JOURNAL OF GEOPHYSICAL RESEARCH: SOLID EARTH, VOL. 118, 2847–2864, doi:10.1002/jgrb.50179, 2013

# Earthquake clusters in southern California I: Identification and stability

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

Received 9 December 2012; revised 25 March 2013; accepted 28 March 2013; published 10 June 2013.

[1] We use recent results on statistical analysis of seismicity to present a robust method for comprehensive detection and analysis of earthquake clusters. The method is based on nearest-neighbor distances of events in space-time-energy domain. The method is 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. Analysis of the ETAS model demonstrates that the cluster detection results are *accurate* and *stable* with respect to (1) three numerical parameters of the method, (2) variations of the minimal reported magnitude, (3) catalog incompleteness, and (4) location errors. 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*. The results reproduce the essential known statistical properties of earthquake clusters, which provide overall support for the proposed technique. In addition, systematic analysis with our method allows us to detect several new features of seismicity that include (1) existence of a significant population of *single-event clusters*, (2) existence of foreshock activity in natural seismicity that exceeds expectation based on the ETAS model, and (3) dependence of all cluster properties, except area, on the *magnitude difference* of events from mainshocks but not on their absolute values. The classification of detected clusters into several major types, generally corresponding to singles, burst-like and swarm-like sequences, and correlations between different cluster types and geographic locations is addressed in a companion paper.

**Citation:** Zaliapin, I., and Y. Ben-Zion (2013), Earthquake clusters in southern California I: Identification and stability, *J. Geophys. Res. Solid Earth*, *118*, 2847–2864, doi:10.1002/jgrb.50179.

## 1. Introduction

[2] Earthquake clustering is an essential aspect of seismicity with signatures in space, time, and size (e.g., magnitude, potency/moment, and 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. Clustering in space is exemplified by the concentration of earthquakes along the boundaries of major tectonic plates and regional fault networks [e.g., *Scholz*, 2002; *Utsu*, 2002]. Clustering in time is best seen as a significant increase of seismic activity immediately after large earthquakes leading to *aftershock sequences* [e.g., *Omori*, 1894; *Utsu*, 1961; *Utsu et al*., 1995; *Kisslinger*, 1996]. Earthquake *swarms*, *foreshocks*, *bursts*, *gaps*, and *switching* of seismic activity among spatio-temporal domains are other terms used to denote different types of seismic clustering [e.g., *Richter*, 1958; *Jones and Molnar*, 1979; *Romanowicz*, 1993; *Utsu*, 2002; *Felzer and Brodsky*, 2006; *Vidale et al*., 2006; *Vidale and Shearer*, 2006; *Ben-Zion*, 2008; *Shearer*, 2012].

[3] 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 article is a companion to *Zaliapin and Ben-Zion* [2013] doi:10.1002/jgrb.50178.

Additional supporting information may be found in the online version of this article.

[1]Department of Mathematics and Statistics, University of Nevada, Reno, Nevada, USA.

[2]Department of Earth Sciences, University of Southern California, Los Angeles, California, USA.

Corresponding author: I. Zaliapin, Department of Mathematics and Statistics, University of Nevada, Reno, NV 89557, USA. (zal@unr.edu)

©2013. American Geophysical Union. All Rights Reserved.
2169-9313/13/10.1002/jgrb.50179

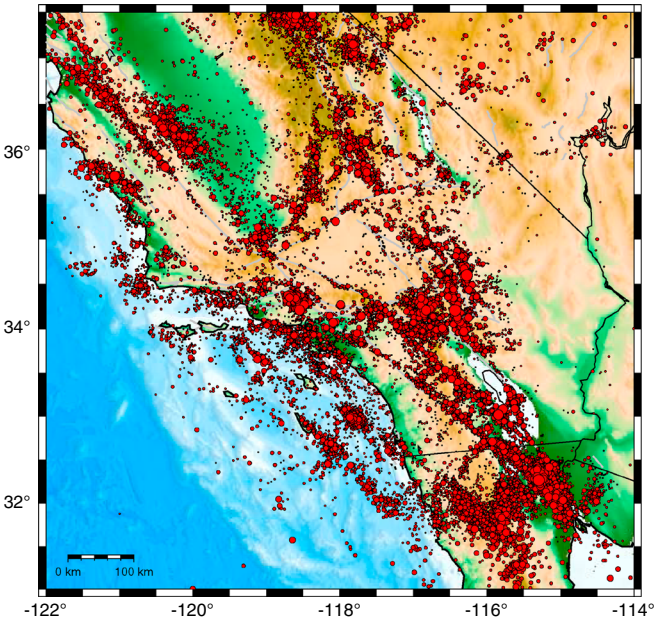


**Figure 1.** Map of earthquake epicenters, $m \geq 2$, from the relocated catalog of *Hauksson et al*. [2012]. Circle size is proportional to magnitude. Major faults are shown by gray lines.

[4] This study takes advantage of recent results on statistical cluster identification [*Zaliapin et al*., 2008; *Zaliapin and Ben-Zion*, 2011], and recent empirical evidence [e.g., *Vidale et al*., 2006; *Vidale and Shearer*, 2006; *Enescu et al*., 2009; *Holtkamp et al*., 2011; *Shearer*, 2012], to develop a comprehensive approach toward objective and robust analysis of seismic clusters. This paper is the first in a series having the following specific goals: (1) identify statistically significant earthquake clusters in southern California, (2) classify the detected clusters into several main types according to their statistical properties, and (3) relate the detected cluster types to key governing properties of the crust. The present paper focuses on the first goal; the other two will be addressed in follow-up papers.

[5] The employed data and basic methods of catalog analysis are outlined in section 2. The cluster detection that forms the core of the study follows from results of *Zaliapin et al*. [2008] and is described in section 3. The proposed method is based on a generally observed bimodal distribution of nearest-neighbor earthquake distances in a combined space-time-magnitude domain (section 3; sections A and B in the supporting information). The nearest-neighbor distances quantify the deviations of observed seismicity from a SIP process. We show in supporting information section B that the observed bimodal distribution cannot result from marginal spatial or temporal clustering of earthquakes and hence cannot be attributed, for instance, to the complexity of fault networks. Instead, it is fundamentally due to dependent space-time seismicity structures associated primarily with foreshock-mainshock-aftershock sequences. The observed bimodal distribution of earthquake distances provides a natural tool for partitioning an examined earthquake catalog into separate individual *clusters*: events within a cluster are abnormally close to their nearest-neighbors, while events from distinct clusters are relatively far from each other.

[6] The clusters are naturally divided into *singles* that contain just one event and *families* having multiple events that are subclassified into *foreshocks, mainshocks, and aftershocks* (section 3.5). The suggested classification is consistent overall with traditional expert definition of different types of earthquakes (mainshocks, foreshocks, and aftershocks). Our methodology, however, is based solely on statistical properties of the data and can be used to analyze seismicity objectively and systematically in different geographic regions, time intervals, and magnitude ranges.

[7] The employed cluster technique is characterized by (1) soft parameterization that uses only three easily estimated parameters (the *b*-value of the magnitude distribution, the spatial dimension of epicenters, and the threshold that separates "very close" from other distances); (2) the ability to uniformly analyze clusters associated with mainshocks of greatly different magnitude; (3) demonstrated high stability of the cluster detection with respect to the employed parameters, minimal reported magnitude, catalog incompleteness, and location errors; and (4) the 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]. The proposed algorithm is *objective*, since it is based on a general intrinsic property of natural seismicity (bimodal earthquake distance distribution) rather than on any ad hoc division criteria, and it provides a robust tool for systematic analysis of earthquake clusters that span wide regions of space, time and sizes.

[8] Section 4 presents analysis of various statistical properties of the detected clusters. This analysis has two goals. First, it confirms that our technique reproduces the essential known properties of seismic clusters, which is important for the validation of the proposed approach. Second, it reveals several interesting new or not well-documented features of seismic clusters. These include the existence of a prominent population of single-event clusters (given the employed catalog resolution); similarity of the distributions of magnitude differences between the mainshock and largest aftershock and foreshock (resulting in the Båth law for both aftershocks and foreshocks); closer overall proximity of foreshock magnitudes to that of mainshocks compared to the aftershock magnitudes; and dependence of the cluster structure on the difference between magnitudes of mainshock and cluster events, rather than on their absolute magnitudes. A companion paper [*Zaliapin and Ben-Zion*, 2013] uses the developed approach to demonstrate the existence of several types of seismicity clusters in southern California that are characterized by distinct topological properties and geographic location and, therefore, likely associated with different failure processes.

## 2. Data and Basic Methods

### 2.1. Data

[9] We work with the relocated southern California earthquake catalog of *Hauksson et al*. [2012], available via the SCEC data center (http://www.data.scec.org/research-tools/downloads.html). Figure 1 shows the epicenters of 111,981 earthquakes with magnitude $m \geq m_c = 2$ used in this study. The time-latitude map of seismicity with $m \geq 3$ in Figure 2 illustrates visually various changes of seismic intensity, most clearly related to aftershock sequences of large earthquakes.

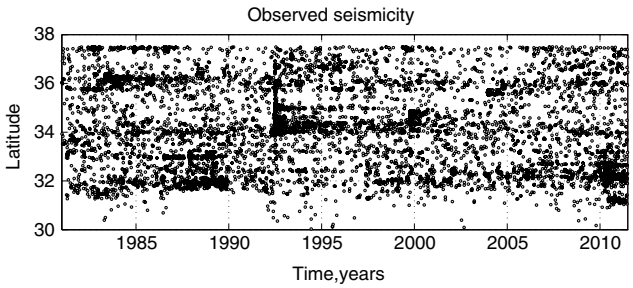


**Figure 2.** Epicenters of the earthquakes with $m \geq 3$ as a function of time and latitude.

### 2.2. Completeness

[10] The employed magnitude threshold $m_c = 2$ is lower than the completeness magnitude for southern California, which is estimated to be above 3.0 [*Felzer*, 2008; *Schorlemmer and Woessner*, 2008]. However, we demonstrate in sections D and E in the supporting information material that the cluster structure of the events is insensitive to the catalog incompleteness as well as to the minimal reported magnitude. This supports the assumption that the recovered cluster structure is close to the one that would be observed in a complete catalog. The results for $m_c = 3$ (not shown) are qualitatively similar to the ones reported in this study, yet the number and size of the clusters is insufficient for presenting visually clear results.

### 2.3. Δ-analysis

[11] Any aftershock analysis is intrinsically affected by the existence of the catalog lower cutoff magnitude $m_c$. For instance, if we analyze earthquakes with magnitudes $m \geq m_c = 2$, then an earthquake of magnitude $m = 2$ cannot have aftershocks of a smaller magnitude, while an $m = 6$ event may have aftershocks with magnitudes $2 \leq m \leq 6$. To equalize the magnitude ranges for potential fore/aftershocks of mainshocks with different magnitudes, we often perform a Δ-*analysis* that (1) only considers mainshocks with magnitude $m \geq m_c + \Delta = 4$ and (2) only considers fore/aftershocks with magnitude within $\Delta = 2$ units below that of the mainshock. The fore/aftershocks detected by this analysis are called Δ-fore/aftershocks. The conventional analysis that considers all events is referred to as *regular analysis*.

## 3. Earthquake Clustering: Nearest-Neighbor Approach

[12] We detect earthquake clusters based on analysis of nearest-neighbors in a multidimensional domain that includes the location, time, and size of earthquakes. It is shown below (section 3.3 and Figure 4) that the nearest-neighbor distances of recorded earthquakes in this combined domain are separated clearly into two subpopulations. The first population is comprised of *clustered events* that occur unusually close (in a sense to be rigorously defined) in time and space to their nearest neighbors. The second population is comprised of *background events* that happen farther away from their nearest neighbors; the spatio-temporal distribution of background events is reminiscent of that for a SIP process. We note that the term *background* is not equivalent to *homogeneous*. In fact, a rigorous analysis [e.g., *Luen and Stark*, 2012] can reject the hypothesis that the background events are a realization of a SIP process. Nevertheless, it will be demonstrated below (Figures 3, 4, and 9) that the deviations from a SIP realization in the background subpopulation are *orders of magnitude* less than in the clustered population. This motivates us to focus on the clustered population and consider its objective statistical identification.

[13] We emphasize that the problem considered in this study is different from *catalog declustering*, which is formulated as removing some events from a catalog in order to obtain a homogeneous remaining point field. We focus, instead, on identifying individual statistically significant clusters and analyzing (1) the properties of the clusters and (2) the properties of the point field represented by the single maximal event of each cluster, whether or not this field is Poissonian. The history of seismicity cluster analysis and the existing declustering approaches [e.g., *Gardner and Knopoff*, 1974; *Reasenberg*, 1985; *Molchan and Dmitrieva*, 1992; *Dzwinel et al.*, 2005; *Marsan and Lengline*, 2008; *Zhuang et al.*, 2002] suggest that the two problems are related. A traditional approach to the declustering problem,

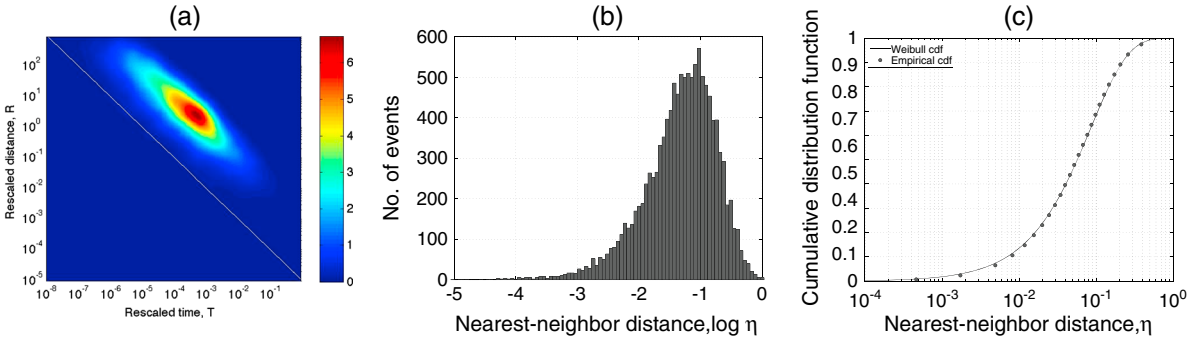


**Figure 3.** Cluster analysis for a time-stationary space-homogeneous Poisson process with exponential magnitudes; the number of events is 12,105, the same as the number of $m \geq 3$ events in the *Hauksson et al.* [2012] catalog. (a) Joint distribution of the rescaled time and distance components ($T$,$R$) of the nearest-neighbor distance $\eta$. The distribution has single mode located along the line $\log R + \log T = \text{const}$. (b) Histogram of the nearest-neighbor distances $\eta_{ij}$; the distribution is clearly unimodal. (c) Empirical cdf (circles) of the nearest-neighbor distance $\eta_{ij}$ and the Weibull cdf (black line), which provides a very close approximation to the data.

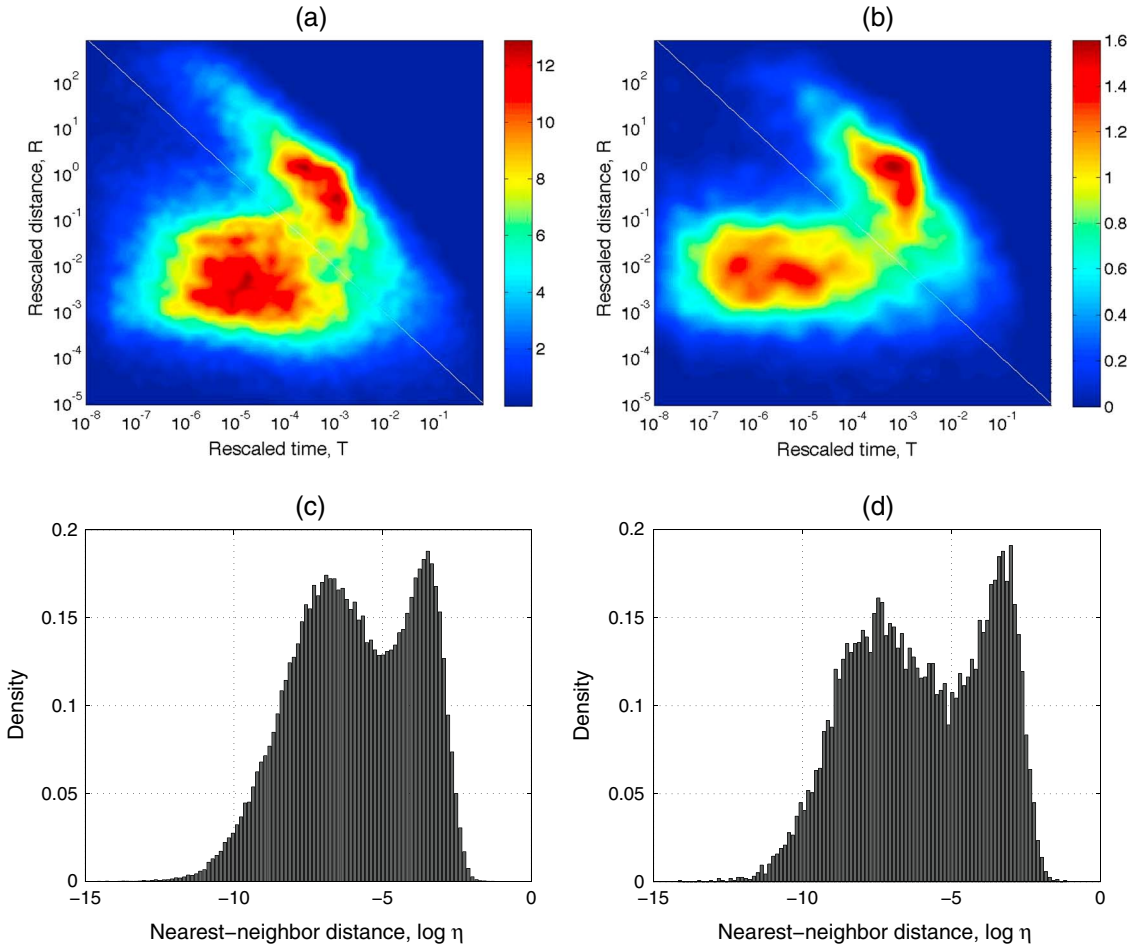


**Figure 4.** Distribution of the nearest-neighbor distance $\eta$ in southern California using the relocated catalog of *Hauksson et al*. [2012]. (a, b) The joint distribution of the rescaled time and space components ($T$,$R$). (c, d) Histogram of the nearest-neighbor distance $\eta$; the values are normalized to sum up to unity. Different panels refer to different values of the magnitude cutoff $m_c$ and, accordingly, to different number $n$ of examined events: (a, c) $m_c = 2$, $n = 111{,}981$; (b, d) $m_c = 3$, $n = 12{,}105$. The bimodal distribution is clearly seen in each panel. Notably, the location of the upper mode, as well as the vertical location of the lower mode, is independent of the magnitude cutoff. The line $\log R + \log T = -5$ that separates the two modes is shown in white in Figures 4a and 4b (cf. Figure 3).

pioneered by *Gardner and Knopoff* [1974], is to detect and remove "aftershocks" and to test the remaining field, mostly comprised of mainshocks, for stationarity and homogeneity. These authors have hypothesized in their classical paper that the resulting field is a realization of a SIP process. This gave a strong impetus to a tradition in statistical seismology to test declustering algorithms against a stationary Poisson outcome. A recent study by *Luen and Stark* [2012] demonstrates, however, that the SIP hypothesis is rejected in analysis of currently available catalogs and declustering methods in southern California. This conclusion is not surprising if one takes into account numerous mechanisms leading to time-dependent evolution of seismicity not related to aftershock clustering. These include seismic migration, swarms, regional changes of seismic intensity, switching of seismic activity between different faults, and technical problems with routine recording of events, in particular in the immediate vicinity of a large event. This provides additional motivation for the approach used in this study, whose primary focus is on the properties of clusters as opposed to that of a declustered catalog.

### 3.1. Distance Between Earthquakes

[14] Consider an earthquake catalog where each event $i$ is characterized by occurrence time $t_i$, hypocenter ($\phi_i$, $\lambda_i$, $d_i$), and magnitude $m_i$. Our initial goal is to identify for each earthquake $j$ its possible *parent*, which is an earlier earthquake $i$ that is the closest, in some sense, to $j$ among all earlier events. This motivates us to consider a distance that is asymmetric in time. Following *Baiesi and Paczuski* [2004], the distance between earthquakes $i$ and $j$ is defined as

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

[15] Here $t_{ij} = t_j - t_i$ is the inter-occurrence time in years, which is positive if earthquake $i$ happened before event $j$ and negative otherwise, $r_{ij} \ge 0$ is the spatial distance between the earthquake hypocenters in kilometers, and $d_f$ is the (possibly fractal) dimension of the earthquake hypocenter distribution. In the main analysis of this paper, we compute

the distance $\eta_{ij}$ with parameters $b = 1$, $d_f = 1.6$. The depth of the earthquakes is ignored, and $r_{ij}$ is computed as the surface distance between the event epicenters.

[16] It will be convenient to represent the scalar distance $\eta$ in terms of its space and time components normalized by the magnitude of the parent event $i$ [*Zaliapin et al.*, 2008]:

$$T_{ij} = t_{ij}10^{-qbm_i}; R_{ij} = \left(r_{ij}\right)^{d_f} 10^{-(1-q)bm_i}. \quad (2)$$

[17] It is readily seen that $\eta_{ij} = T_{ij}R_{ij}$ or, equivalently, $\log_{10}\eta_{ij} = \log_{10} T_{ij} + \log_{10} R_{ij}$. In this work, we always use $q = 0.5$. Figure A1 (supporting information section A) illustrates the connection between the normalized time $T$ and time in years for events of different magnitudes $m_i$.

[18] The *nearest-neighbor distance* (NND) for a given event $j$ is the minimal distance among $\eta_{ij}$ where $i$ goes over all earlier events in the catalog. The event $i$ that corresponds to the nearest-neighbor distance is called the *nearest-neighbor*, or *parent*, of event $j$. We use for the NND the same notation $\eta_{ij}$ as for the general distance in equation (1), which should create no confusion.

[19] Our cluster analysis is based on significant deviations of the observed NND $\eta$ from the values expected in the absence of clustering. The next section reviews the properties of the distance $\eta$ and its 2-D expansion ($T$,$R$) for a stationary homogeneous Poisson point process that by construction has no clustering.

### 3.2. Homogeneous Poisson Process

[20] Consider a marked Poisson point process that is homogeneous in $d_f$-dimensional space, stationary in time, and has magnitudes (marks) that follow the Gutenberg-Richter distribution; we will refer to this process as a *Stationary Homogeneous Poisson* (*SHP*) *process*. *Zaliapin et al*. [2008] and *Hicks* [2011] demonstrated that the NND $\eta_{ij}$ in SHP can be closely approximated by the Weibull distribution. The traditional NND for a Poisson point field in a Euclidean space also has the Weibull distribution [*Feller*, 1970]. This relates our cluster analysis in space-time-magnitude domain and classical results on clustering in multidimensional spaces. *Zaliapin et al*. [2008] also demonstrated that the joint 2-D distribution of ($T$, $R$) in a SHP process is unimodal and is concentrated along the line $\log_{10} T + \log_{10} R = \text{const}$.

[21] Figure 3 shows results for a SHP process with 12,105 events, which is the number of $m \geq 3$ events in the relocated catalog of *Hauksson et al*. [2012]. Figure 3a displays the joint distribution of ($T$, $R$) estimated using a Gaussian kernel smoothing; the distribution is clearly unimodal. The white line corresponds to $\log_{10} R + \log_{10} T = -5$ and is shown for visual comparison of these synthetic results with later results for southern California. Figure 3b shows the histogram of the NND on a logarithmic scale; it further emphasizes the unimodal shape of the distribution. Figure 3c juxtaposes the empirical cumulative distribution function (cdf) of the NND $\eta$ and the theoretical cdf for the Weibull distribution, which provides a very close approximation for the observations [*Zaliapin et al*., 2008; *Hicks*, 2011].

### 3.3. Observed Seismicity

[22] In contrast to the above results for a SHP process, analysis of the observed seismicity reveals a prominently bimodal distribution of $\eta$ and of the joint distribution of ($T$, $R$), with an additional mode located closer to the origin. Figure 4 shows the distribution of the NND $\eta$ for two different cutoff magnitudes in southern California. The first striking observation is the existence of two modes: One is extended along and above the white diagonal line in Figures 4a and 4b; this mode is reminiscent of the distribution for a SHP process shown in Figure 3. We refer to this mode as *background*. The other mode is located closer to the origin and has horizontally elongated shape in the 2-D plot. We call this mode *clustered*. We note that the location of the background mode is independent of the magnitude cutoff of the analysis; the vertical (space) location of the cluster mode is also independent of the magnitude cutoff.

[23] The bimodal distribution of the NND is a general feature of observed seismicity. *Hicks* [2011] used the NEIC catalog 1973–2011, $m \geq 4$, and found similar bimodal distribution for the worldwide seismicity, as well as the regional seismicity of Japan, New Zealand, and Africa. *Bautista* [2011] demonstrated a bimodal distribution for seismicity in Nevada, using the catalog produced by the Nevada Seismological Laboratory.

[24] Supporting information section B shows that the reported bimodal distribution cannot be explained by spatial or temporal marginal distributions of the observed earthquakes; instead, it is associated with *dependent* spatio-temporal structures. The existence of the cluster mode is ultimately caused by groups of earthquakes that happen within highly localized regions in both space and time. Such groups mainly correspond to foreshock-mainshock-aftershock sequences or swarms. In the next section we use the bimodality of the ($T$,$R$) distribution to identify individual space-time clusters of seismicity.

### 3.4. Spanning Network, Forest of Earthquakes

[25] By connecting each earthquake $j$ in the catalog to its nearest neighbor (parent) $i$ according to the NND $\eta$, we form a single cluster that contains all examined events. In the graph-theoretical language, the nearest-neighbor links form a *spanning network*. The spanning network for the nearest-neighbor analysis of earthquakes with $m \geq 4$ in southern California is shown in Figure 5a. Statistical properties of the spanning network for California seismicity according to the ANSS catalog were studied by *Baiesi and Paczuski* [2004]. From a topological point of view, the spanning network is a *tree*, which means that it does not have loops. The tree structure of the nearest-neighbor spanning network can be not very intuitive, since this property does not hold in Euclidean spaces, so we prove it in supporting information section C.

[26] Each link in the spanning tree is assigned a *strength* inversely proportional to the respective NND $\eta$. This allows separating all the links in the spanning tree into *weak* and *strong*, in accordance with the bimodal distribution of $\eta$. Specifically, weak links (large distances) are defined by the condition $\eta \geq \eta_0$; they correspond (Figure 4) to the background part of the bimodal distribution of ($T$,$R$). Strong links (short distances) are defined by the complementary condition $\eta < \eta_0$; they correspond to the cluster part of the bimodal distribution of ($T$,$R$). A visual inspection of Figure 4 suggests $\eta_0 = 10^{-5}$ as a reasonable separation threshold. *Hicks* [2011] performed a formal analysis to establish the best boundary between the two modes, considering a Gaussian mixture model with the background and cluster modes.

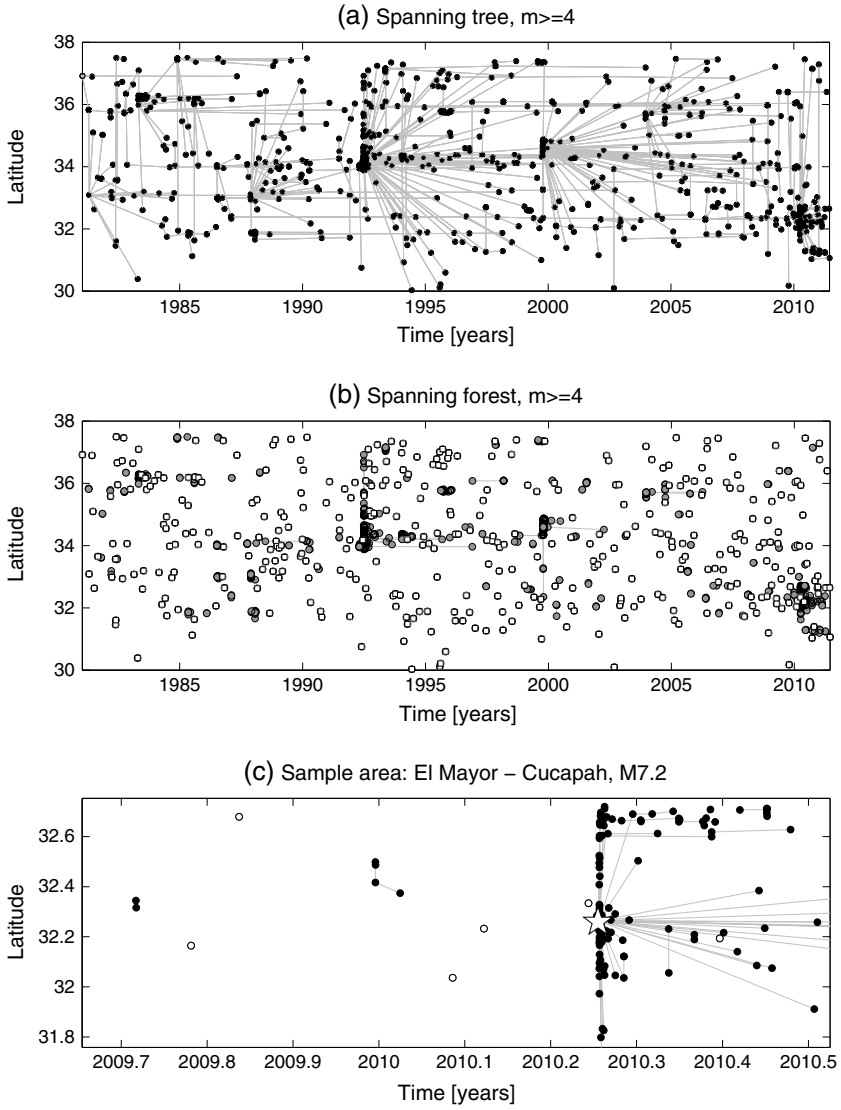


**Figure 5.** Nearest-neighbor analysis of earthquakes with $m \geq 4$ from the relocated catalog of *Hauksson et al*. [2012]. The nearest-neighbor links are shown by grey lines. (a) Spanning tree of nearest-neighbors. (b) Nearest-neighbor forest obtained by removing the weak links from the spanning tree. Open circles represent single events, filled circles represent clusters. (c) Sample area around the El Mayor – Cucapah M7.2 earthquake of 10 April 2010 (white star).

The model was estimated using the Expectation Maximization approach [*Hastie et al*., 2001]. The analysis was performed in 1-D for values of $\log_{10} \eta$ and in 2-D for the joint distribution of ($\log_{10} T$, $\log_{10} R$). Such a formal analysis suggests also that $\eta_0 \approx 10^{-5}$ .

[27] We now remove the weak links from the spanning tree and keep only the strong links. This results in a *spanning forest* (Figure 5b), which is a collection of distinct trees that span all events in the catalog. The forest contains many single-event trees, which we call *singles* and show by open circles. The other events are connected in multievent clusters, which we call *families*. By construction, each family is comprised of highly coupled events that are connected to their parents by strong links. There are 1146 earthquakes with $m \geq 4$ in the examined catalog. Of these, 685 (60%) have strong links to parents and 461 (40%) have weak links. There are 373 (33%) singles; the other 773 (67%) events form 89 families. The size of the families ranges from 2 to 166 events. Figure 5b, however, creates a visual impression that most of the events are singles (empty circles), while the family events (filled circles) comprise a small fraction of the catalog. This impression is due to the fact that the more numerous family events are highly clustered in space and time, while the less numerous singles are distributed more uniformly. As an illustration, Figure 5c zooms on an area around the M7.2 El-Mayor–Cucapah earthquake of 4 April 2010. The aftershock sequence of this earthquake contains numerous events highly clustered in space and time; they all belong to a single family.

[28] Next, we focus on the internal structure of the clusters in the spanning nearest-neighbor forest.

### 3.5. Mainshocks, Aftershocks, Foreshocks

[29] We introduce the following terminology for the events within a family (Figure 6). The earthquake with the

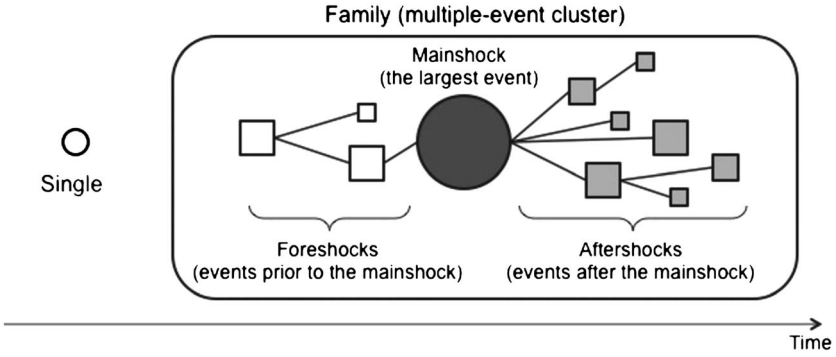


**Figure 6.** Event classification: an illustration. The figure illustrates the definition of single (empty circle) and family, which consists of the mainshock (dark circle), aftershocks (dark squares), and foreshocks (empty squares). The same symbols and color code is used below in Figures 7 and 8.

largest magnitude in a family is called *mainshock*. If there are several earthquakes with the largest magnitude within a family, the first one is considered to be the mainshock; hence, each family has a single mainshock. All events in a family that occurred after the mainshock are called *aftershocks*. All events that occurred prior to the mainshock are called *foreshocks*. This terminology closely resembles the one commonly used in the literature on earthquake cluster analysis. Altogether, we have (Figure 6) two types of clusters—singles and families—and three types of events within families—mainshocks, aftershocks, and foreshocks. The event classification depends on the catalog magnitude cutoff. For instance, with a lower cutoff some singles may become mainshocks (after being connected to possible foreshocks and aftershocks), while with a higher cutoff, some mainshocks may become singles. Other changes of event types are also possible.

[30] Table 1 summarizes the individual event identification in the nearest-neighbor analysis of earthquakes with magnitude $m \geq 2$ in southern California. The proportion of events of different type is very stable for magnitudes below 5, with about 37% of mainshocks and singles, 56% aftershocks, and 7% foreshocks. These proportions change for larger events, giving significant preference to the mainshocks, which seems very intuitive (see also section 4.1). According to a Fisher test [*Agresti*, 2002], this proportion change can hardly be attributed (not shown) to the decrease of the event number within the large magnitude ranges and hence represents an actual effect. Table 2 summarizes the event classification in Δ-analysis. Figure 7 shows the spatial distribution of the mainshocks and singles in Δ-analysis.

[31] Figure 8 shows events of different types in time-latitude coordinates. The figure demonstrates that the nearest-neighbor analysis detects the most obvious clusters, mostly related to aftershock sequences of large earthquakes. At the same time, there are still some clear variations in the mainshock intensity, like the one during 1987–-1990 around latitude 32 in Figure 8a. However, the spatio-temporal scales of the groups of mainshocks and singles are much larger than that of the clusters identified by the nearest-neighbor analysis. To illustrate this, we perform the nearest-neighbor analysis only for the mainshocks and singles with different magnitude thresholds (Figure 9). The joint distribution of $(T,R)$ in this case is clearly unimodal, with a single mode located above the line $\log_{10} T + \log_{10} R = -5$, which separates the background and cluster populations in the analysis of the entire catalog (Figure 4). The cluster mode is absent in the analysis of mainshocks and singles, indicating that the essential clustering has been already successfully identified and removed. The single mode of $(T,R)$ is largely spread and does not look like the ellipsoid mode of a Poisson process shown in Figure 3a; this indicates that the mainshock/singles field is not Poissonian. This statement is further confirmed by the fact that the NND do not obey the Weibull distribution (not shown). At the same time, the single mode looks similar to the background mode of the original catalog (Figures 4a and 4b); this suggests that the cluster identification procedure does not distort the nonhomogeneous and possibly nonstationary background events. Identifying and possibly removing the weak remaining clusters of mainshocks and singles is a separate problem, which also can be approached by the nearest-neighbor methodology.

**Table 1.** Statistics of Singles, Mainshocks, Aftershocks, and Foreshocks in the Nearest-Neighbor Analysis of Events With $m \geq 2$

| Magnitude Range | Singles | | Families | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Mainshocks (= No. of Families) | | Aftershocks | | Foreshocks | |
| | No. | % | No. | % | No. | % | No. | % |
| All events: $m \geq 2$ | 34,836 | 31 | 6,557 | 6 | 62,441 | 56 | 8,147 | 7 |
| $2 \leq m < 3$ | 32,376 | 32 | 4,392 | 4 | 55,682 | 56 | 7,426 | 7 |
| $3 \leq m < 4$ | 2,379 | 22 | 1,794 | 16 | 6,115 | 56 | 671 | 6 |
| $4 \leq m < 5$ | 80 | 8 | 312 | 30 | 596 | 58 | 47 | 4 |
| $5 \leq m < 6$ | 1 | 1 | 48 | 49 | 47 | 48 | 2 | 2 |
| $m \geq 6$ | 0 | 0 | 11 | 85 | 1 | 8 | 1 | 8 |

**Table 2.** Statistics of Singles, Mainshocks, Aftershocks, and Foreshocks in the Nearest-Neighbor Δ-Analysis of Events With $m \geq 2$

| Magnitude Range | Singles | | Families | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Mainshocks (= No. of Families) | | Aftershocks | | Foreshocks | |
| | No. | % | No. | % | No. | % | No. | % |
| All events: $m \geq 2$ | 102 | 2 | 350 | 8 | 3260 | 72 | 807 | 18 |
| $2 \leq m < 3$ | - | - | - | - | 2029 | 79 | 532 | 21 |
| $3 \leq m < 4$ | - | - | - | - | 992 | 81 | 229 | 19 |
| $4 \leq m < 5$ | 96 | 15 | 296 | 47 | 199 | 31 | 43 | 7 |
| $5 \leq m < 6$ | 6 | 7 | 43 | 48 | 39 | 43 | 2 | 2 |
| $m \geq 6$ | 0 | 0 | 11 | 85 | 1 | 7 | 1 | 7 |

However, the clustering of mainshocks and singles lacks the bimodality of the original catalog, so analysis of such data requires additional *ad-hoc* rules. The present study does not address this problem.

### 3.6. Quality and Stability of Cluster Identification

[32] The proposed cluster detection technique is based on the earthquake distance of equation (1) and the cluster threshold $\eta_0$. We note that the parameter $q$ of equation (2) is only used for visual purposes (to define and plot rescaled time $T$ and space $R$) and is not involved in the cluster identification. The algorithm is completely parameterized by the triplet $(b,d_f,\eta_0)$ whose values are estimated from the observations, so it does not involves any *ad-hoc* choices or tuning parameters. Nevertheless, there exist statistical variability in the estimation of each parameter—*Marzocchi and Sandri* [2003] give a review of $b$-value estimation with numerous references; the results on estimating fractal distribution of epicenters are reviewed by *Harte* [1998], *Kagan* [2007], and *Molchan and Kronrod* [2009]; the estimation of the threshold $\eta_0$ is discussed by *Hicks* [2011]. The results of the cluster detection might be also affected by the catalog completeness magnitude and earthquake location errors.

[33] To test the performance and stability of the proposed cluster identification, we apply it in supporting information section D to catalogs generated by the ETAS model [*Ogata*, 1998]. This analysis has three goals. First, it demonstrates that the proposed technique is able to correctly identify spatio-temporal clusters in a model with known underlying cluster structure, and to perform well for a wide range of model parameters. Second, it demonstrates stability of the cluster identification with respect to the algorithm parameters, catalog magnitude completeness threshold, and earthquake location errors. Third, it illustrates some statistical effects related to the adopted conditional definition of event types: mainshocks, aftershocks, and foreshocks. Supporting information section E demonstrates the stability of our cluster technique with respect to the above potential sources of error in the observed catalog of southern California seismicity.

[34] In this study, we consider a version of the ETAS model with isotropic spatial kernel and homogeneous spatial background distribution, which is commonly used in analysis of observed seismicity [e.g., *Veen and Schoenberg*, 2008; *Wang et al*., 2010; *Chu et al*., 2011]. Although the isotropy and homogeneity assumptions are violated in the observed seismicity, such model captures the essential aspects of self-excited seismicity and can serve to illuminate basic similarities and differences between the observed and synthetic clusters. We expect that the main conclusions of this work will remain valid for a spatially inhomogeneous/anisotropic version of ETAS. We also notice that the examined ETAS model does not reproduce possible incompleteness of the observed catalogs immediately after large events (so-called short-term aftershock incompleteness).

[35] The results of the quality and stability analyses suggest that (1) the earthquake catalog of southern California has a cluster structure reminiscent in general of that generated by the ETAS model; additional details described in section 4, (2) the cluster structure can be robustly recovered by the proposed technique, and (3) the cluster structure is stable with respect to various choices and numerical parameters of the proposed algorithm.

## 4. Statistics of Detected Clusters

[36] In this section, we examine various statistics of the detected earthquake clusters. The analysis has two goals: (1) to reproduce the known statistical features of aftershock/foreshock series in order to validate the proposed cluster technique and (2) to reveal new properties of earthquake clusters.

### 4.1. Magnitude Distribution

[37] Figure 10 illustrates the magnitude distribution of mainshocks/singles, aftershocks, and foreshocks. Figure 10a

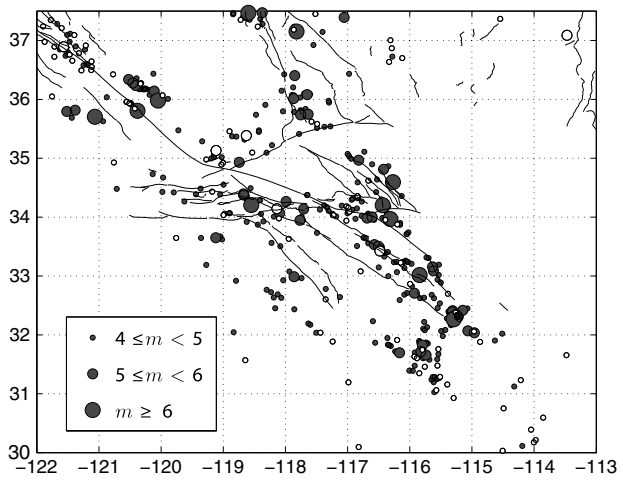


**Figure 7.** Spatial distribution of mainshocks and singles with magnitudes $m \geq 4$ in the nearest-neighbor Δ-analysis of events with $m \geq 2$. Filled circles correspond to family mainshocks, open circles to singles; circle size corresponds to event magnitude, as indicated in the legend. There are no singles with magnitude above 6.

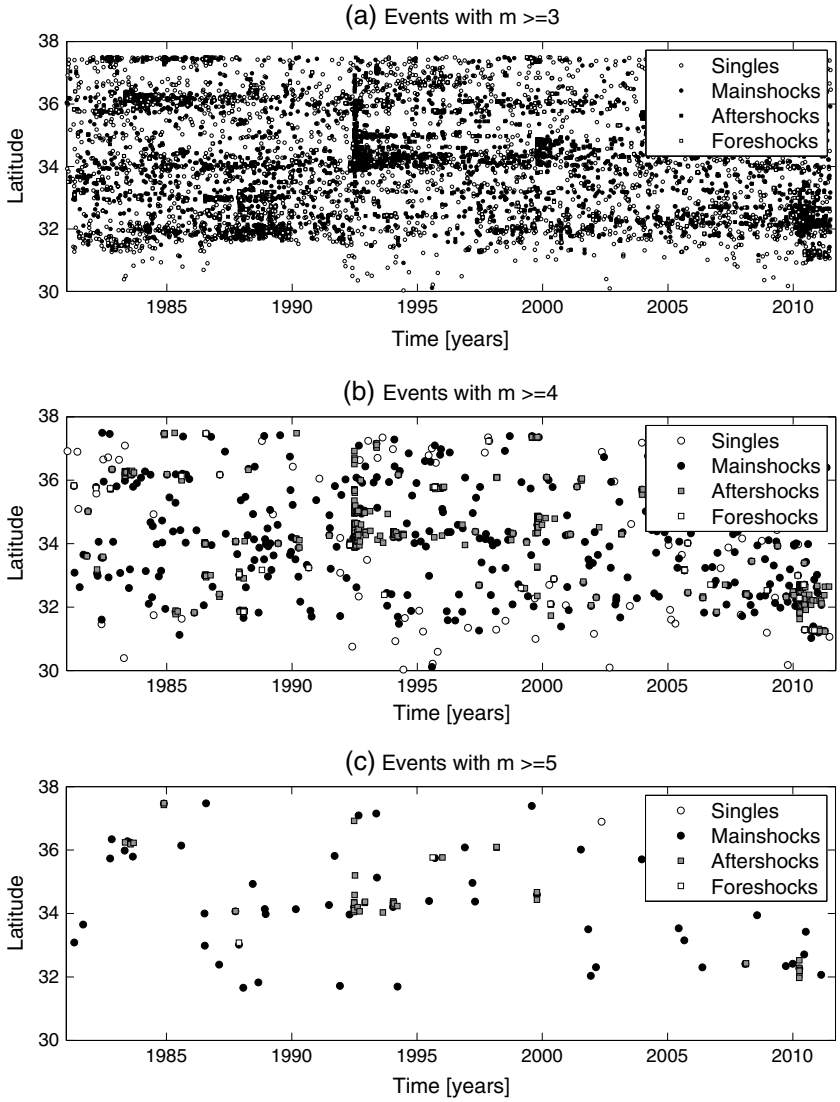


**Figure 8.** Event classification in the nearest-neighbor analysis of earthquakes with $m \geq 2$. Different panels show events of different magnitudes (while the analysis is always done for $m \geq 2$), as indicated in panel titles. Different symbols correspond to different event types, as described in the legend.

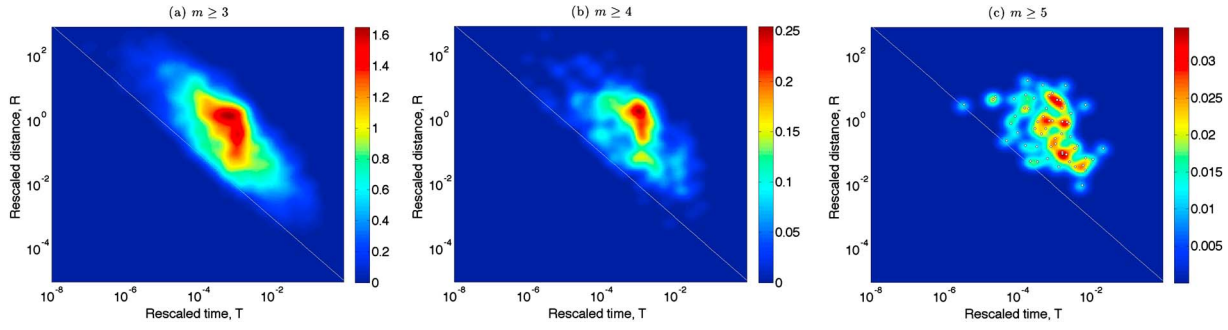


**Figure 9.** The joint distribution of rescaled time and space components ($T$,$R$) of the nearest-neighbor distance $\eta$ for the nearest-neighbor analysis of only mainshocks and singles. Different panels refer to different values of the magnitude cutoff $m_c$ and, accordingly, to different number $n$ of examined events (a) $m_c = 3$, $n = 4{,}569$; (b) $m_c = 4$, $n = 441$; and (c) $m_c = 5$, $n = 59$; individual values of ($T$,$R$) are shown by circles in this panel. The distribution is unimodal and located above the line $\log R + \log T = -5$ that separates the two modes in the analysis of the entire catalog (Figure 4).

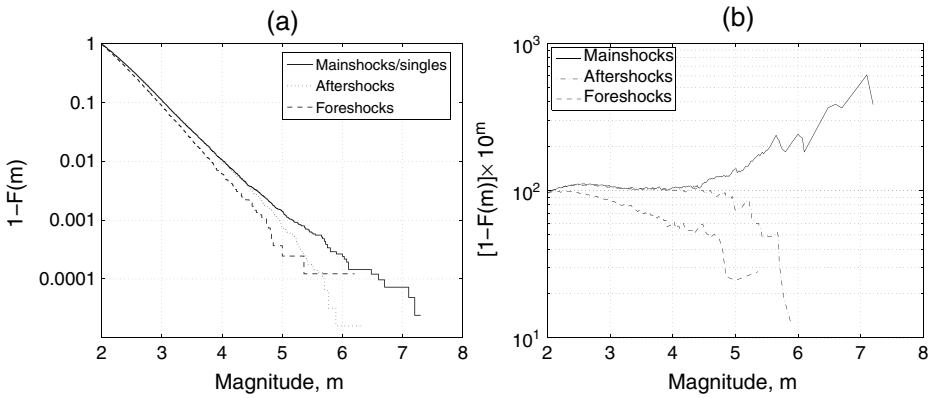


**Figure 10.** Magnitude distribution for different event types. (a) Proportion 1-$F(m)$ of events with magnitude equal to or above $m$. (b) Normalized proportion $[1\text{-}F(m)]\times 10^m$ of events with magnitude equal or above $m$. Different event types are shown by different line styles, as indicated in the legend.

shows the proportion $1-F(m)$ of earthquakes with magnitude above or equal to $m$, where $F(m)$ denotes the empirical cumulative distribution function of magnitudes. Figure 10b shows the cumulative proportion normalized by the magnitude, $[1-F(m)]\times 10^m$. This transformation is convenient to emphasize changes of the exponential index: It converts a pure exponential distribution $F(m)=1-10^{-bm}$ with $b=1$ to a horizontal line; a downward slope indicates an exponential distribution with index $b>1$, while an upward slope indicates an exponential distribution with index $b<1$. The following observations are noteworthy. (1) There exists a downward bend in all three distributions within the magnitude range $2\le m\le 2.5$; the bend reflects the catalog incompleteness. (2) All three distributions are approximately exponential (i.e., are closely approximated by a straight line in the chosen coordinates) within the range $2.5\le m\le 4.5$ with index $b\approx 1$ for mainshocks/singles and aftershocks and a larger index $b>1$ for foreshocks. (3) There is a prominent upward (downward) bend in the mainshock/singles (aftershock) distribution at $m=4.5$; there is also a less prominent downward bend for foreshocks at the same point.

[38] Table 3 reports the estimation of the $b$-value for events of different types, using events with magnitude $m\ge 3.0$ to eliminate the effects of catalog incompleteness reported above. The estimations are done with the method of *Tinti and Mulargia* [1987] that takes into account the discreteness of reported magnitudes; the magnitude step in the analyzed catalog is $\delta=0.01$. We refer to *Marzocchi and Sandri* [2003] for comprehensive discussion and tests of this method. The estimations confirm the visual impression from the cumulative plots of Figure 10: The $b$-values for mainshocks/singles and aftershocks are the same within the intermediate magnitude range and are significantly lower than the $b$-value for foreshocks. While the estimated $b$-values and respective confidence intervals would depend on the employed estimation method (different methods may lead to deviations of estimated values within $\pm 0.03$), this qualitative conclusion remains the same. We also notice that despite the fact that the observed deviations of the magnitude distribution from a pure exponential law at large values do affect slightly the estimated $b$-values, they are not large enough to mask the special behavior of foreshocks.

[39] The location of the bend at $m=4.5$ is consistent with results of *Knopoff* [2000], who reported an upward bend for the mainshocks in southern California at $m=4.7$ using a different method for mainshock identification. This effect is also observed in the ETAS model (supporting information section D and Figure D9) and hence is due at least in part to the finite spatio-temporal domain of analysis and conditional event type definition: larger events have a slightly higher chance of becoming mainshocks than smaller events. The upward deviation in the number of mainshock/singles from the pure exponential law is also an expected outcome of the growth of stress concentration in elastic solid with the rupture size [*Ben-Zion and Rice*, 1993; *Ben-Zion*, 1996]. When ruptures reach a critical size $R_c$ for which the stress transfer to the edge is comparable to a typical stress drop, they can generate at the propagating front sufficient stress to continue to propagate through areas that just sustained a stress drop. Such ruptures can become "runaway events" that continue to grow (statistically) to a size limited by strong heterogeneities or the overall fault dimensions [*Ben-Zion*, 2008, Figure 13]. The actual value of $R_c$ depends on the level of heterogeneities and average stress drops. The results of *Knopoff* [2000] and those shown in Figure 10 might indicate that $m\approx 4.5$ is sufficiently large to produce statistically runway events in southern California. The enhancement in the number of mainshocks for events larger than $m\approx 4.5$ leads to corresponding reduction in the number of other event types (foreshocks and aftershocks), as shown in Table 1. The precise origin of the observed upward deviation in the mainshock magnitude distribution remains unclear and may be due to either physical processes or statistical artifacts, or a mixture of the two.

**Table 3.** Estimated $b$-Value for Events of Different Type for $m\ge 3$, Using Estimator of *Tinti and Mulargia* [1987]

| | $b$-Value | 95% CI | $n$ |
|---|---|---|---|
| All | 0.992 | 0.974 – 1.010 | 12,105 |
| Mainshocks | 1.003 | 0.974 – 1.032 | 4,625 |
| Aftershocks | 1.004 | 0.980 – 1.028 | 6,759 |
| Foreshocks | 1.129 | 1.047 – 1.211 | 721 |

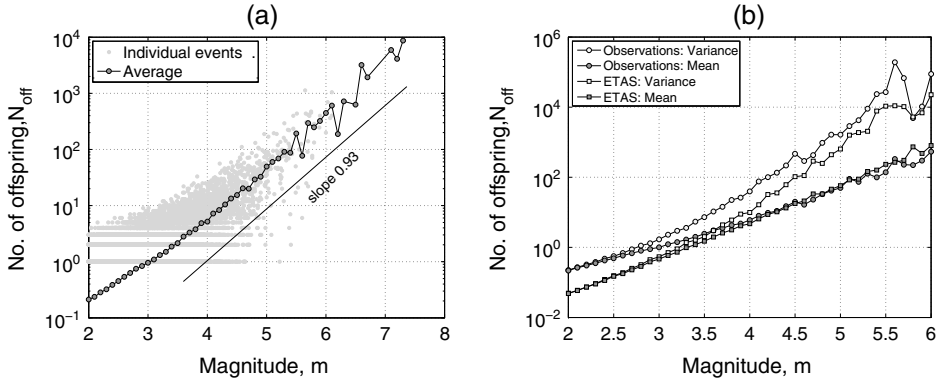


**Figure 11.** Number $N_{\text{off}}$ of direct offspring. (a) Individual offspring numbers (light circles) for all events in the catalog and the averaged offspring numbers (dark circles) within nonoverlapping magnitude intervals of length 0.1. The line has slope 0.93 that corresponds to the scaling of the average offspring numbers within the interval $4 \le m \le 6$. (b) Average (dark) and variance (light) for the offspring number for the observed catalog (circles) and ETAS model (squares). Each point represents all events within magnitude interval of length 0.1.

[40] Importantly, there exist two observations that are not reproduced by the ETAS catalogs considered in this study (supporting information section D, Figure D9, and Table D3): (1) comparable *b*-values for aftershocks and mainshocks within intermediate magnitude range and (2) higher *b*-value for foreshocks than for aftershocks and mainshocks. This suggests that the behavior of natural seismicity clusters, and foreshocks in particular, cannot be completely explained within the ETAS framework; or at least by its spatially homogeneous version used here. Evidently, there are features that are not due to the conditional definition of events; these features might reflect important physical processes within earthquake clusters.

### 4.2. Number of Offspring

[41] Figure 11a shows the estimated number $N_{\text{off}}$ of direct offspring for each event in the observed catalog (light dots) and the averaged offspring number within nonoverlapping magnitude intervals of length 0.1 (dark circles). The offspring number scales with the event magnitude as

$$N_{\text{off}} \propto 10^{cm}, c = 0.93 \pm 0.06.$$

[42] Here the estimation is done within the range $4 \le m \le 6$, and the margins of error correspond to a 95% CI. The distribution of $N_{\text{off}}$ for fixed $m$ can be closely approximated by the negative binomial distribution, and it deviates significantly from a Poisson distribution. This is consistent with results of *Kagan* [2010] and is confirmed by the chi-square goodness-of-fit test for $2 \le m \le 6$ with magnitude step 0.1 (not shown). The analysis for $m = 4.5$ is illustrated in Figure 12a, which shows the empirical cumulative distribution of the offspring number (circles) and its best maximum likelihood approximations by the negative binomial (solid line) and Poisson (dashed line) distributions. Clearly, the negative binomial model provides a very close fit, while the Poisson model is not applicable. Similar results are seen at all other magnitudes. For comparison, Figure 12b repeats the same analysis in an ETAS catalog with 146,432 events described in supporting information section D.5. Although the actual offspring numbers in the ETAS model have a Poisson distribution, the estimated numbers of direct offspring have larger variance and are better approximated by a negative binomial distribution; this issue is further discussed in supporting information section D.5. An important observation is that the variance of the offspring distribution in the observed catalog seems to be much larger than that in the considered ETAS model. This observation is further confirmed in Figure 11b where we show the

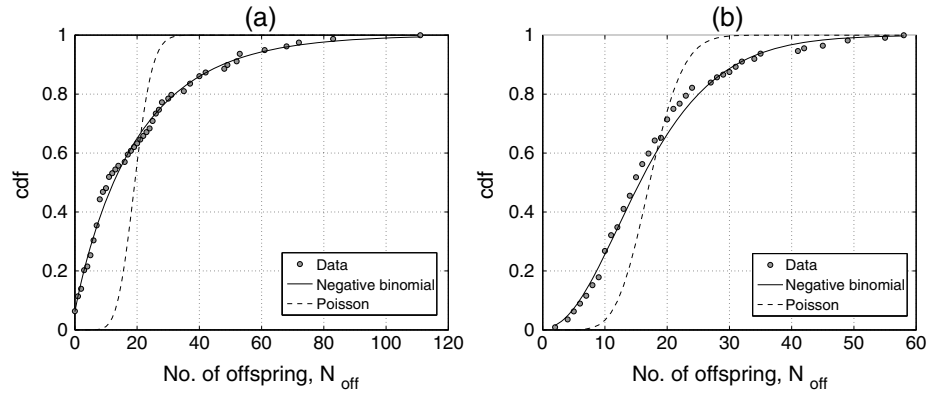


**Figure 12.** Distribution of the offspring number $N_{\text{off}}$ in the observed catalog (a) and ETAS model (b). The empirical cumulative distribution function (filled circles) is juxtaposed with the best maximum likelihood Poisson (dashed line) and negative binomial (solid line) models.

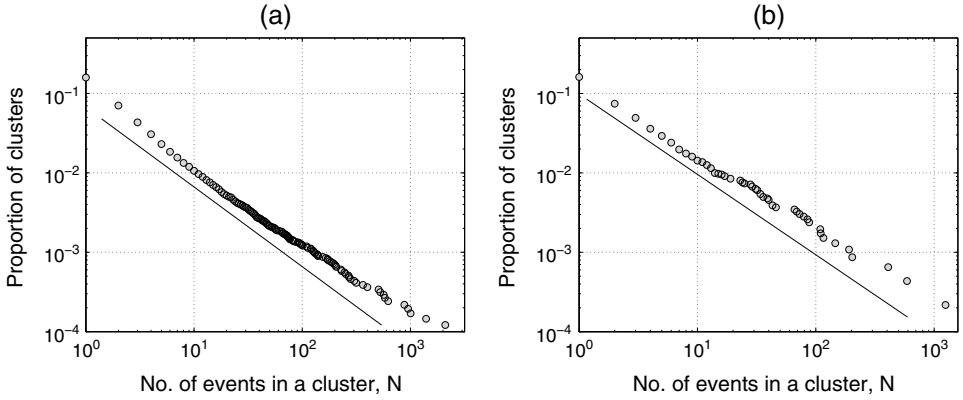


**Figure 13.** Distribution of the cluster size $N$ in the nearest-neighbor forest (Figure 5b). The figure shows the tail distribution function $1-F(N)$ versus $N$; this plot should be linear for a Pareto distribution. Evidently, Pareto distribution provides a reasonable approximation to the data, independently of the magnitude cutoff for the analysis. Each panel shows a line with slope $-1$. Different panels correspond to different lower magnitude cutoffs used in the nearest-neighbor analysis: (a) $m \geq 2$ and (b) $m \geq 3$.

estimated mean and variance of the offspring number $N_{\text{off}}$ for the observed and ETAS catalogs. We note the following features: (1) The average offspring number for $m > 4$ is the same in the ETAS and observed catalogs. (2) On the other hand, the average offspring number for $m < 4$ is significantly smaller in the ETAS results than in the observed catalog. (3) The variance of $N_{\text{off}}$ in the observed catalog is always larger than that in the ETAS catalog. (4) The ratio between the variance and average increases from 1 to about 100 as the magnitude $m$ increases from 2 to 6 in both the ETAS and observed catalogs; this further emphasizes inappropriateness of the Poisson model, for which the ratio is unity.

[43] An important consequence of the increased variability of the offspring number in the observed catalog compared to the ETAS model is the existence of a large population of *singles*—mainshocks with no offspring (Tables 1 and 2). The existence of singles cannot be explained solely by catalog artifacts such as incompleteness or the minimal reported magnitudes. Singles comprise 84% of all detected clusters (as opposed to 31% of all events, as reported in Table 1), 53% of the clusters with $m \geq 3$, and 17% of the clusters with $m \geq 4$. The largest single has magnitude 5.0.

### 4.3. Cluster Size, Number of Foreshocks, Aftershocks

[44] The distribution of cluster size $N$ is shown in Figure 13; it can be closely approximated by a Pareto distribution with index $a \approx -1$. Recall that the Pareto cumulative distribution function can be written as $F(x) = 1 - Ax^{-a}$, with $x \geq A^{1/a}$ for some $A$, $a > 0$. This is equivalent to $1 - F(x) = Ax^{-a}$, or $\log[1 - F(x)] = -a \log[x] + \log A$. Hence, the log-log plot of the tail function $1 - F(x)$ versus $x$ for the Pareto distribution is linear. This is seen in Figure 13 for two different minimal magnitudes of the nearest-neighbor cluster analysis. The approximate Pareto distribution of cluster sizes with index $a \approx -1$ is reproduced in the analysis of the ETAS model (Figure D10b) and can be readily explained by the combination of exponential mainshock magnitude distribution with a given $b$-value (index) and exponential number of offspring for a given mainshock with index $\alpha \approx b$ [*Saichev et al.*, 2005]. We emphasize that Figure 13 merely illustrates that the cluster size distribution can be approximated by a Pareto law; it is not intended to validate or invalidate a more delicate assumption about the precise equality of $\alpha$ and $b$, which has no effect on the results of this study.

[45] Figure 14 illustrates the distribution of the number of aftershocks and foreshocks per cluster. This analysis includes families with no aftershocks and/or foreshocks as well as singles; this makes it possible for the average number of fore/aftershocks to be less than 1. Figure 14a shows the number $N$-1 of the foreshocks and aftershocks in a cluster as a function of the cluster mainshock magnitude $m$. The total number of foreshocks and aftershocks scales as ($N$-1) $\propto 10^{\beta m}$ with $\beta = 0.95 \pm 0.06$ (95% CI); a line with slope 0.95 is shown for visual convenience. The same analysis is done separately for aftershocks and foreshocks in Figure 14b. The number of aftershocks (circles) per cluster is much larger than that of foreshocks (diamonds). The number $N_{\text{A}}$ of aftershocks still scales with the mainshock magnitude as $N_{\text{A}} \propto 10^{\beta m}$ with $\beta \approx 0.99 \pm 0.06$ (95% CI). The number of foreshocks does not exhibit a clear exponential scaling; nevertheless, the best exponential fit would have index $\beta \approx 0.6$ that is smaller than that for aftershocks. The estimation of slopes is done here within the magnitude range $m \geq 4$; the slope for the lower magnitudes is slightly larger in both analyzes. Similar productivity distribution is observed in the ETAS model (supporting information section D and Figure D10a).

[46] The observed increase of the fore/aftershock number with the cluster mainshock magnitude is ultimately caused by the existence of the catalog lower cutoff magnitude. To demonstrate this, we examine in Figure 15 the cluster size $N$ for different mainshock magnitudes in the $\Delta$-analysis, which only considers foreshocks and aftershocks within $\Delta$=2 magnitude units from the respective mainshock. The cluster size seems to be independent of the mainshock magnitude; this visual impression is confirmed by the ANOVA [*Freedman*, 2005] and Kruskal-Wallis [*Kruskal and Wallis*, 1952] tests summarized in Table 4. Specifically, we test the null hypothesis $H_0$: *Cluster sizes have the same mean (ANOVA) or median (Kruskal-Wallis) in different groups according to the cluster mainshock magnitude*. We run three

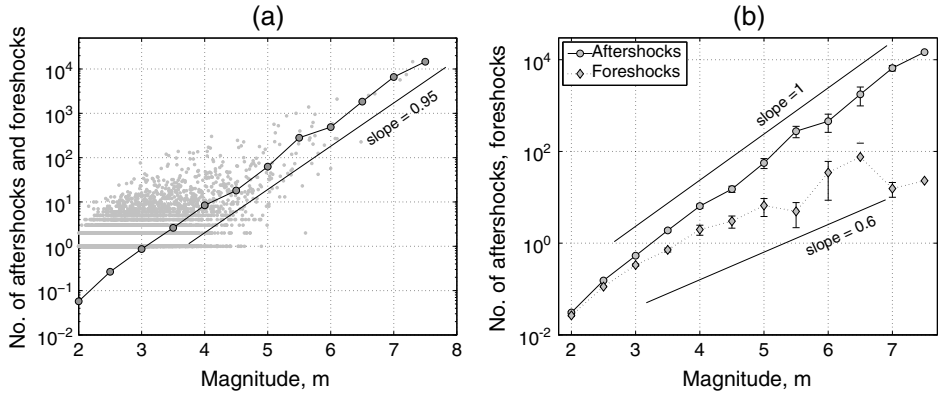


**Figure 14.** Number of aftershocks and foreshocks in a cluster as a function of the cluster mainshock magnitude $m$. (a) The average total number of aftershocks and foreshocks, $N$-1, per cluster. The small grey dots represent individual clusters; the filled circles show the average number of offspring within the nearest 0.5 magnitude range. The best least-square fit to the averages values within the range $m \geq 4$ has the slope of 0.95. (b) The average number of aftershocks (circles) and foreshocks (diamonds) per cluster. Singles have zero aftershocks and foreshocks; this makes it possible for the average to be less than unity.

series of tests, each of which corresponds to three consecutive lines of the table. Each test is done (1) using ANOVA approach (columns 4 and 5) and (2) using Kruskal-Wallis approach (columns 6 and 7). In ANOVA tests, we use $\log_{10}(N)$ to better satisfy the assumption of sample Normality; the Kruskal-Wallis test refers to the sample median and hence gives the same results for $N$ and $\log_{10}(N)$. The first series of tests (lines 1–3) compares the cluster size $N$ among several groups of clusters binned into equidistant mainshock magnitude intervals. The null hypothesis cannot be rejected if the magnitude bins have lengths below 1 (lines 2 and 3). For longer magnitude bins (line 1), the null is rejected. A more detailed analysis (not shown) suggests that the reported rejection is due to a slight decrease of the cluster size at the lowest mainshock magnitudes. The observed decrease is in fact caused by the singles ($N = 1$), which possibly form a particular population as discussed in section 4.2. This is demonstrated in the second series of tests (lines 4–6) that excludes the singles from the analysis; i.e., it considers only earthquake families. The null cannot be rejected in any of the six tests. Binning of clusters into equidistance magnitude intervals (as is done in the above tests) results in samples of significantly different sizes, which may affect the testing results. To avoid this effect, the third series of tests (lines 7–9) compares the family sizes (excluding singles) in bins of equal size. The null hypothesis cannot be rejected in any of the six tests. These results suggest that the observed cluster structure is robust with respect to the earthquake magnitudes; namely, the family size depends only on the magnitude difference of mainshock and other family events, and not on their absolute values. We also provide additional support to the hypothesis that singles form a special population that should be analyzed separately from the families.

### 4.4. Temporal Structure of Families

[47] Figure 16 shows the intensity of events in the clusters vs. time, averaged over all the detected clusters in regular and Δ-analyses. Both analyses recover the conventional structure of a foreshock-mainshock-aftershock sequence, with a lower number of foreshocks and higher number of aftershocks. Both aftershock and foreshock intensities decrease away from the time of mainshock. This is further illustrated in Figure 17 that shows a power-law decay of both aftershock and foreshock intensities away from the time of the mainshock. Specifically, the figure shows results for families with mainshock magnitude $m \geq 4$, for Δ-aftershocks (Figure 17a, black dots) and Δ-foreshocks (Figure 17b), with $\Delta = 2$ and within 50 days from the mainshock. Figure 17a also shows (light squares) the intensity of first-generation offspring for parents with magnitude $m \geq 4$, within Δ=2 magnitude units from the parent. The results are consistent with those obtained in the ETAS model (cf. supporting information section D and Figure D11): The slopes for both aftershock and foreshock decays are comparable, and the slope for aftershock decay is lower than that for the first-generation offspring, which is due to the existence of secondary, tertiary, *etc*. aftershocks. The intensity slopes remain the same (within statistical margins of error) if the analysis is done separately for different mainshock

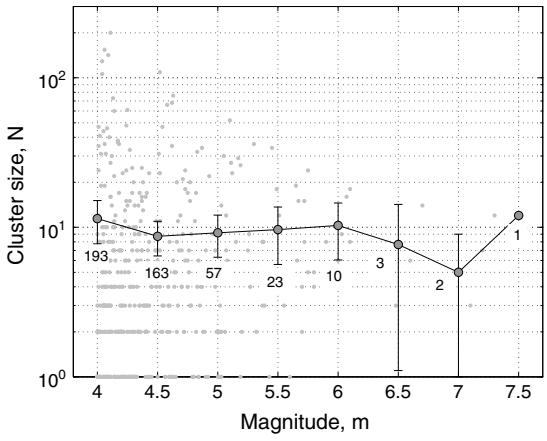


**Figure 15.** Cluster size $N$ for clusters with different main shock magnitude in Δ-analysis. Small grey dots represent individual clusters; filled circles show average cluster size for different magnitude groups; error bars indicate the 95% confidence interval for the mean. The number of clusters in each group is indicated in the figure.

**Table 4.** ANOVA and Kruskal-Wallis Tests of the Hypothesis $H_0$: *The Size of a Δ-Cluster Is Independent of the Cluster Magnitude*

| Experiment | Magnitude Bin Size | No. of Bins | ANOVA | | Kruskal-Wallis | |
|---|---|---|---|---|---|---|
| | | | *P*-value | Reject $H_0$ at 5% Level? | *P*-value | Reject $H_0$ at 5% Level? |
| Bins by magnitude, all clusters | 1 | 4 | 0.03 | Yes | 0.006 | Yes |
| | 0.5 | 8 | 0.35 | No | 0.08 | No |
| | 0.25 | 13 | 0.47 | No | 0.13 | No |
| Bins by magnitude, no singles | 1 | 4 | 0.62 | No | 0.32 | No |
| | 0.5 | 8 | 0.96 | No | 0.80 | No |
| | 0.25 | 13 | 0.97 | No | 0.77 | No |
| Equisized bins, no singles | - | 2 | 0.87 | No | 0.43 | No |
| | - | 5 | 0.72 | No | 0.61 | No |
| | - | 15 | 0.52 | No | 0.46 | No |

magnitudes (not shown). The same analysis for regular aftershocks and foreshocks (as opposed to Δ-aftershocks/foreshocks) leads to different decay slopes for different mainshock magnitudes, with significant deflation of the estimated slopes at large mainshock magnitudes (not shown); this is explained by significant increase of secondary, etc. aftershocks as well as incomplete registration of small events in the vicinity of a large one.

[48] The above results are consistent with the Omori-Utsu law for the intensity of both aftershocks and foreshocks [*Omori*, 1894; *Utsu*, 1961; *Papazachos*, 1973; *Jones and Molnar*, 1979; *Utsu et al*., 1995; *Helmstetter et al*., 2003]:

$$\Lambda(t) = \frac{K}{(t+c)^p}. \tag{3}$$

[49] Our observations indicate that $p = 0.82 \pm 0.04$ for Δ-aftershocks within $0.1 \le t \le 10$ days, $p = 1.02 \pm 0.02$ for the first-generation offspring within $0.01 \le t \le 10$ days; and $p = 0.89 \pm 0.13$ for Δ-foreshocks within $0.1 \le t \le 10$ days. All margins of errors refer to a 95% confidence interval. Recall that the productivity index $K$ of fore/aftershocks in regular analysis (when all events reported in a catalog are considered in the analysis) scales with the mainshock magnitude $m$ as

$$K = 10^{\beta m}, \tag{4}$$

with $\beta \approx 1$ for aftershocks and $\beta < 1$ for foreshocks (Figure 14). In the Δ-analysis (which only includes aftershocks and foreshocks within Δ magnitude units from the respective mainshock), the productivity $K = K(\Delta)$ is a constant that depends on Δ but not on the mainshock magnitude. These results are consistent with the existing knowledge about properties of aftershocks and foreshocks [*Utsu*, 1961; *Papazachos*, 1973; *Jones and Molnar*, 1979; *Kisslinger and Jones*, 1991; *Utsu et al*., 1995; *Helmstetter and Sornette*, 2002; *Helmstetter et al*., 2003].

[50] The following observation from Figure 16 is noteworthy: The aftershock intensity in the regular analysis is order of magnitude larger than that in the Δ-analysis (as expected). In contrast, however, the difference between the foreshock intensities in the regular and Δ-analyses is much smaller. A more focused examination of this is illustrated in Figure 18a that shows the distribution of the magnitude differences $d_m = m_{\text{mainshock}} - m_{\text{event}}$ of family events and their respective mainshocks. The foreshock magnitudes are prominently closer to that of the mainshock compared to the aftershock magnitudes. Accordingly, more foreshocks remain in the family in the Δ-analysis. The observed difference between aftershocks and foreshocks magnitude distributions cannot be explained by the conditional definition of event types. This is illustrated in supporting information section D (Figure D12a) by analysis of the ETAS model. Interestingly, as shown in Figure 18b, the distributions of the difference between the magnitude of the mainshock and the *largest* foreshock or aftershock are statistically the same. This observation is discussed further in the next section. The difference between the relative numbers of foreshocks and aftershocks, and their magnitude differences from the mainshocks, are consistent with statistical existence of

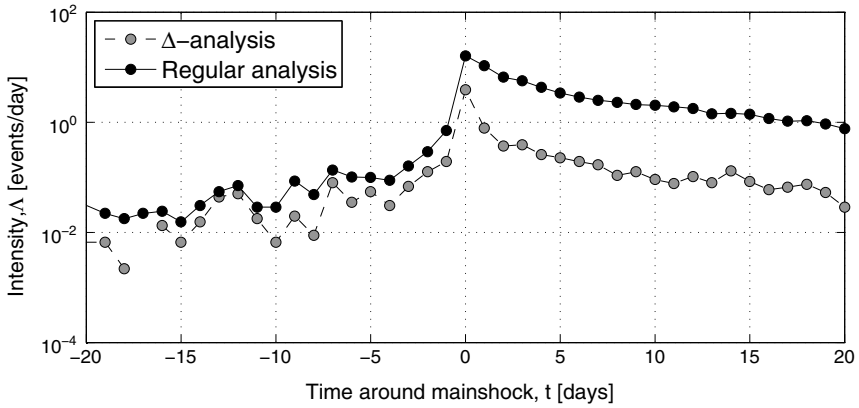


**Figure 16.** Intensity of events in a cluster, in events per day per cluster; mainshock is placed at $t = 0$. The figure shows results for all clusters, including singles, with mainshock magnitude $m \ge 4$.

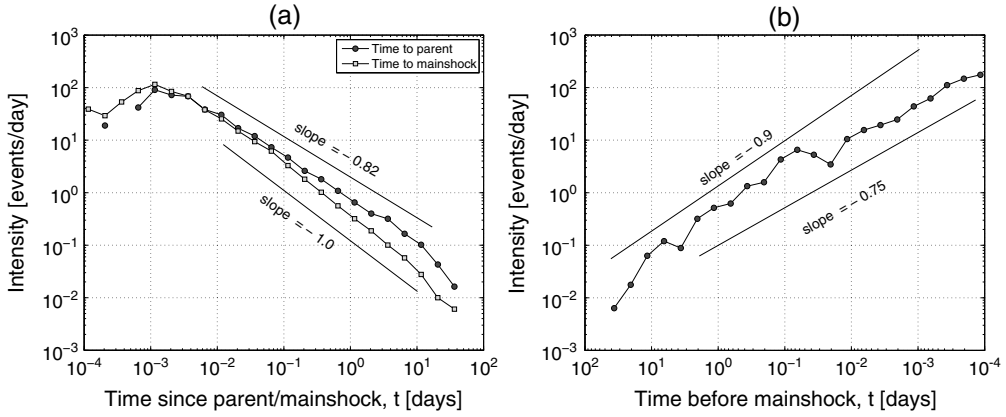


**Figure 17.** Intensity of (a) aftershocks and (b) foreshocks in events per day per family. The figure shows results for families with mainshock magnitude $m \geq 4$, for Δ-aftershocks and foreshocks, with $\Delta = 2$, and within 50 days from the mainshock. Figure 17a also shows (light squares) the intensity of the first-generation offspring for parents with magnitude $m \geq 4$, within Δ=2 magnitude units from the parent.

some accelerated failure process as the time of mainshocks is approached [e.g., *Mogi*, 1969; *Keilis-Borok and Kossobokov*, 1990; *Bufe and Varnes* 1993; *Ben-Zion and Lyakhovsky*, 2002; *Turcotte et al*., 2003].

### 4.5. Båth Law for Foreshocks and Aftershocks

[51] It has been observed in analysis of aftershock sequences of large earthquakes [e.g., *Båth*, 1965; *Kisslinger and Jones*, 1991; *Shcherbakov and Turcotte*, 2004; *Shearer*, 2012] that there is a systematic difference between the magnitudes of a mainshock and the largest aftershock. The value of the magnitude difference reported in the literature is close to 1.2 and is independent of the mainshock magnitude. Our analysis confirms (Figure 18b) the Båth law for aftershocks, as well as for foreshocks, with average magnitude differences of 1.1 and 1.2, respectively. Interestingly, the distribution of the differences $\Delta_m = m_{\text{mainshock}} - m_{\text{largest-event}}$ is almost uniform for both aftershocks and foreshocks within the range [0, 2], with a few sporadic values above 2 that correspond to a fast decaying tail of the distribution. The uniform range [0, 2] seems to be independent of the magnitudes of the examined families, as illustrated by the aftershock analysis of families with mainshock magnitude $m \geq 5$ shown in the inset of Figure 18b. The uniform distribution within the range [0, 2], with a fast decaying tail at larger values, explains the mean value of the magnitude difference in the Båth law that is slightly larger than 1 [e.g., *Ben-Zion*, 2008]. An almost uniform distribution of $\Delta_m$ or aftershocks is also observed in the ETAS model (supporting information section D and Figure D12b). However, the magnitude distribution for the foreshocks in the examined ETAS models always exhibits significant deviations from a uniform $\Delta_m$. The data examined in this study does not allow resolving whether the ETAS model has systematic deviations from observations in regards to the foreshock magnitude distribution. This should be tested further using additional observed and ETAS catalogs.

### 4.6. Area and Duration of Families

[52] We define the area $A$ of a group of earthquakes as the area of the minimal convex hull that contains these earthquakes. The area of foreshock and aftershock sequences in

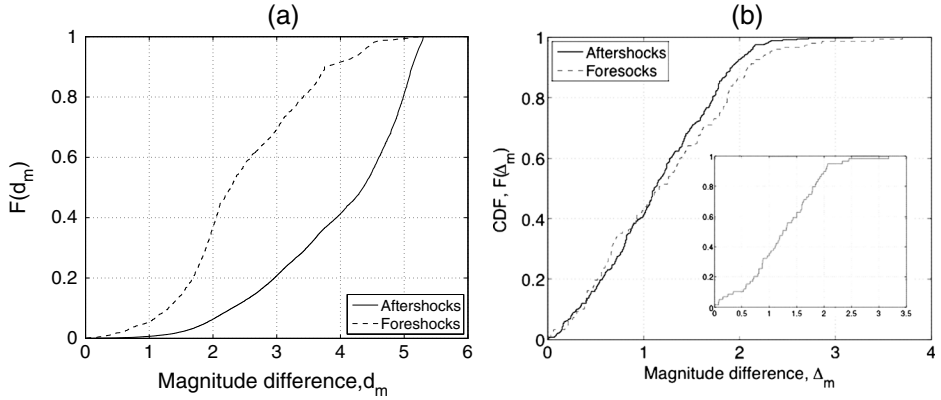


**Figure 18.** Magnitude of family events relative to the family mainshock magnitude. (a) Difference $d_m$ between magnitude of the family mainshock and each foreshock (dashed line) and aftershock (solid line). (b) Difference $\Delta_m$ between magnitudes of the family mainshock and the largest foreshocks (dashed line) and aftershocks (solid line). Only families with mainshock magnitude $m \geq 4$ are analyzed in both panels. The inset in Figure 18b shows the aftershock analysis for families with mainshock magnitude $m \geq 5$.

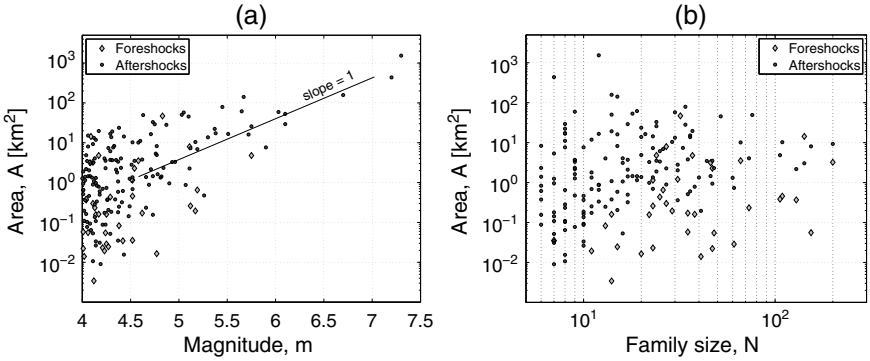


**Figure 19.** Area $A$ occupied by foreshocks (diamonds) and aftershocks (circles) vs. family mainshock magnitude m (panel a) and family size $N$ (panel b). The figure shows 147 aftershocks series and 38 foreshock series in Δ-analysis with at least five events.

Δ-analysis is illustrated in Figure 19. The analysis is done for 147 aftershock sequences and 38 foreshock sequences with at least five events. The area $A$ scales with the family mainshock magnitude $m$ as $A \propto 10^{\gamma m}$, $\gamma \approx 1$ (Figure 19a) and is independent of the family size $N$ (Figure 19b). This suggests the existence of a *damage zone* around the parent rupture with a linear size that scales with the mainshock magnitude $m$ as $10^{\lambda m}$, $\lambda \approx 0.5$. These results are consistent with the empirical scaling relation $\log_{10} P_0 = \alpha M_L + const$, where $P_0$ is the scalar seismic potency defined as $A$ times slip $\Delta u$, $\alpha \approx 1.5$, and $\Delta u \propto A^{1/2}$ for crack-like ruptures [e.g. *Ben-Zion*, 2008; *Kanamori and Anderson*, 1975]. The observations show that the foreshock area, on average, is order of magnitude smaller than the aftershock area, independent of the family size $N$.

[53] We also observed (not shown) that the family duration $D$ is independent of the family mainshock magnitude and seems to slightly increase with the family size. This suggests a magnitude-independent mechanism of stress relaxation after a mainshock. The results are consistent with a process that is dominated by elastic (rather than viscous) stress transfer. The slight increase of duration with family size is related to the fact that larger families tend to be concentrated in relatively hot areas where viscous processes play a larger role. This issue will be illustrated in more detail in a subsequent paper.

## 5. Discussion

[54] We present a statistical methodology for detecting earthquake clusters based on the generally observed bimodal distribution of nearest-neighbor distances in a combined space-time-size domain [*Zaliapin et al*., 2008; *Zaliapin and Ben-Zion*, 2011; *Hicks*, 2011; *Bautista*, 2011; *Mignan*, 2012; C. Gu et al., Triggering cascades and statistical properties of aftershocks, in review, 2012] and apply the bimodality to identify systematically individual seismicity clusters in a relocated catalog of 111,981 events with magnitudes $m \geq 2$ in southern California [*Hauksson et al*., 2012]. The following features of the proposed technique are noteworthy. *Soft parameterization*: the algorithm uses only three numeric parameters that can be closely estimated from observations; variations of the parameters within wide limits, largely exceeding their statistical variability, do not seriously affect the cluster identification. *Stability*: The technique is stable with respect to location errors, minimal reported magnitude, and catalog incompleteness. *Absence of an underlying model*: The technique does not assume any particular form of earthquake clustering. The latter is the case, for instance, in the cluster approach of *Zhuang et al*. [2002] that is based on the assumption that natural seismicity is fully accounted for by the ETAS model (supporting information Section D). *Absence of ad-hoc rules*: The technique is self-adapted to observed seismicity; it does not use any expert-defined thresholds or tuning parameters. The latter is the case for the classical window method of *Gardner and Knopoff* [1974] and its ramifications.

[55] The iterative approach of *Marsan and Lengline* [2008], which is based on a softly-parameterized technique with no *apriori* model of clustering, may be considered similar in spirit to but is quite different than the statistical nearest-neighbor method used in this study. A particular advantage of our approach is that the division of the examined seismicity into *background* events and *clustered* events is governed by the intrinsic and clearly seen bimodal distribution of earthquake distances (e.g., Figures 4 and D3). The bimodal distribution is observed in various regional and global catalogs [*Zaliapin et al*., 2008; *Zaliapin and Ben-Zion*, 2011; *Hicks*, 2011; *Bautista*, 2011; *Mignan*, 2012; Gu et al., in review, 2012], as well as in the ETAS model (supporting information section D) [*Zaliapin et al*., 2008; Gu et al., in review, 2012]. We have shown in supporting information section B that the bimodal distribution of earthquake distances cannot be explained by independent spatial or temporal catalog inhomogeneities but is rather a feature of *local dependent* space-time structures.

[56] We define *clustered* events as groups of events with abnormally short space-time distances to their nearest neighbors (the lower-left mode of the bimodal distribution of Figures 4a and 4b) and *background* seismicity as the collection of events that do not contribute to the clustered mode (the upper right mode of the bimodal distribution of Figures 4a and 4b). The existence of clustered events suggests a natural way of identifying *individual clusters* (section 3) that are the main subject of this study. The background events, by construction, correspond to a (possibly) nonstationary inhomogeneous Poisson process with independent time and space marginals. This is consistent with the recent results of *Luen and Stark* [2012] that declustered catalogs in southern California cannot be approximated by a stationary inhomogeneous Poisson process.

[57] The quality and stability of the proposed cluster detection algorithm are demonstrated using synthetic and

observed catalogs in supporting information sections D and E. In particular, we show that our method closely reconstructs the cluster structure of a spatially isotropic ETAS model (supporting information section D), simulated as a branching process with well-defined (although not directly observed) parent-child attributions. There exists an alternative interpretation of the ETAS model as a marked point process with the conditional intensity of equation (D1); in this interpretation the model has no explicit cluster structure. Noticeably, our technique reveals the existence of a statistically significant population of *clustered events* in synthetic ETAS catalogs (Figure D3), and it does not rely on a particular interpretation of the ETAS model. We emphasize that this general observation (existence of clustered events) is independent of a particular way of forming individual clusters.

[58] We analyze in detail the statistical properties of the detected individual clusters and compare our findings with related available results in the literature. Importantly, the event identification based here on purely statistical analysis is consistent with the traditional identification of different types of earthquakes for the largest most conspicuous clusters. In particular, we demonstrate the validity of the Omori-Utsu (Figures 14, 16, and 17) and Båth (Figure 18b) laws for aftershocks [e.g., *Omori*, 1989; *Utsu et al*., 1995] in the detected clusters, and confirm that the same laws hold for foreshocks [e.g., *Jones and Molnar*, 1979; *Helmstetter et al*., 2003]. These laws are observed clearly for foreshocks only when data of many sequences are stacked [e.g., *Papazachos*, 1973; *Jones and Molnar*, 1979], given the small number of foreshocks in individual sequences. Our results, furthermore, point to an origin of the Båth law related to a specific distribution (Figure 18b) of magnitude differences between the mainshock and the largest aftershock/foreshock; this distribution is almost uniform for differences between 0 and 2 and rapidly decays for larger values. To the best of our knowledge, the existing statistical explanations of the Bath law [e.g., *Helmstetter and Sornette*, 2003; *Saichev and Sornette*, 2005; *Vere-Jones*, 2008] do not refer to this particular form of the magnitude difference distribution.

[59] The spatial extent of clusters is shown to increase with mainshock magnitude in agreement with empirical potency-magnitude scaling relation (Figure 19a). On the other hand, the other properties of the clusters structure (except area) depend on the *magnitude difference* of an event and its mainshock but not on their absolute values (e.g., Figure 15), and there is weak-to-no dependency of cluster durations on the mainshock magnitude. We observe a power law distribution of the number of events in the clusters, with dominance of the single-event clusters (Figure 13), which may be a general feature of seismicity. In section 4.2, we discuss the existence of significant population of singles – mainshocks with no foreshocks or aftershocks, which cannot be completely explained by catalog incompleteness or minimal reported magnitude, and hence presents an interesting feature of the data. Finally, the observations include statistical evidence for an accelerated failure process before mainshocks. This is associated with increasing rate of foreshocks (Figures 16 and 17) and smaller magnitude differences between foreshocks and mainshocks than between aftershocks and mainshocks (Figure 18). Clarifying the details of this process will be done in a dedicated future study.

[60] The results of this study provide a foundation for a more focused analysis of the structure of the detected clusters that is performed in a companion paper [*Zaliapin and Ben-Zion*, 2013]. The analysis done in that paper demonstrates the existence of three dominant cluster types, corresponding generally to singles, burst-like and swarm-like sequences, and that the largest mainshocks are associated with sequences that are likely a mixture of these three basic types.

[61] **Acknowledgments.** We are grateful to Rick Schoenberg for sharing with us the R-code for simulating the ETAS model. The manuscript benefited from constructive comments of three anonymous referees. The study was supported by the Southern California Earthquake Center (based on NSF Cooperative Agreement EAR-0529922 and USGS Cooperative Agreement 07HQAC0008) and the National Science Foundation (grants EAR-0908903 and DMS-0934871). The SCEC contribution number for this paper is 1718.

## References

Agresti, A. (2002), *Categorical Data Analysis*, 2nd ed., p. 734, Wiley, New York.

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

Båth, M. (1965), Lateral inhomogeneities in the upper mantle, *Tectonophysics*, *2*, 483–514.

Bautista, J. (2011), Nearest-neighbor analysis of marked point fields with applications to cluster analysis of seismicity, Undergraduate Honors Thesis, Department of Mathematics and Statistics, University of Nevada, Reno, May, 2011.

Ben-Zion, Y. (1996), Stress, slip and earthquakes in models of complex single-fault systems incorporating brittle and creep deformations, *J. Geophys. Res.*, *101*, 5677–5706.

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

Ben-Zion, Y., and V. Lyakhovsky (2002), Accelerated seismic release and related aspects of seismicity patterns on earthquake faults, *Pure Appl. Geophys.*, *159*, 2385–2412.

Ben-Zion, Y., and J. R. Rice (1993), Earthquake failure sequences along a cellular fault zone in a three-dimensional elastic solid containing asperity and nonasperity regions, *J. Geophys. Res.*, *98*, 14109–14131.

Bufe, C. G., and D. J. Varnes (1993), Predictive modeling of the seismic cycle of the great San Francisco bay region, *J. Geophys. Res.*, *98*, 9871–9883.

Chu, A., F. Schoenberg, P. Bird, D. Jackson, and Y. Kagan (2011), Comparison of ETAS parameter estimates across different global tectonic zones, *Bull. Seismol. Soc. Am.*, *101*(5), 2323–2339.

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. 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. Seismol. Soc. Am.*, *99*, 3114–3123, doi:10.1785/0120080038.

Feller, W. (1970), *An Introduction to Probability Theory and Its Applications*, vol. *2*, 2nd ed., p. 669, Wiley, New York.

Felzer, K. R. (2008), Calculating California seismicity rates, in *The Uniform California Earthquake Rupture Foreast,* version 2 (UCERF 2), Appendix I, U.S. Geological Survey Open-File Report 2007-1437I and California Geological Survey Special Report 203I, 42 pp.,

Felzer, K. R., and E. E. Brodsky (2006), Decay of aftershock density with distance indicates triggering by dynamic stress, *Nature*, *441*, 735–738.

Freedman, D. A. (2005), *Statistical Models: Theory and Practice*, p. 246, Cambridge University Press, Cambridge, U.K.

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

Harte, D. (1998), Dimension estimates of earthquake epicenters and hypocenters, *J. Nonlinear Sci.*, *8*(6), 581–618.

Hastie, T., R. Tibshirani, and J. Friedman (2001), 8.5 The EM algorithm, in *The Elements of Statistical Learning*, pp. 236–243, Springer, New York.

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.

Helmstetter, A., and D. Sornette (2002), Subcritical and supercritical regimes in epidemic models of earthquake aftershocks, *J. Geophys. Res.*, *107*(B10), Art. No. 2237.

Helmstetter, A., and D. Sornette (2003), Bath's law derived from the Gutenberg-Richter law and from aftershock properties, *Geophys. Res. Lett.*, *30*(20), Art. No. 2069.

Helmstetter, A., D. Sornette, and J.-R. Grasso (2003), Mainshocks are aftershocks of conditional foreshocks: How do foreshock statistical properties emerge from aftershock laws, *J. Geophys. Res.*, *108*(B1), Art. No. 2046.

Hicks, A. (2011), Clustering in multidimensional spaces with applications to statistical analysis of earthquake clustering, MSc Thesis, Department of Mathematics and Statistics, University of Nevada, Reno, August, 2011.

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

Jones, L. M., and P. Molnar (1979), Some characteristics of foreshocks and their possible relationship to earthquake prediction and premonitory slip on fault, *J. Geophys. Res.*, *84*, 3596–3608.

Kagan, Y. (2007), Earthquake spatial distribution: The correlation dimension, *Geophys. J. Int.*, *168*(3), 1175–1194.

Kagan, Y. (2010), Statistical distributions of earthquake numbers: consequence of branching process, *Geophys. J. Int.*, *180*, 1313–1328.

Kanamori, H., and D. L. Anderson (1975), Theoretical basis of some empirical relations in seismology, *Bull. Seismol. Soc. Am.*, *65*, 1073–1095.

Keilis-Borok, V. I., and V. G. Kossobokov (1990), Premonitory activation of earthquake flow: Algorithm M8, *Phys. Earth Planet. Inter.*, *61*, 73–83.

Kisslinger, C. (1996), Aftershocks and fault-zone properties, *Adv. Geophys.*, *38*, 1–36.

Kisslinger, C., and L. M. Jones (1991), Properties of aftershock sequences in southern California, *J. Geophys. Res.*, *96*, 11,947–11,958.

Knopoff, L. (2000), The magnitude distribution of declustered earthquakes in southern California, *Proc. Natl. Acad. Sci.*, *97*(22), 11,880–11,884.

Kruskal, W., and W. A. Wallis (1952), Use of ranks in one-criterion variance analysis, *J. Am. Stat. Assoc.*, *47*(260), 583–621.

Luen, B., and P. B. Stark (2012), Poisson tests of declustered catalogs, *Geophys. J. Int.*, doi:10.1111/j.1365-246X.2012.05400.X.

Marsan, D., and O. Lengline (2008), Extending earthquakes' reach through cascading, *Science*, *319*(5866), 1076–1079, doi:10.1126/science.1148783.

Marzocchi, W., and L. Sandri (2003), A review and new insights on the estimation of the b-value and its uncertainty, *Ann. Geophys.*, *46*(6), 1271–1282

Mignan, A. (2012), Functional shape of the earthquake frequency-magnitude distribution and completeness magnitude, *J. Geophys. Res.*, *117*, B08302.

Mogi, K. (1969), Some features of recent seismic activity in and near Japan 2: Activity before and after great earthquakes, *Bull. Eq. Res. Inst. Univ. Tokyo*, *47*, 395–417.

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

Molchan, G., and T. Kronrod (2009), The fractal description of seismicity, *Geophys. J. Int.*, *179*(3), 1787–1799.

Ogata, Y. (1998), Space-time point-process models for earthquake occurrences, *Ann. Stat.*, *50*(2), 379–402.

Omori, F. (1894), On the aftershocks of earthquakes, *J. College Sci., Imperial Univ. Tokyo*, *7*, 111–200.

Papazachos, B. C. (1973), The time distribution of reservoir-associated foreshocks and its importance to the prediction of the principal shock, *Bull. Seismol. Soc. Am.*, *63*, 1973–1978.

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

Richter, C. F. (1958), *Elementary Seismology*, Freeman, San Francisco, pp. 768.

Romanowicz, B. (1993), Spatiotemporal patterns in the energy release of great earthquakes, *Science*, *260*, 1923–1926.

Saichev, A., and D. Sornette (2005), Distribution of the largest aftershocks in branching models of triggered seismicity: Theory of the universal Bath law, *Phys. Rev. E*, *71*(5), 056127.

Saichev, A., A. Helmstetter, and D. Sornette (2005), Power-law distributions of offspring and generation numbers in branching models of earthquake triggering, *Pure Appl. Geophys.*, *162*, 1113–1134.

Scholz, C. (2002), *The Mechanics of Earthquakes and Faulting*, p. 471, Cambridge Univ. Press, Cambridge U. K.

Schorlemmer, D., and J. Woessner, (2008), Probability of detecting an earthquake, *Seismol. Soc. Am.*, *98*, (5), 2103–2117.

Shcherbakov, R., and D. L. Turcotte (2004), A modified form of Bath's law, *Bull. Seismol. Soc. Am.*, *94*, 1968–1975.

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.

Tinti, S., and F. Mulargia (1987), Confidence intervals of b values for grouped magnitudes, Bulletin of the Seismological Society of America, *77* no. 6 p., 2125–2134.

Turcotte, D. L., W. I. Newman, and R. Shcherbakov (2003), Micro and macroscopic models of rock fracture, *Geophys. J. Int.*, *152*(3), 718–728.

Utsu, T. (1961), A statistical study of the occurrence of aftershocks, *Geophys. Mag.*, *30*, 521–605.

Utsu, T. (2002), Statistical features of seismology, in *International Handbook of Earthquake and Engineering Seismology, Part A*, edited by W. HK Lee, H. Kanamori, P. C. Jennings, and C. Kisslinger, pp. 719–732, Academic Press, London, U. K.

Utsu, T., Y. Ogata, and R. S. Matsu'ura (1995), The centenary of the Omori formula for a decay law of aftershock activity, *J. Phys. Earth*, *43*, 1–33.

Veen, A., and F. Schoenberg (2008), Estimation of space-time branching process models in seismology using an EM-type algorithm, *J. Am. Stat. Assoc.*, *103*, 614–624.

Vere-Jones, D. (2008), A limit theorem with application to Bath's law in seismology, *Adv. Appl. Probab.*, *40*(3), 882–896.

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

Vidale, J. E., and P. M. Shearer (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*, B05312.

Wang, Q., F. P. Schoenberg, and D. Jackson (2010), Standard errors of parameter estimates in the ETAS model, *Bull. Seismol. Soc. Am.*, *100*(5A), 1989–2001.

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

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. Int.*, *185*, 1288–1304, doi:10.1111/j.1365-246X.2011.04995.X.

Zaliapin, I., and Y. Ben-Zion (2013), Earthquake clusters in southern California II: Classification and relation to physical properties of the crust, *J. Geophys. Res.*, *Solid Earth*, *118*, doi:10.1002/jgrb.50178.

Auxiliary material for

# Earthquake clusters in southern California I: Identification and stability

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

[1]Department of Mathematics and Statistics, University of Nevada, Reno, 89557 (zal@unr.edu)
[2] Department of Earth Sciences, University of Southern California, Los Angeles, 90089-0740 (benzion@usc.edu)

*J. Geophys. Res.*

## Section A. *The earthquake distance: Motivation*

The definition (1) of the earthquake distance [*Baiesi and Paczuski*, 2004] is motivated by the intuitive expectation that the value of $\eta_{ij}$ should be small if earthquake $j$ might be related to earthquake $i$, and it should be larger if there is no relationship between earthquakes $i$ and $j$. To illustrate, consider a situation when $N(m)$ earthquakes with magnitude above $m$ happen independently of each other in $d_f$-dimensional space and time and obey the Gutenberg-Richter relation $\log_{10}N(m) = a - bm$. Then the expected number of earthquakes with magnitude $m$ within the time interval $t$ and distance $r$ from any given earthquake is proportional to $tr^{d_f}10^{-bm}$, which is an essential component of the definition (1). In other words, the distance (1) is the number (up to a constant) of earthquakes of magnitude $m$ that are expected within the time $t$ and distance $r$ from the earthquake $j$ in a process with no clustering. If the distance $\eta_{ij}$ is significantly smaller than most pair-wise distance within the catalog, this means that earthquake $j$ has happened abnormally close to $i$; this motivates one to consider $i$ as a parent for $j$. Naturally, this approach only reveals statistical, not causal, relationships between earthquakes. Figure A1 illustrates the connection between the normalized time $T$ (see Eq. (2) of the main text) and the calendar time in years.

## Section B. *The origin of the bimodal distribution of nearest-neighbor distances*

The goal of this section is to shed some light on the origin of the bimodal distribution of the nearest-neighbor distance shown in Fig. 4 of the main text. Comparison of the results for the observed seismicity (Fig. 4) with that for a homogeneous Poisson process (Fig. 3) suggests that the bimodality is related to earthquake clustering. There are several primary types of clustering in the catalogs: time-independent space clustering mainly related to the fault network geometry, space-independent time clustering related to (possible) global changes of seismic activity, and dependent space-time clustering mainly related to the foreshock-aftershock sequences or swarms. We demonstrate below that the cluster mode of the distribution in Fig. 4 *cannot* be explained by temporal or spatial clustering of earthquakes alone. The existence of this mode is ultimately caused by the *clusters* with dependent spatio-temporal structure that are due to the groups of earthquakes that happen within localized spatio-temporal regions; mainly to the foreshock-aftershock sequences or swarms.

Towards this goal, we consider three models of seismicity that retain the marginal spatial and/or temporal distributions of the real earthquakes while exhibiting no dependent spatio-temporal clustering. We start with the catalog of observed earthquakes with $m \geq 3$, which contains 12,105 earthquakes. The first randomized catalog is obtained by independent uniform random reshuffling of times and locations of the observed events. Reshuffling means that the event times $s_i$, $i = 1,\ldots, n,$ in the new catalog are obtained from the original times $t_i$, $i = 1,\ldots, n,$ as $s_i = t_{\sigma(i)}$, where $\sigma(i)$ denotes a uniform random permutation of the sequence $[1,\ldots,n]$. An independent reshuffling procedure is then applied to the epicenter locations $(\phi_i, \lambda_i)$. The time-latitude map of seismicity from this catalog is shown in Fig. B1a; the joint distribution $(T,R)$ of

the rescaled time and space components of the nearest-neighbor distance is shown in Fig. B2a. By construction, this randomized catalog has the same marginal time and space distributions as the observed seismicity. For instance, in Fig. B1a one can see significant variations of seismic activity along the latitude, which is related to the fault network geometry, as well as the most prominent time variations related to the aftershocks activity in the original catalog. At the same time, we have destroyed all possible clusters with *dependent* spatio-temporal structure. For example, when randomized seismic activity increases in 1992, it affects the entire region, and not only the vicinity of the Landers earthquake as in the original catalog (cf. Fig. 2). Figure B2a shows that this randomization suffices to destroy the bimodal structure of the joint distribution $(T,R)$: the randomized catalog is characterized by a unimodal distribution of $(T,R)$ located along a diagonal line.

The second randomized catalog (Figs. B1b and B2b) is obtained by reshuffling the events locations and using independent uniform random times within the duration of the original catalog. This catalog retains the marginal spatial distribution (and fault-related clustering) of events, while removing all the temporal inhomogeneities. The joint distribution $(T,R)$ is again unimodal; in addition it is more compact and is better separated from the origin, comparing to that of the randomized catalog from Fig. B2a. These differences are related to removing the temporal clustering of events.

The third randomized catalog (Figs. B1c and B2c) is obtained by retaining the original times of events and using random locations that are uniformly distributed between 30 – 37.5N and 113 – 122W. This catalog retains the temporal clustering of the original catalog while removing all the spatial inhomogeneities. The joint distribution of $(T,R)$ is bimodal in this case, with a weak second mode caused by the temporal clusters. The events that comprise this mode tend to happen close in time to their parents ($T \approx 10^{-6}$) and far away from the parents in space ($R \approx 10^{0.5}$). This spatial separation is two orders of magnitude higher than that observed in the original catalog (Fig. 5b). A noteworthy observation is that the time clustering of the observed seismicity is "stronger" than the spatial clustering, as illustrated by the comparison of the joint distributions $(T,R)$ in Figs. B2b and B2c.

## Section C. *Proof of the tree structure of the spanning earthquake network*

Recall that the NND $\eta$ is asymmetric: The parent $i$ of event $j$ must happen earlier: $t_i < t_j$. Hence, if we start at any earthquake $j$ in the catalog and repeatedly move from each event to its parent, we never can reach $j$ again. This implies that each possible nearest-neighbor cluster is a tree (a graph without cycles). Next, we show that we only have a single spanning tree. Each nearest-neighbor cluster (tree) must have a *root* – an earthquake without the parent. But we have only one such earthquake – the first event in the catalog; all other events have well-defined parents. This completes the proof.

## Section D. *Quality and stability of cluster identification in ETAS model*

### D.1 Model specification and parameters

The ETAS belongs to the class of Marked Point Processes (MPP). Traditionally, the main object of MPP analysis is the conditional intensity $m(t,\mathbf{f},m|H_t)$ of a process $Z_t=\{t_i,\mathbf{f}_i,m_i\}$ given its history $H_t = (\{t_i,\mathbf{f}_i,m_i\} : t_i < t)$ up to time $t$. Here $t_i$ represents earthquake occurrence times, $\mathbf{f}_i$ their coordinates (e.g., epicenter, hypocenter, or centroid) and $m_i$ the magnitudes. It can be shown [*Daley and Vere-Jones*, 2002] that conditional intensity completely specifies the process $Z_t$. The statistical analysis and inference for $Z_t$ are done using the conditional likelihood

$$\log L_t = \sum_{t_i<t} \log \mu\left(t_i, \mathbf{f}_i, m_i \mid H_t\right) - \int_0^t \int_M \int_F \mu\left(t, \mathbf{f}, m \mid H_t\right) dt\, dm\, d\mathbf{f}\,, \tag{D1}$$

where $M$ and $F$ denote the magnitude range and spatial domain of events, respectively. The ETAS assumes a particular self-exciting mechanism of earthquake generation. Namely, some background events (immigrants) occur according to a homogeneous stationary Poisson process. Each earthquake in a catalog generates offspring (first generation events), these offspring generate their own offspring (second generation events), and so on. The resulting seismic flow is a compound of immigrants and triggered events from all generations. The main body of the work on ETAS operates under the assumption that the magnitudes of events are independent and drawn from the Gutenberg-Richter (exponential) distribution with a constant $b$-value. This reduces conditional intensity to the following special form, which allows various particular parameterizations [*Ogata*, 1998, 1999]:

$$\mu\left(t, \mathbf{f} | \mathrm{H}_t\right) = \mu_0\left(t, \mathbf{f}\right) + \sum_{i: t_i<t} g\left(t - t_i, \mathbf{f} - \mathbf{f}_i, m_i\right).$$

We use in this study a homogeneous background intensity $\mu_0 = \mu$ and the following parameterization for the response function $g$ suggested by *Ogata* [1998, Eq. (2.3)]:

$$g\left(t, x, y, m\right) = \frac{K}{\left(t + c\right)^p} \frac{\exp\left(\alpha\left(m - m_0\right)\right)}{\left(x^2 + y^2 + d\right)^q}\,. \tag{D2}$$

Here $m_0$ is the lowest considered magnitude, and $(x,y)$ are Cartesian coordinates of the epicenters. The model is specified by 8 scalar parameters $\theta = \{\mu, b, K, c, p, \alpha, d, q\}$.

It has been shown [*Sornette and Werner*, 2005; *Veen and Schoenberg*, 2008; *Wang et al*., 2010] that estimation of the ETAS model is affected by the catalog's lowest magnitude cutoff, which may lead to a serious bias in the numerical values of the estimated parameters. It is also known that the ETAS parameters depend on the tectonic environment [*Chu et al*., 2011] and local physical properties of the lithosphere [*Enescu et al*., 2009]. These are some of the reasons why there are no commonly accepted "standard" values of the ETAS parameters for a given region. In this study, we generate synthetic ETAS catalogs using a range of parameters consistent with those reported in the literature [e.g., *Wang et al*., 2010; *Chu et al*., 2011; *Marzocchi and Zhuang*, 2011].

**D.2 Clustering in ETAS model**

An ETAS catalog can be naturally divided into individual clusters according to the model's explicit parent-offspring relationships. Namely, a cluster is defined as a group of events that have the common *ancestor* (grand-parent of arbitrary order), which itself is a background event (has no parent). This unique cluster's ancestor is also included in the cluster; by construction it is always the first event in a cluster. According to this definition, some clusters consist of a single background event, while the others include several generation of offspring. Within each cluster, we assign the following event types, same as in analysis of observed catalogs. *Mainshock* is the first largest event in a cluster, *foreshocks* are all events before the mainshock, and *aftershocks* are all events after the mainshock.

We next explore how the cluster technique of Sect. 3 can recover (i) the partition of an ETAS catalog into individual clusters, (ii) the event type (main/fore/aftershock) assignment and (iii) the parent-offspring assignment. The analysis is done using the observed catalog of events that reports only their occurrence time, magnitude and location. It should be noted that while we do study the parent-offspring assignment, it plays secondary role in the context of our study, comparing to the partition into individual clusters and event type. In the subsequent analysis, the event types, as well as parent and cluster assignments that correspond to the actual ETAS model structure will be called *true*; while those estimated using our cluster technique will be called *estimated*.

**D.3 Cluster identification: quality**

The analysis in this study was done using multiple ETAS catalogs with a range of realistic parameter values. We found that the results in different catalogs are qualitatively very similar to each other, with quantitative differences being directly related to the model parameters (e.g., different *b*-value, *p*-value, *etc.*) In this and the next section we illustrate the results using a particular ETAS catalog that corresponds to parameters $\mu = 0.003$ (km$^2$ year)$^{-1}$, $b = \alpha = 1$, $K = 0.007$ (km$^2$ year)$^{-1}$, $c = 0.00001$ year, $p = 1.1$, $q = 1.7$, $d = 30$ km$^2$; the simulations are done within a region of 500×500 km during 10 years. The synthetic catalog is illustrated in Figs. D1a, D2a that show, respectively, the magnitude and *X* coordinate of events as a function of time. The catalog consists of 29,761 events, of which 7,545 (25%) are background events. Figure D3 shows the joint 2-dimensional distribution of the temporal (*T*) and spatial (*R*) components of the nearest-neighbor distance $\eta$ (panel a) as well as the distribution of the scalar values of $\eta$ (panel b). The figure clearly demonstrates prominent bimodality of the nearest-neighbor distance, similar to the one reported for the observed seismicity (cf. Fig. 4). A bimodal distribution of the nearest-neighbor distance $\eta$ in ETAS model has been also reported by *Zaliapin et al.* (2008) and *Gu et al.* (2012).

The time-magnitude and time-coordinate sequence of mainshocks identified by the analyzed cluster technique are illustrated in Figs. D1b and D2b, respectively. Visually, our cluster procedure makes a decent job in identifying and removing the clusters from the original ETAS catalog. Tables D1, D2 and Fig. D4 assess the cluster detection in a quantitative way. Table D1 cross-classifies the events in the catalog according to their true vs. estimated type: 88% of events have been correctly classified into fore/main/aftershocks; the majority of the misclassified events (7%) are aftershocks recognized as mainshocks. The latter misclassification is due to the long-range triggering, when offspring occur at large time and/or distance from their parents. This long-range triggering is caused by the power-law tails of the temporal and spatial offspring kernels use in ETAS model. In the presence of a non-zero background the long-range offspring are mixed with the background events and cannot be correctly identified by a purely statistical procedure; the number of misclassifications increases with the background intensity. Table D2 illustrates similar cross-classification for 279 events with magnitude above 5. Clearly, the quality of detection increases with magnitude of analyzed events. Figure D4 shows the proportion of various misclassifications among events with magnitude above *m*: Black dots show proportion of events with misspecified parent, open circles – proportion of events assigned to a wrong cluster, squares – proportion of misclassified types (the same as Tables D1, D2), diamonds – proportion of misclassified mainshocks. Notably, the proportion of events with misspecified parents is about 40% for events of magnitude below 6, which is much higher than the proportion of other misclassification types. In particular, the cluster is correctly recognized for over 88% of events; the proportion of respective errors decreases to zero as magnitude *m* increases to 5.8. This shows that although it can be difficult to detect the true ETAS parents, one can still closely reconstruct the cluster structure of a catalog. This is an important observation, since the clusters present the primary object of the analysis in this study.

**D.4 Cluster identification: stability**

This section assesses and illustrates the stability of cluster identification with respect to the parameters of the algorithm, minimal reported magnitude, catalog incompleteness, and errors in event location.

First, we consider the three numerical parameters that are used in the cluster detection procedure: fractal dimension of epicenters $d_f$, *b*-value, and cluster detection threshold $\eta_0$. The value of the threshold $\eta_0$ is estimated in each experiment from the Gaussian mixture model [*Hicks*, 2011], except the experiments when we explicitly vary $\eta_0$. We intentionally choose wide ranges for the parameter values:

$$1 \le d_f \le 3,\ 0 \le b \le 2, \text{ and } -6 \le \eta_0 \le -2.$$

The chosen ranges are much wider than the respective statistical margins of error that correspond to estimating these parameters in ETAS model or in observations. This is done in order to test the general limits of applicability of the proposed cluster technique. Recall that the main version of the analysis uses

the true ETAS values $d_f = 2$ and $b = 1$ and the corresponding threshold $\eta_0 = -4.476$ from the Gaussian mixture model; we refer to these parameters as *standard*.

Figure D5 summarizes the results of 1D stability analysis where we vary a single parameter and keep the rest at their standard values. A rather surprising observation is that the total proportion of misspecified event types, shown in panels (a-c), *never* exceeds 33%, even for obviously outrageous parameter values. For the parameters close to their standard values (shown by stars), the proportion of misspecified events is within 10% – 15%, which is very close to the error of 12% observed in the main version of the analysis. Panel (d) shows individually the proportion of misspecified mainshocks (squares) and aftershocks (triangles) as a function of the threshold $\eta_0$. This panel emphasizes the broadness of the parameter range considered – the proportion of misspecified mainshocks changes from 0 to 100% within the considered range. The panel also illustrates that most of the aftershocks are very well separated from the mainshocks: even when the threshold is so low that *all* mainshocks are properly specified, the proportion of misspecified aftershocks is only 40%. The same conclusion can be derived, of course, from visual analysis of the bimodal distribution in Fig. D3.

Figure D6 illustrates a 2D stability analysis; it shows the proportions of misspecified mainshocks (panel a) and aftershocks (panel b) as a function of the pair $(b,d_f)$ on a 20x20 grid; the threshold $\eta_0$ is estimated in each experiment from a Gaussian mixture model. Similar to the 1D stability experiments, the proportion of errors is a smooth function of the algorithm parameters, so that the error remains close to the one observed for the main version of algorithm. The proportion of misspecified mainshocks in *all* experiments is within 5%-10%. A significant increase of misspecified aftershocks, to 30%, is only observed for clearly "wrong" values of parameters, e.g. $b \approx 0$, $d_f \approx 1$.

We now analyze stability of cluster detection with respect to the minimal reported magnitude. Specifically, we perform the cluster analysis for a truncated catalog, only using magnitudes $m \geq m_0$ (starting with computing nearest-neighbor distances, etc.), and then compare the event types estimated in the truncated catalog with the true event types. The results are shown in Fig. D7. The proportion of misspecified events decreases with completeness magnitude $m_0$ from the original 11.57% to 0 at $m_0 = 5.7$; in other words, the cluster detection quality increases with magnitude of event. The same conclusion can be drawn from the analysis of Fig. D4 above. We notice that the analysis of Fig. D4 differs from the one performed here in that in Fig. D4 we always use the event types estimated in a complete catalog, and only report proportions of errors for different magnitude thresholds. Here, in contrast, we perform the complete cluster and event type estimation in each truncated catalog.

Next, we analyze stability of cluster detection with respect to the catalog incompleteness. For that, we perform thinning of the original ETAS catalog so that each event with magnitude $3 \leq m \leq 5$ has probability $P(m) = (5-m)/2$ to be removed. More specifically, all events with magnitude $m \leq 3$ are definitely removed; all events with magnitude $m \geq 5$ are definitely retained; all other events has removal probability $P(m)$ that decreases linearly with magnitude. Figure D8a compares the magnitude distribution in the original and a thinned catalog. The thinning in this experiment is quite severe: it retains only about 20% of events in the catalog. We generate 100 thinned catalogs according to this procedure and compute the proportion of misspecified events in each of them. An event is called misspecified if (i) it has been retained in the catalog after thinning, and (ii) its type in the analysis of the thinned catalog is different from the type of this event in the analysis of the actual catalog. The proportion of misspecified events is $0.1249 \pm 0.009$ (95%CI); its distribution is shown in Fig. D8b. Comparing this with the original misspecification proportion of 0.1157 (see Sect. D3, Table D1), we conclude that the catalog incompleteness has a very weak effect on the cluster detection quality.

Finally, we analyze the effects of location errors. For that, we randomly shift the epicenters of events in the ETAS catalog by adding independent 2D Gaussian errors with independent components of zero mean and standard deviation $\sigma$. We then perform cluster analysis on a randomized catalog and compare the estimated results with the true ones, focusing on the proportion of the events with misclassified types. We considered 100 randomized catalogs for each value of $\sigma$. Recall that the cluster identification in the true catalog corresponds to the proportion 0.1157 of misclassified events (see Sect.

D3, Table D1). The proportion of misclassified events in randomized catalogs for $\sigma$ = 0.1km, 0.3km, and 1.0km is, respectively, 0.1167±0.001, 0.1170±0.002, and 0.1187±0.002 (95%CI). This shows that random location errors produce practically negligible effect on cluster detection and event classification.

**D.5 Basic cluster statistics**

This section focuses on basic statistics of the detected clusters. The ETAS catalog we use here is longer than the one in the previous sections, to be a better match to the observed catalog in southern California. Specifically, we use an ETAS model with the same parameters as above: $\mu = 0.003$ (km$^2$ year)$^{-1}$, $b = \alpha = 1$, $K = 0.007$ (km$^2$ year)$^{-1}$, $c = 0.00001$ year, $p = 1.17$, $q = 1.7$, $d = 30$ km$^2$; the simulations are done within a region of 500×500 km during 15 years. The catalog consists of 146,432 earthquakes. The bimodal distribution of the nearest-neighbor distance and cluster identification quality (not shown) are similar to those reported in the previous sections for a shorter ETAS catalog.

Figure D9 illustrates the frequency-magnitude distribution for mainshocks/singles and aftershocks (true and estimated). The true mainshock and aftershock distributions are distinctly different, each being closely approximated by an exponential (GR) law with different *b*-values. We also observe upward (downward) deviations from the exponential laws at largest magnitudes. The estimated distributions are very close to the true ones (see legend). Panel (a) shows the cumulative distribution function (cdf), panel (b) shows the normalized cdf in order to emphasize the deviations from a pure exponential law. Table D3 reports the maximum likelihood estimations of the *b*-values for different event types together with the respective uncertainties. A noteworthy observation is that the estimated *b*-value for aftershocks is larger than that for mainshocks and foreshocks; the same difference is seen in other ETAS catalogs as well (not shown). This difference is due to the conditional assignment of event types, which deflates the *b*-value for mainshocks (largest events in respective clusters), and, accordingly, inflates it for aftershocks. The *b*-value for foreshocks is smaller than that for aftershocks since larger events have higher chance to become parents for mainshocks, according to the employed earthquake distance of Eq. (1).

Figure D10 illustrates cluster productivity: the number of foreshocks and aftershocks per mainshock. Panel (a) shows the cluster size $N$ as a function of cluster mainshock magnitude $m$; the data is closely approximated by the exponential line $N \propto 10^{\beta m}$. The exponent index $\beta$ estimated within the intermediate magnitude ranges $3.0 \le m \le 6.0$ is 1.09 ± 0.02, where the error margins correspond to a 95% confidence interval (95% CI). We also show for comparison the number of first-generation offspring per parent (squares), which by ETAS construction has exponent index 1. Panel (b) shows the cumulative distribution of the cluster size $N$ (circles) and the number of first-generation offspring (squares). Both distributions have a power-law tail. The distribution of the offspring is closely approximated by a Pareto law $F(x) = cx^{-a}$, $c>0$, $a \approx 1$. The cluster size distribution deviate from this scaling due to finite size effects: The largest events in the catalog tend to attract a larger number of offspring, while the smallest events cannot attract enough offspring because of the catalog's magnitude cutoff. The value of the scaling exponent $a \approx 1$ is related to the chosen values of the ETAS parameters $b = \alpha = 1$. It is readily seen (e.g., *Saichev et al.*, 2005) that the combination of exponential frequency-magnitude relationship with $b = 1$ and exponential offspring productivity with $\alpha = 1$ leads to the power law cluster size distribution with index $a = b/\alpha = 1$. It must be noted though that this argument concerns only the first-generation offspring, while we work with offspring of all generations. We notice, however, that in the examined catalog clusters with only first generation offspring comprise 77% of all non-single clusters, and clusters with the average leaf depth smaller than 2 (hence, with a significant fraction of the first generation offspring) comprise 86% of all non-single clusters. Similar proportions hold for the other examined ETAS catalogs. Hence, the first order

approximation to the cluster size distribution can be done under the assumption of single generation offspring.

The intensity of foreshocks and aftershocks within 50 days of the mainshock is shown in Fig. D11. Black dots refer to aftershocks (panel a) and foreshocks (panel b) of mainshocks with magnitude $m \geq 4$. The slope of aftershock decay estimated for $t \geq 0.5$ day, is -0.93 ± 0.09 (95%CI); the slope of foreshock decay is harder to estimate due to large fluctuations of the respective intensities. The deviation of the aftershock slope from $p = 1.1$ used in ETAS simulations is explained by existence of secondary, ternary, *etc.* aftershocks. Panel (a) shows for comparison (light squares) the intensity of the first-order offspring in ETAS model. The slope estimated within $t \geq 0.5$ day is -1.1 ± 0.01 (95%CI).

Figure D12 shows the distribution of magnitude differences between mainshock and aftershock/foreshocks in families with mainshock magnitude $m \geq 4$: panel (a) refers to all aftershocks and foreshocks; panel (b) refers to the largest aftershock/foreshock in a family. The first observation (panel a) is that the majority of aftershocks and foreshocks have rather large magnitude difference from the mainshock: $d_m \geq 4$ for 80% of aftershocks and $d_m \geq 3$ for 80% of foreshocks. It is also noteworthy that the difference $\Delta_m$ between the mainshock and the largest aftershock (panel b) is almost uniform within the range $0 \leq \Delta_m \leq 2$, while the foreshock difference shows larger fluctuations.

Finally, we analyze the distribution of the number $N_{\text{off}}$ of direct offspring. According to the ETAS definition, the actual number $N_{\text{off}}$ of offspring of an event of magnitude $m$ has Poisson distribution with intensity $\lambda \propto 10^m$. The coefficient of proportionality is determined by the space-time kernel of Eq. (D2). The distribution of the estimated number of offspring though significantly deviates from a pure Poisson. This is explained by the existence of the actual offspring of event $i$ that were attached to other events during the estimation, as well as the offspring of other events that were attached to $i$. These effects create additional variability in the estimated number $N_{\text{off}}$, which can be closely approximated by a *negative binomial* distribution, as illustrated in Fig. 12b of the main text.

## Section E. *Stability of cluster identification in southern California*

This section assesses the stability of cluster identification in the observed catalog. Here, unlike the analysis of ETAS model, we do not know the "true" cluster structure, so the *quality* of cluster identification cannot be directly assessed. At the same time, we can assess its *stability*. For that, we vary parameters of the algorithm and compare results with the ones obtained in the main version of the analysis, which is done here with $d_f = 1.6$, $b = 1$, minimal magnitude $m_0 = 3$, and threshold $\eta_0$ estimated from the Gaussian mixture model. The use of adaptive estimation of the threshold is important in these experiments, since its values depend (although weakly) on the other three parameters of the algorithm. Figure E1 shows the proportion of events with estimated type different from that obtained in the main version of analysis, as a function of each of the parameters. Similarly to the ETAS stability analysis, we intentionally use very wide ranges for parameter variation, in order to explore the general limits of algorithm stability:

$$1 \leq d_f \leq 2,\ 0 \leq b \leq 2,\ 3 \leq m_0 \leq 6, \text{ and } -6 \leq \eta_0 \leq -4.$$

The proportion of misspecified types is below 7% for *all* experiments within the following parameter ranges:

$$1.1 \leq d_f \leq 2,\ 0.5 \leq b \leq 1.3,\ 3 \leq m_0 \leq 6 \text{ and } -5.5 \leq \eta_0 \leq -4.55.$$

The errors larger than 7% are only observed for the parameter values that are clearly inconsistent with the available observations, like $b > 1.5$. Notably, the proportion of errors *never* exceeds 18% in our experiments.

Next, we analyze the stability of cluster detection with respect to the event location error. Specifically, we generate 100 catalogs by randomly altering the locations of events. The location error is modeled by a 2D Normal random variable with zero mean, independent components, and standard deviation for both component given by the standard error of event location reported by *Hauksson et al.* (2012). The proportion of misspecified event types (compared to the analysis of true event locations) is 0.044±0.005 (95% CI); the maximal observed proportion is 0.051. This shows that the proposed algorithm is stable with respect to the location uncertainties.

The *stability* results of this section are consistent with that obtained above in ETAS model. This supports a conjecture that the *quality* of cluster detection, if one assumes that there exists a *true* cluster structure in observed catalogs, is also good, similar to that in ETAS analysis.

Table D1: Cross-classification of event types (true vs. estimated) in ETAS catalog: All 29,671 events are considered

| | | True | | |
|---|---|---|---|---|
| | | Foreshock | Mainshock | Aftershock |
| Estimated | Foreshock | 2760 (9%) | 77 (0.2%) | 157 (0.5%) |
| | Mainshock | 331 (1%) | 7007 (24%) | 2198 (7%) |
| | Aftershock | 242 (0.8%) | 461 (2%) | 16438 (55%) |

Table D2: Cross-classification of event types (true vs. estimated) in ETAS catalog: 279 events with magnitude $m \geq 5$ are considered

| | | True | | |
|---|---|---|---|---|
| | | Foreshock | Mainshock | Aftershock |
| Estimated | Foreshock | 31 (11%) | 1 (0.4%) | 1 (0.4%) |
| | Mainshock | 6 (2%) | 90 (32%) | 11 (4%) |
| | Aftershock | - | 4 (1%) | 135 (48%) |

Table D3: Estimated *b*-values for different event types in ETAS catalog (maximum likelihood estimation and confidence interval)

| | True | | Estimated | |
|---|---|---|---|---|
| | *b*-value | 95% CI | *b*-value | 95% CI |
| Mainshocks | 0.932 | 0.91 – 0.95 | 0.957 | 0.94 – 0.97 |
| Aftershocks | 1.006 | 1.00 – 1.01 | 1.006 | 1.00 –1.01 |
| Foreshocks | 0.960 | 0.92– 1.00 | 0.935 | 0.89 – 0.98 |

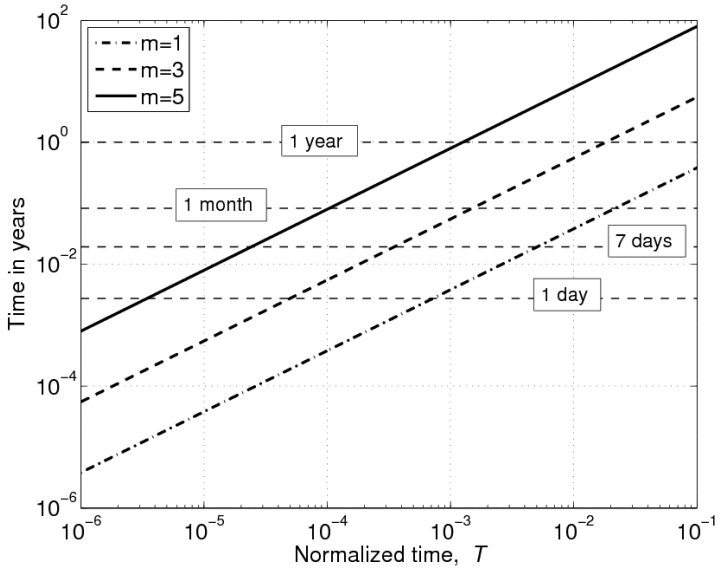


Figure A1: Correspondence between the normalized time $T$ of Eq. (2) ($x$-axis) used in the 2-D cluster analysis and time in years ($y$-axis) for earthquakes of different parent magnitudes, $m$ = 1, 3, and 5. Horizontal lines indicate times of 1 day, 7 days, 1 month, and 1 year.

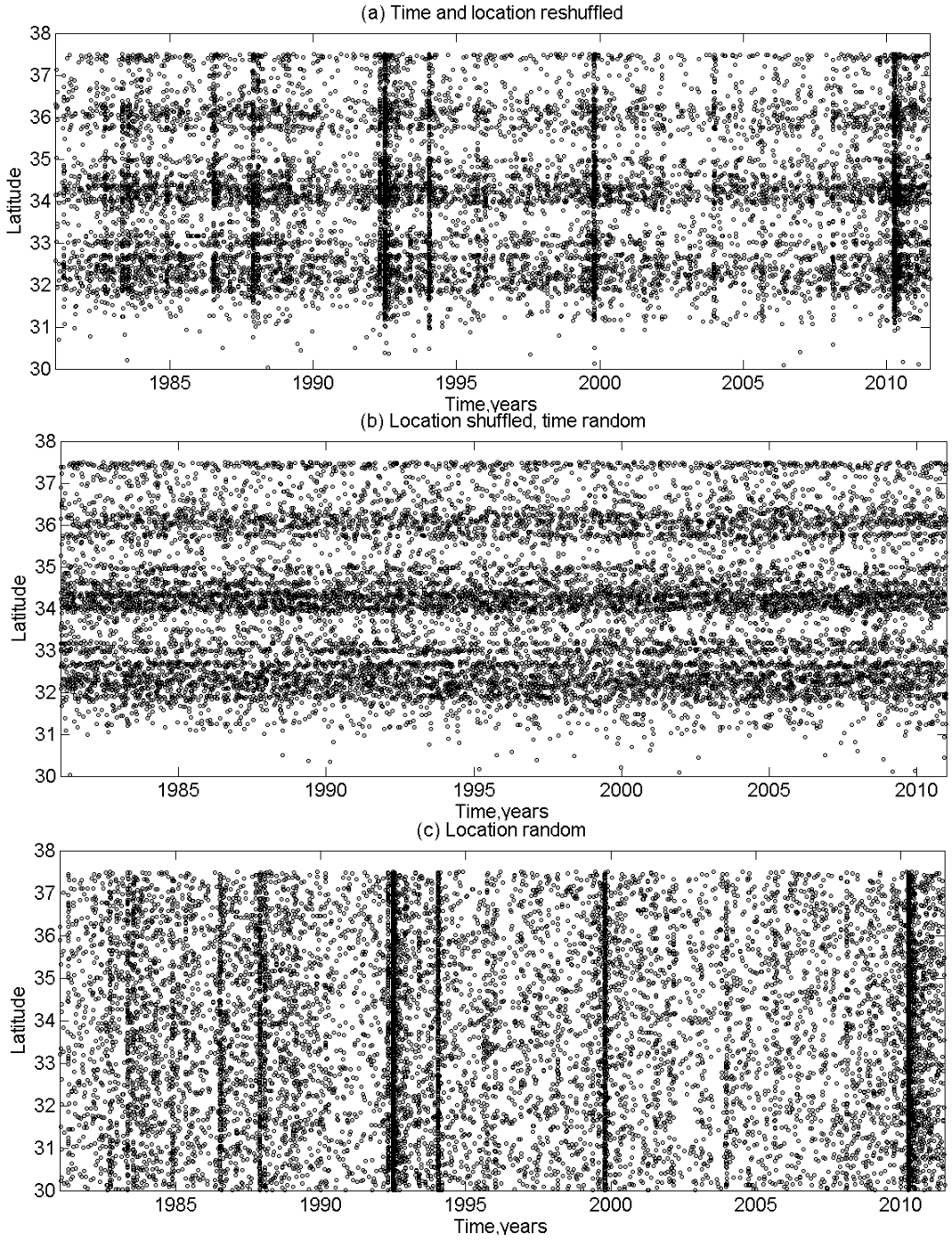


Figure B1: Time-latitude map of earthquakes from randomized catalogs. (a) Times and locations of the observed events are randomly reshuffled. (b) Locations are randomly reshuffled; times are uniform random variables. (c) Locations are uniform random variables, original times.

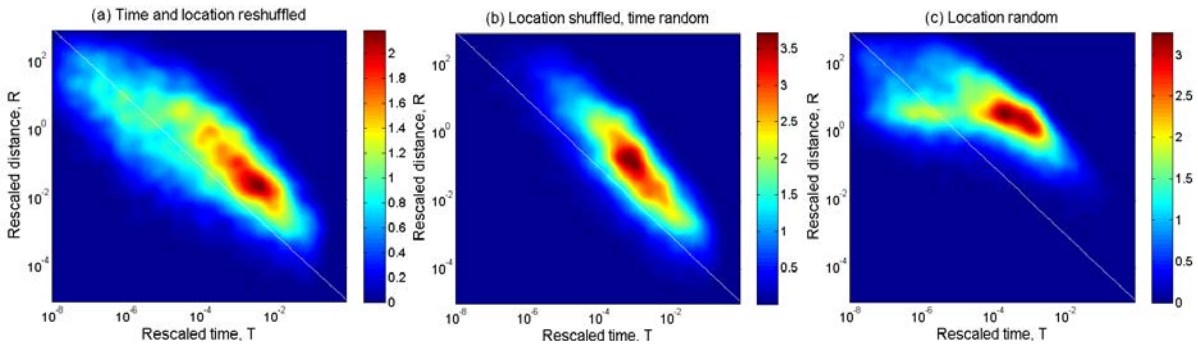


Figure B2: The joint distribution of rescaled time and space components ($T$,$R$) of the nearest-neighbor distance $\eta$ in randomized catalogs. (a) Times and locations are randomly reshuffled. This catalog retains the marginal spatial and temporal distributions of the observed seismicity, while removing their local interactions. (b) Locations are randomly reshuffled; times are uniform random variables. This catalog retains the spatial clustering, while removing all the time inhomogeneities. (c) Locations are uniform random variables, original times. This catalog retains the temporal clustering, while removing all the space inhomogeneities.

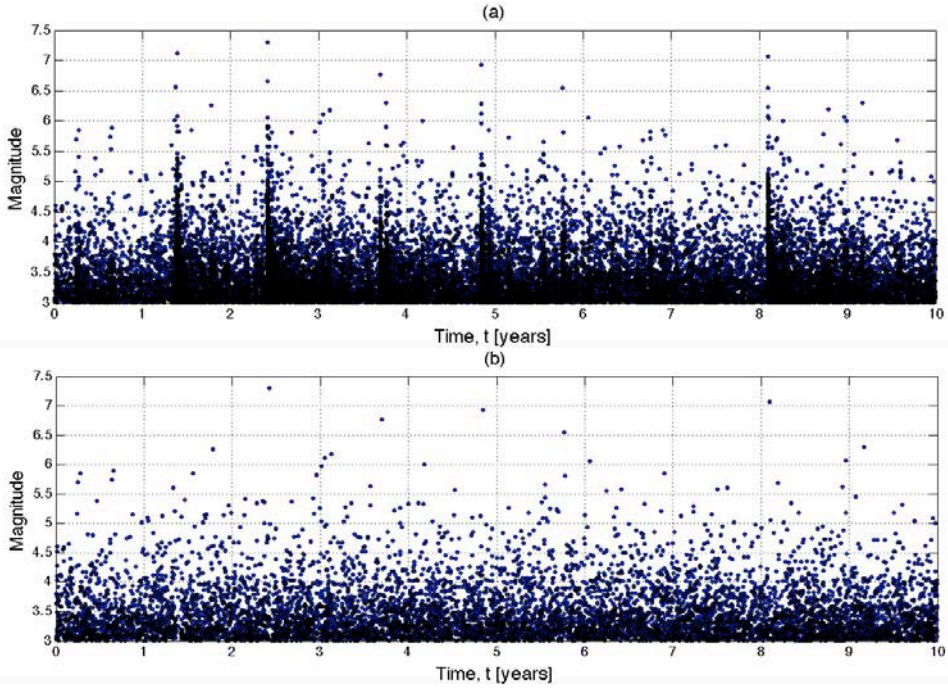


Figure D1: ETAS model – an example of declustering. Figure shows the time-magnitude sequence for events with $m \geq 3$. (a) All events, $n$ = 29,671; (b) Mainshocks, $n$ = 9,536.

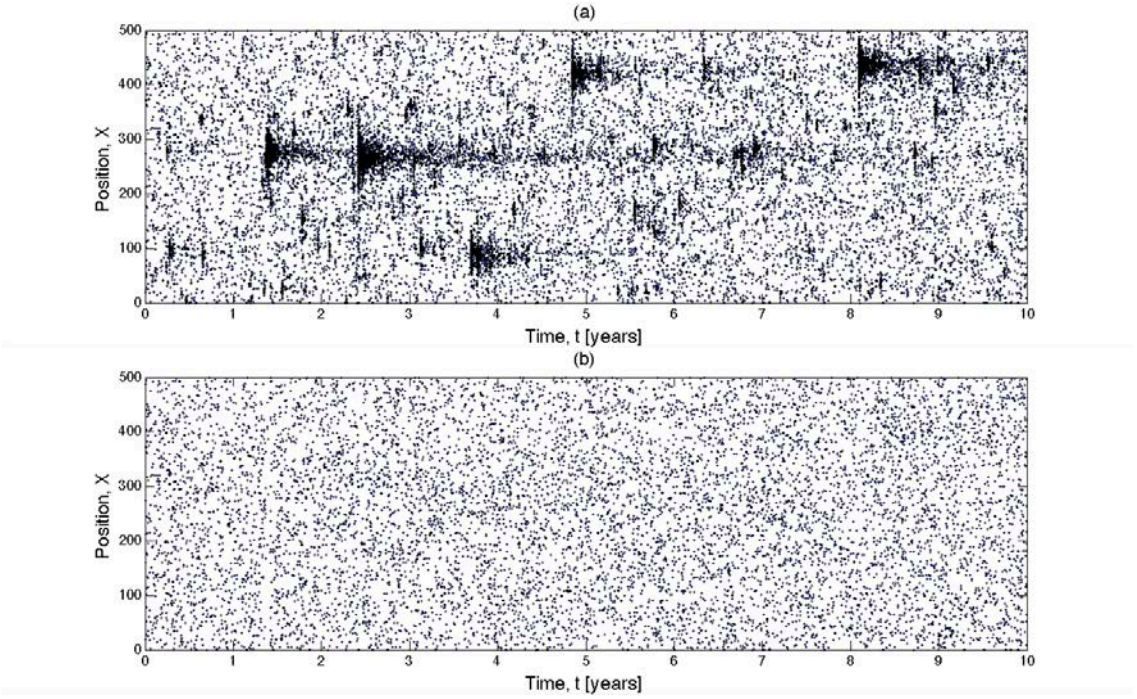


Figure D2: ETAS model – an example of declustering. Figure shows the $X$ coordinate of epicenters vs. time for all events in the catalog. (a) All events, $n$ = 29,671; (b) Mainshocks, $n$ = 9,536.

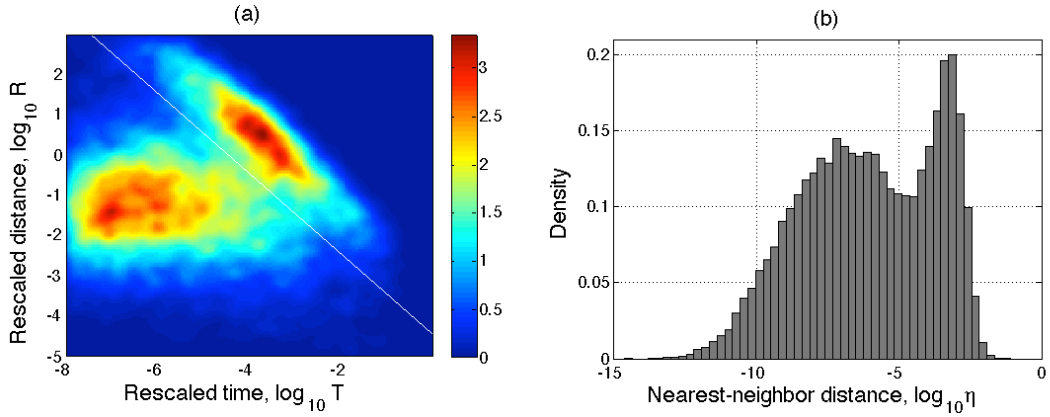


Figure D3: ETAS model – nearest-neighbor distance. (a) Joint distribution of the time and space components ($T$,$R$) of the nearest-neighbor distance $\eta$. (b) Histogram of the log-values of the nearest-neighbor distance $\eta$. Bimodal distribution is clearly seen: the background part is located above the white line in panel (a), and corresponds right mode in panel (b); clustered part is located below the white line in panel (b), and corresponds to left mode in panel (b). The white line in panel (a) corresponds to $\eta = -4.47$.

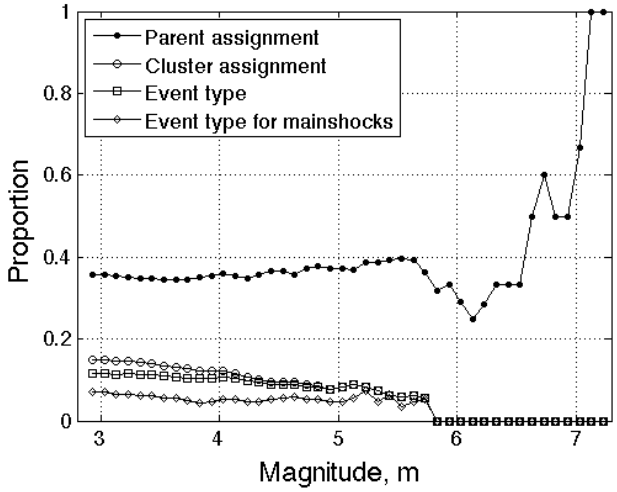


Figure D4: ETAS model – cluster identification errors. The figure shows the proportion of various erroneous identifications for events with magnitude above $m$. Dots – wrong parent assignment; circles – wrong cluster assignment; squares – wrong event type (fore/after/mainshock) assignment, stars – wrong event type assignment for mainshocks only.

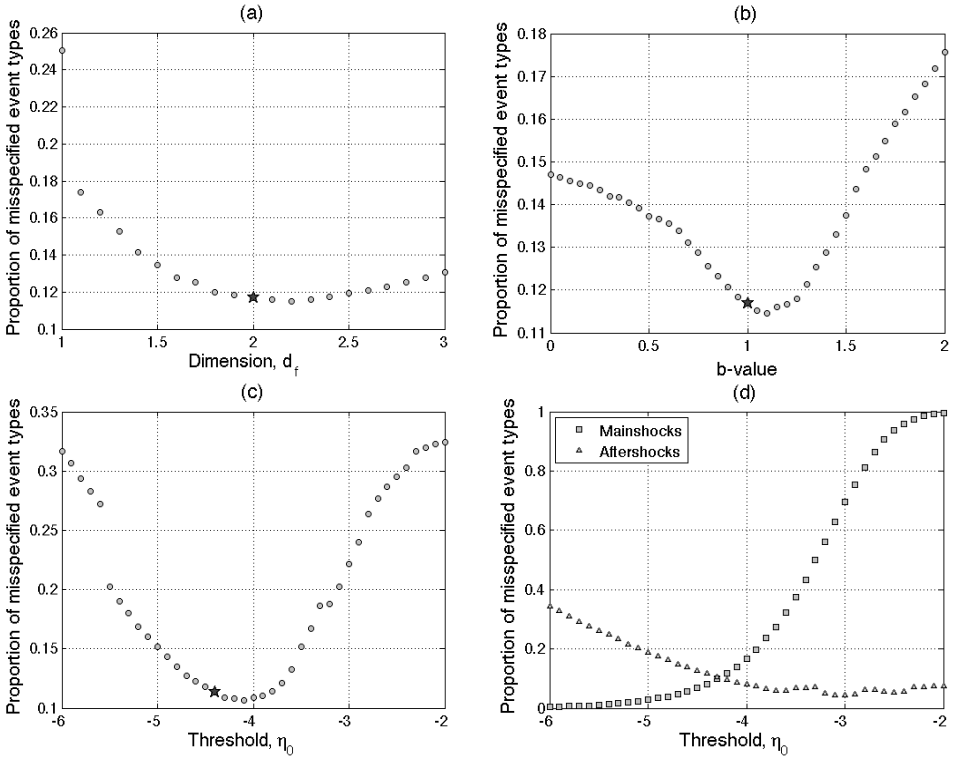


Figure D5: ETAS model – stability of cluster identification. Proportion of events with misspecified event type vs. model numerical parameters. Each panel refers to variation of a single parameter with the other parameters fixed. Stars in panels (a)-(c) refer to the values that correspond to the main version of the analysis, with true values of $d_f = 2$, and $b = 1$, and $\eta_0$ estimated according to the Gaussian mixture model. See text for details. Specifically, we vary (a) the fractal dimension $d_f$ of epicenters, (b) $b$-value, and (c-d) the threshold $\eta_0$. Panels (a-c) show the proportion of all events with misspecified type, panel (d) shows separately the proportion of misspecified mainshocks (squares) and aftershocks (triangles).

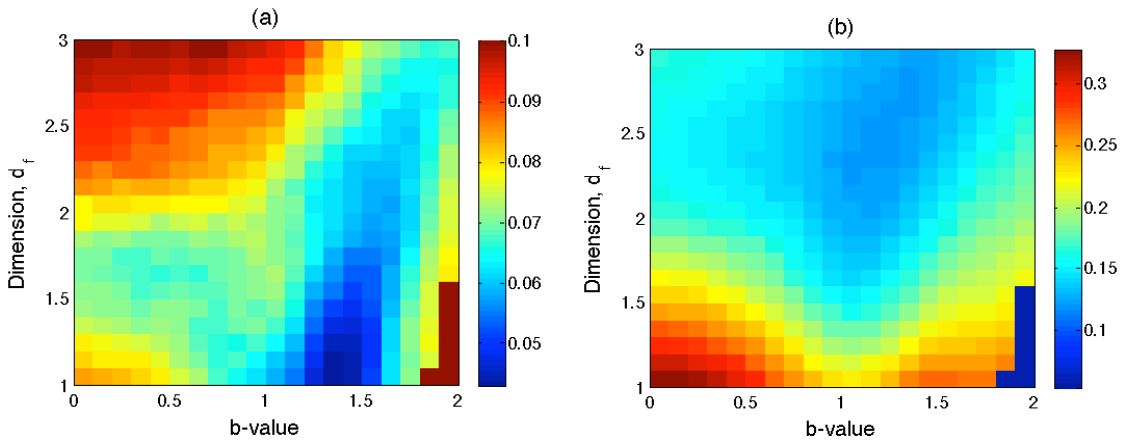


Figure D6: ETAS model – stability of cluster identification. Proportion of misspecified mainshocks (panel a) and aftershocks (panel b) as a function of the pair $(b,d_f)$.

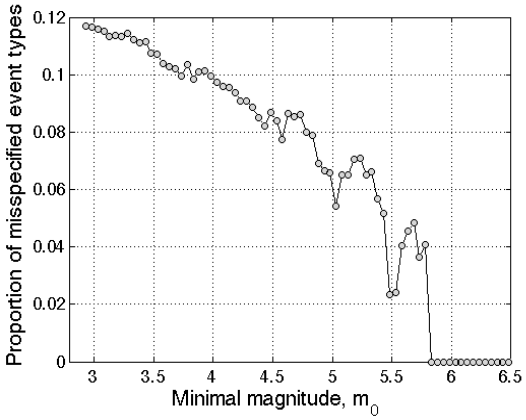


Figure D7: ETAS model – stability of cluster identification. Proportion of events with misspecified types, as a function of minimal magnitude of analysis.

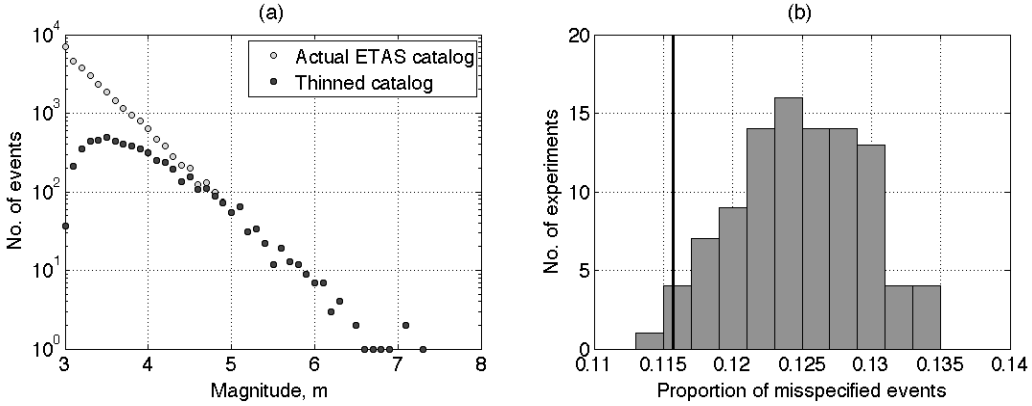


Figure D8: ETAS model – stability of cluster identification in thinning experiment. A thinned catalog is obtained from the actual catalog by removing each event with probability $P(m)$ that decrease linearly from 1 to 0 on the interval $3 \le m \le 5$. (a) Magnitude distribution in the actual (black circles) and a thinned (light circles) catalog. (b) Distribution of the proportion of misspecified events for 100 thinned catalogs. Black vertical line refers to the proportion of misspecified events in the true, complete catalog.

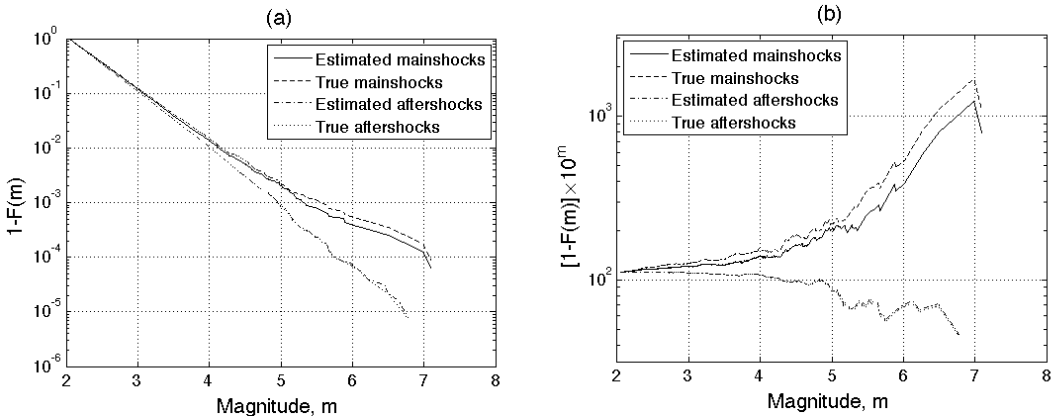


Figure D9: ETAS model – magnitude-frequency distribution. Figure refers to different event types as described in the legend. (a) Proportion 1-$F(m)$ of events with magnitude above $m$, where $F(m)$ is the empirical cumulative distribution function. (b) Weighted proportion of events with magnitude above $m$, $[1\text{-}F(m)]\times 10^m$. Panel (b) emphasizes deviations from an exponential distribution $E(m) = 1\text{-}10^{-m}$ with $b$-value 1, which corresponds to a horizontal line.

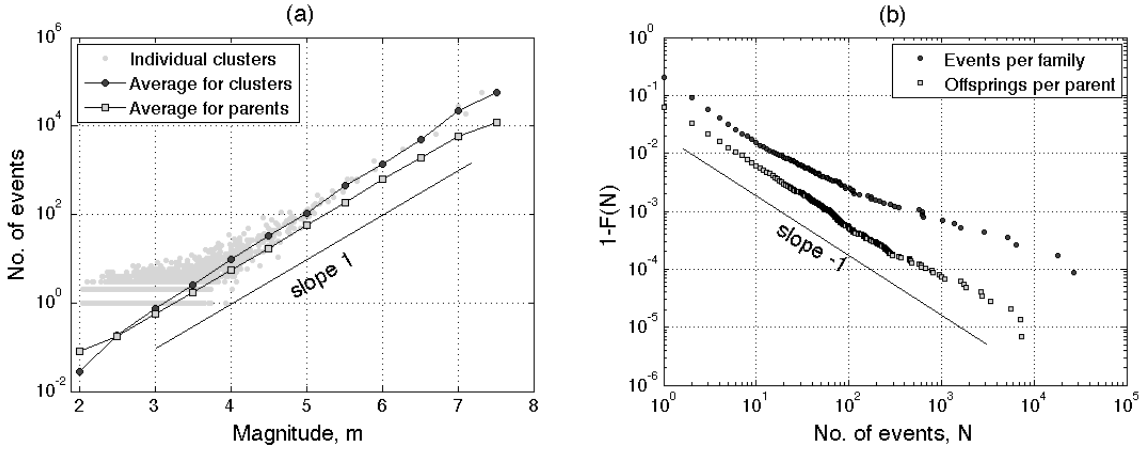


Figure D10: ETAS model – cluster productivity. (a) Number of aftershocks and foreshocks, $N$-1, in a cluster vs. cluster magnitude $m$. Black circles – average number of events in a cluster within magnitude window of length 0.5. Grey dots – individual clusters. Squares – average number of offspring per parent. (b) Distribution of cluster size $N$ (black circles) and the number of offspring per parent (squares).

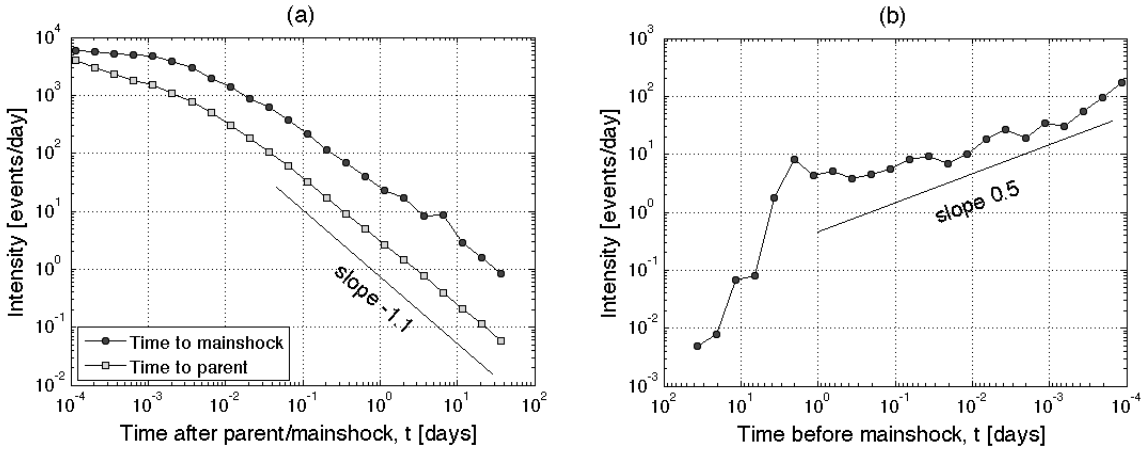


Figure D11: ETAS model – Aftershock and foreshock intensity. (a) Black dots – aftershocks within 50 days of mainshocks with magnitude $m \geq 4$. Squares – first generation offspring. (b) Foreshocks within 50 days of mainshocks with magnitude $m \geq 4$.

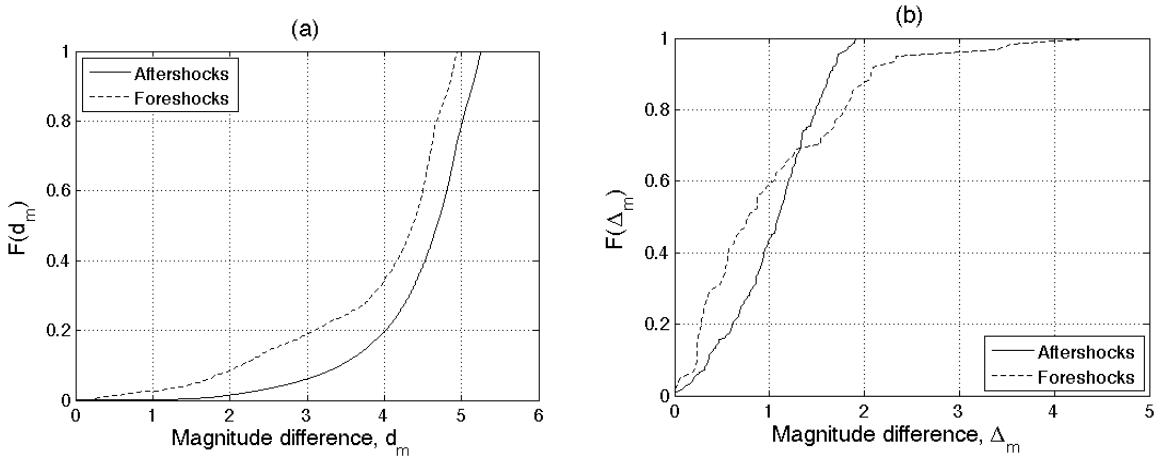


Figure D12: ETAS model – magnitude difference analysis. (a) Magnitude difference $d_m$ between mainshock and each aftershock (solid line) and foreshock (dashed line). (b) Magnitude difference $\Delta_m$ between mainshock and the largest aftershock (solid line) and largest foreshock (dashed line). Families with mainshock magnitude $m \geq 4$ are considered in both panels.

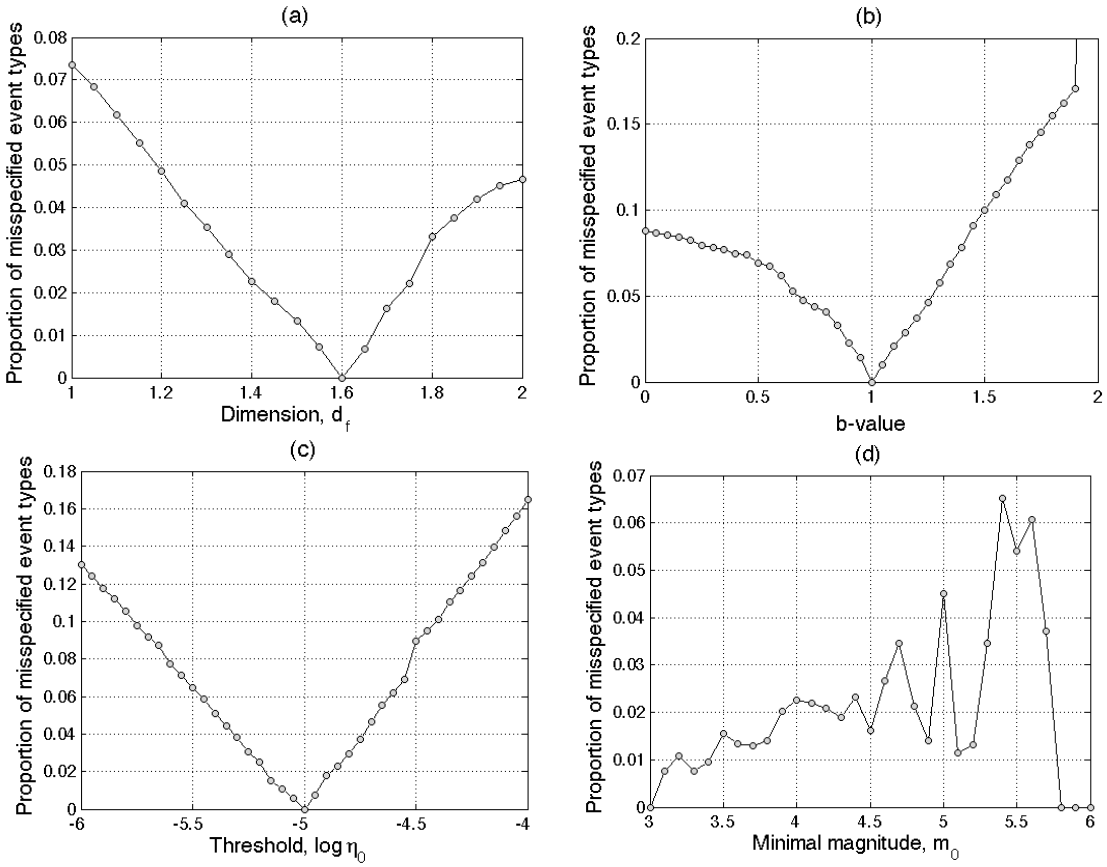


Figure E1: Stability of cluster identification in southern California. Proportion of events with event type different from that obtained in the main version of analysis as a function of algorithm parameter: (a) Fractal dimension of epicenters $d_f$, (b) $b$-value, (c) cluster threshold $\eta_0$, and (d) minimal magnitude of analysis. The main version of analysis uses $d_f = 1.6$, $b = 1$, $m_0 = 3$, and threshold $\eta_0$ estimated from the Gaussian mixture model.