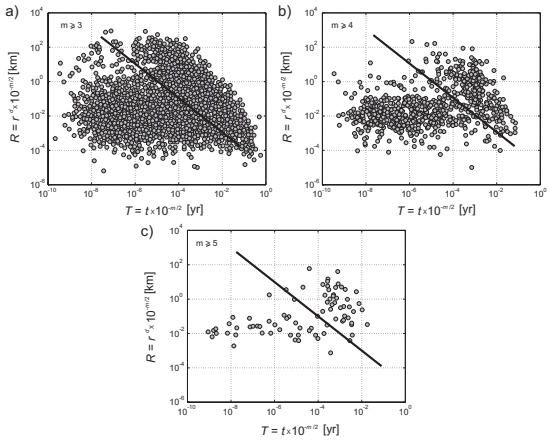

FIG. 4: Distribution of time and space components, (T, R), of the nearest-neighbor distance η^* for the observed seismicity of Southern California during 1984-2004. Notice the bimodal structure; the location of a solid line $\log_{10} T + \log_{10} R = 10^{-5}$ is the same in all panels.

To detect individual aftershocks, we fix a threshold η_0 and remove all the links with $\eta_j^* > \eta_0$ from the tree $\mathcal{T}$. This will result in the forest (set of trees) $\mathcal{F}(\eta_0) = \{\mathcal{T}_i\}_{i=1}^{N(\eta_0)}$. Each tree $\mathcal{T}_i$ in the forest corresponds to a single earthquake cluster: The distance between *linked* elements within any tree is smaller than that between any two elements from

distinct trees. Those 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 mainshocks [4]. Possible earthquake clusters observed prior to the mainshock are then called *foreshocks*. In this situation, it is natural to define i-th *mainshock* as the largest earthquake within the tree $\mathcal{T}_i$, and *aftershocks* (*foreshocks*) as the events from $\mathcal{T}_i$ that occurred later than (prior to) the mainshock. 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. 3,4). The figure focuses on 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 Landers' epicenter; b) the "secondary" aftershocks after the Big Bear earthquake, M=6.4, which itself is the largest aftershock of Landers; c) several earthquakes that occurred immediately after Landers but at large distance from the latter. Such "distant" aftershocks present a special interest in many seismic studies. Both Northridge and Hector Mine aftershock clusters have not been associated with Landers. We emphasize though existence of a distant Landers' aftershock close to the future epicenter of Hector Mine.

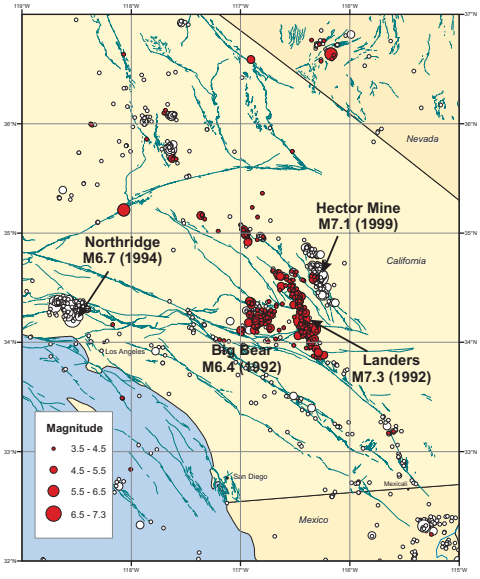

FIG. 5: (Color online) Aftershock identification for Landers earthquake (June 28, 1992, M7.3). The figure shows all earthquakes that occurred after the Landers. Shaded circles mark earthquakes identified as Landers' aftershocks; open circles mark the rest of earthquakes.

CONCLUSION AND DISCUSSION

We demonstrated the existence of statistically distinct clustered and non-clustered parts in the observed seismicity. 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.

The current definition of the distance η remains *ad hoc*; a partial justification for this choice is provided by our result on the distribution for η^* (Proposition 2), which coincides with the Euclidean nearest-neighbor distance distribution for a homogeneous (unmarked) point field. An analog of Proposition 2 is readily proven for any nearest-neighbor distance that depends multiplicatively on spatio-temporal 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 applied part of the proposed clustering analysis.

This study was partly supported by NSF, Grant ATM 0327558 and the Southern California Earthquake Center. SCEC is funded by NSF Cooperative Agreement EAR-0106924 and USGS Cooperative Agreement 02HQAG0008. The SCEC contribution number for this paper is 1137.

* Electronic address: zal@unr.edu
† Electronic address: agabriel@math.purdue.edu
‡ Electronic address: vkb@ess.ucla.edu

[1] B. Romanowicz, Science, 260 (5116), 1923-1926 (1993); J. Dieterich, J. Geophys. Res., 99 (B2), 2601-2618 (1994); F. Press and C. Allen, J. Geophys. Res., 100 (B4), 6421-6430 (1995); K. R. Felzer and E. E. Brodsky, Nature, 441 (7094), 735-738 (2006); G. C. P. King and D. D. Bowman, J. Geophys. Res., 108 (B2), Art. No. 2096 (2003).

[2] J. Rundle, D. Turcotte, and W. Klein (eds), *Geocomplexity and the Physics of Earthquakes.* (AGU, Washington DC, 2000); V. I. Keilis-Borok and A. A. Soloviev (eds), *Nonlinear Dynamics of the Lithosphere and Earthquake Prediction.* (Springer, Heidelberg, 2003); D. Sornette, *Critical Phenomena in Natural Sciences.* 2-nd ed. (Springer-Verlag, Heidelberg, 2004).

[3] M. Baiesi and M. Paczuski, Phys. Rev. E., 69, Art. No. 066106 (2004); M. Baiesi and M. Paczuski, Nonlin. Proc. Geophys., 12, 1-11 (2005); M. Baiesi, Physica A, 360, 534-542 (2006).

[4] L. Knopoff and J. K. Gardner, Geophys. J. R. Astron. Soc., 28, 311313 (1972); V. I. Keilis-Borok, L. Knopoff, I.M. Rotwain, Nature, 283 (5744), 259-263 (1980); P. Reasenberg, J. Geophys. Res., 90, 54795495 (1985); S. D. Davis and C. Frohlich, Geophys. J. Intl., 104 (2), 289-306 (1991); G. M. Molchan and O. E. Dmitrieva Geophys. J. Intl., 109 (3), 501-516 (1992); J. Zhuang, Y. Ogata, D. Vere-Jones, J. Am. Stat. Assoc., 97 (458), 369-380 (2002).

[5] Y. Ogata, Ann. Inst. Stat. Math., 50 (2), 379-402 (1998).

[6] Available at http://quake.geo.berkeley.edu/anss/