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 η_{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 η_{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 (ϕ_i, λ_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 η 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)]:

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

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 η (panel a) as well as the distribution of the scalar values of η (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 η 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 η_0. The value of the threshold η_0 is estimated in each experiment from the Gaussian mixture model [*Hicks*, 2011], except the experiments when we explicitly vary η_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 η_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 η_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 η_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 $P(m) = (5\text{-}m)/2$ 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 ± 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 σ. 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 σ. 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 β 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 Δ_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_{off} of direct offspring. According to the ETAS definition, the actual number N_{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_{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 η_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:

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

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

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

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

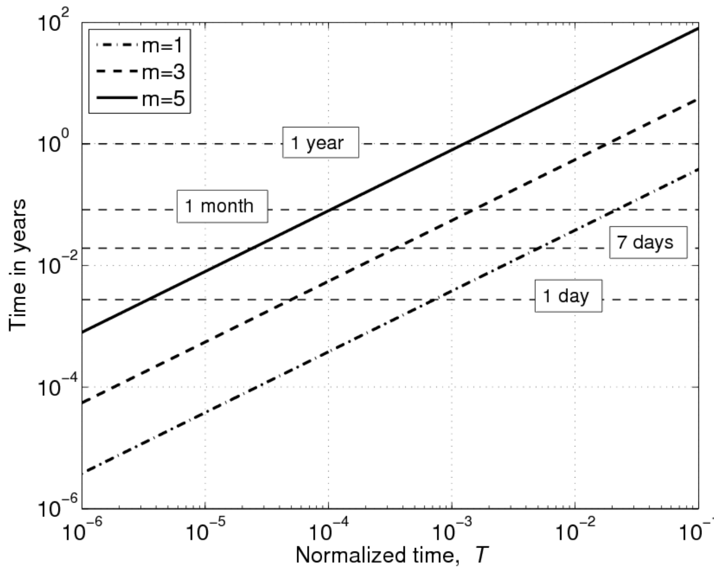

Figure A1: Correspondence between the normalized time T of Eq. (2) (x-axis) used in the 2-D cluster analysis and time in years (y-axis) for earthquakes of different parent magnitudes, m = 1, 3, and 5. Horizontal lines indicate times of 1 day, 7 days, 1 month, and 1 year.

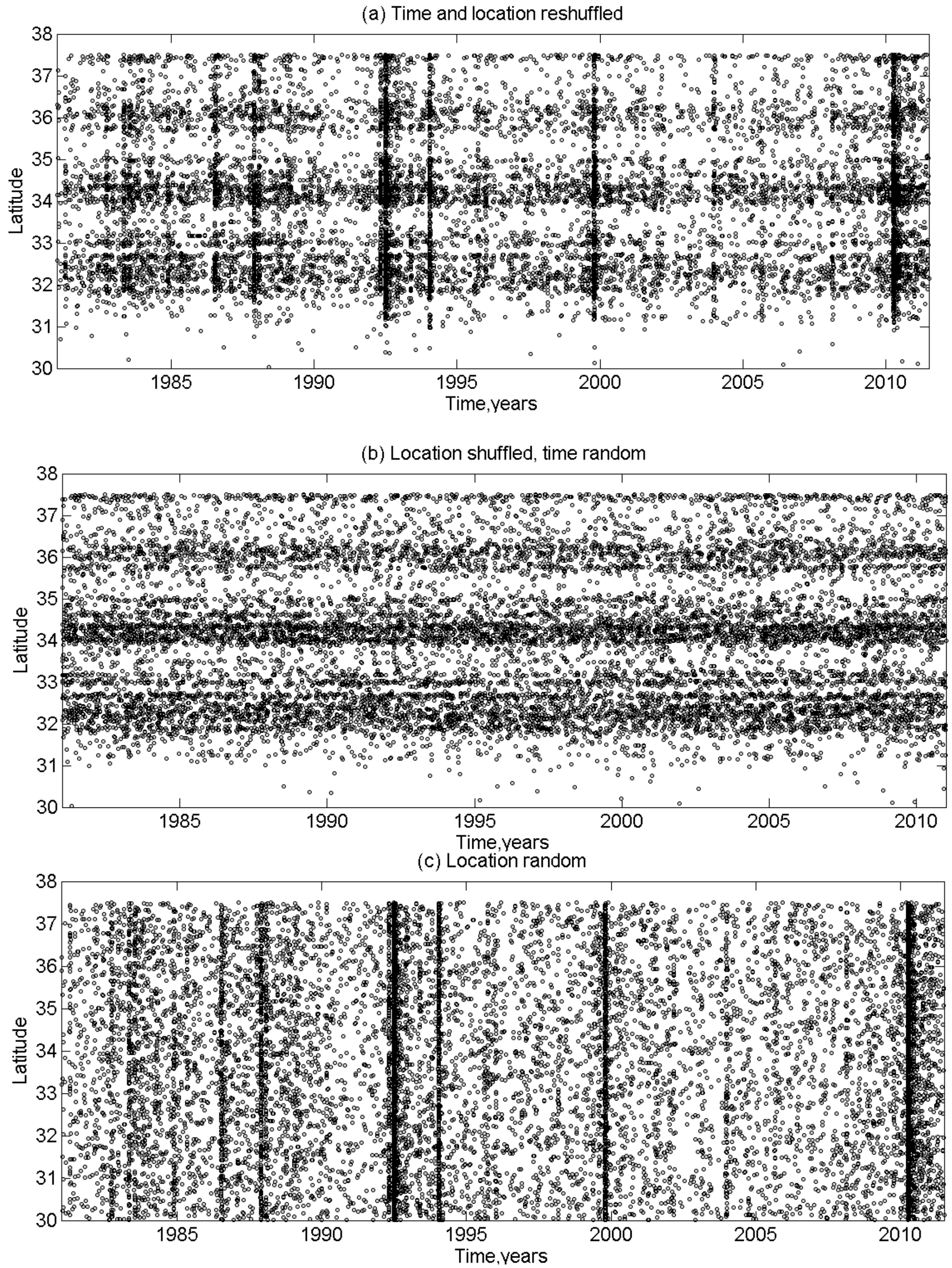

Figure B1: Time-latitude map of earthquakes from randomized catalogs. (a) Times and locations of the observed events are randomly reshuffled. (b) Locations are randomly reshuffled; times are uniform random variables. (c) Locations are uniform random variables, original times.

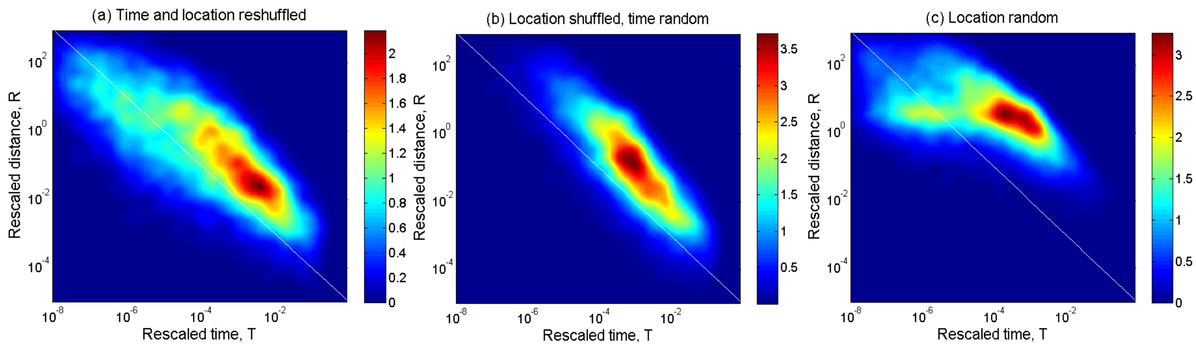

Figure B2: The joint distribution of rescaled time and space components (T,R) of the nearest-neighbor distance η in randomized catalogs. (a) Times and locations are randomly reshuffled. This catalog retains the marginal spatial and temporal distributions of the observed seismicity, while removing their 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.

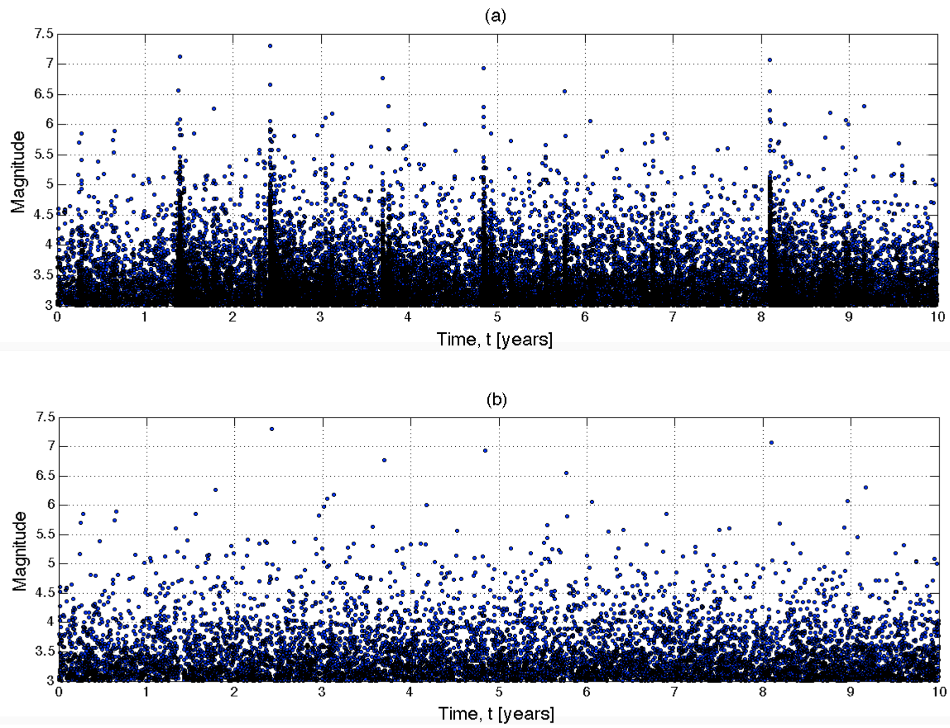

Figure D1: ETAS model – an example of declustering. Figure shows the time-magnitude sequence for events with $m \geq 3$. (a) All events, n = 29,671; (b) Mainshocks, n = 9,536.

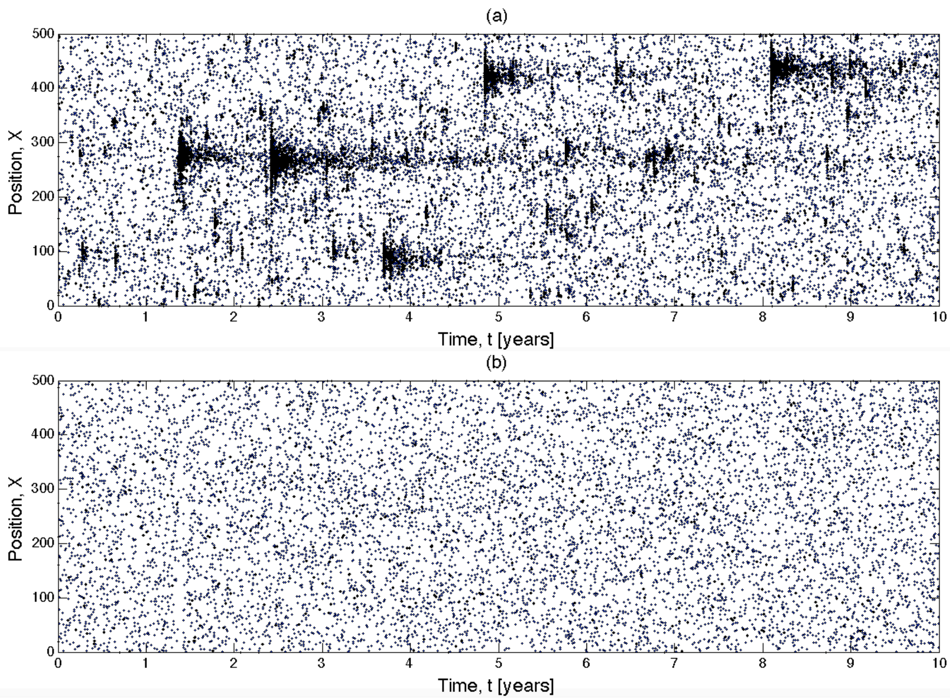

Figure D2: ETAS model – an example of declustering. Figure shows the X coordinate of epicenters vs. time for all events in the catalog. (a) All events, n = 29,671; (b) Mainshocks, n = 9,536.

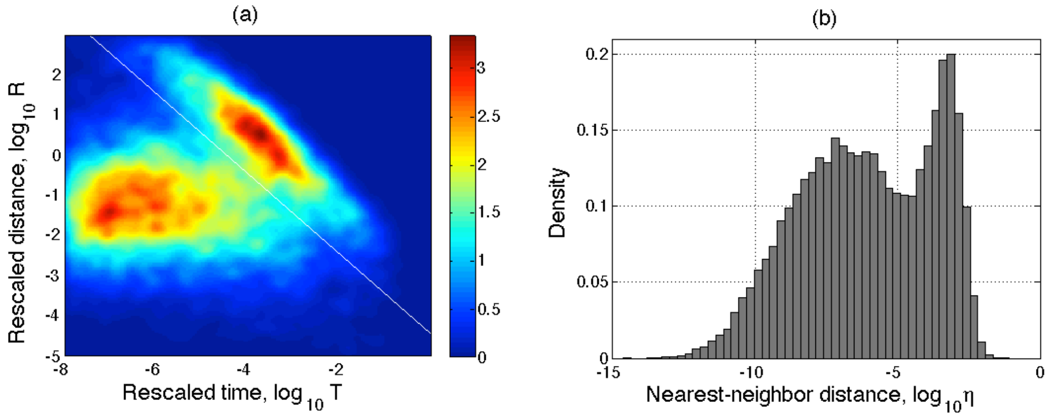

Figure D3: ETAS model – nearest-neighbor distance. (a) Joint distribution of the time and space components (T,R) of the nearest-neighbor distance η. (b) Histogram of the log-values of the nearest-neighbor distance η. Bimodal distribution is clearly seen: the background part is located above the white 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 $\eta = -4.47$.

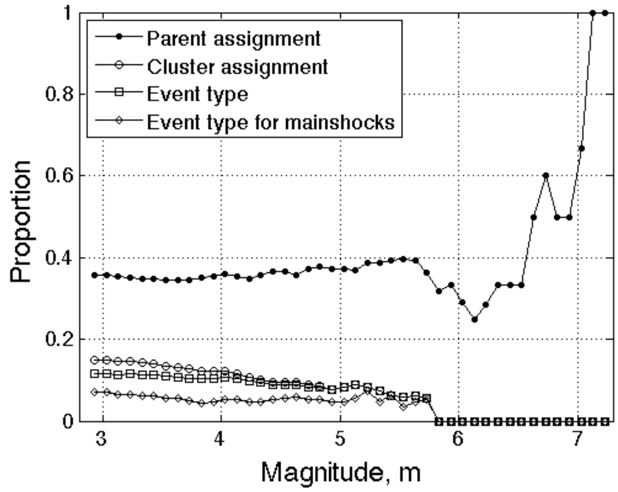

Figure D4: ETAS model – cluster identification errors. The figure shows the proportion of various erroneous identifications for events with magnitude above *m*. Dots – wrong parent assignment; circles – wrong cluster assignment; squares – wrong event type (fore/after/mainshock) assignment, stars – wrong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

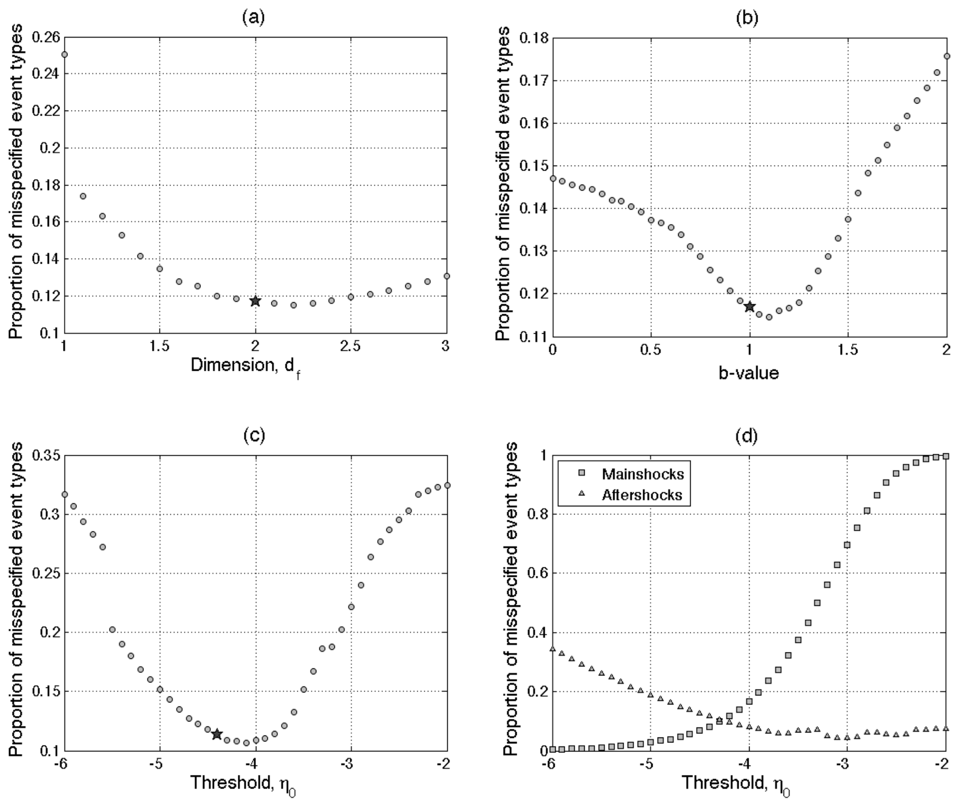

Figure D5: ETAS model – stability of cluster identification. Proportion of events with misspecified event type vs. model numerical parameters. Each panel refers to variation of a single parameter with the other parameters fixed. Stars in panels (a)-(c) refer to the values that correspond to the main version of the analysis, with true values of $d_f = 2$, and $b = 1$, and η_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 η_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).

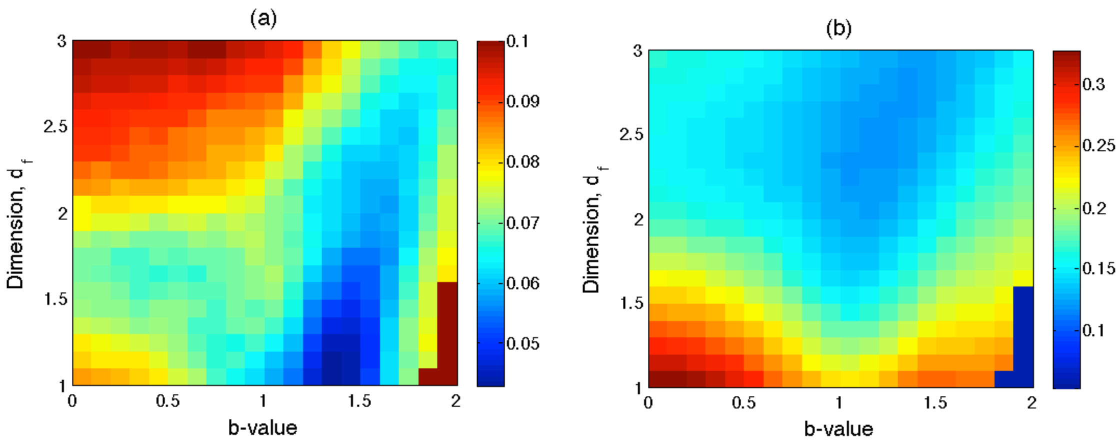

Figure D6: ETAS model – stability of cluster identification. Proportion of misspecified mainshocks (panel a) and aftershocks (panel b) as a function of the pair (b,d_f).

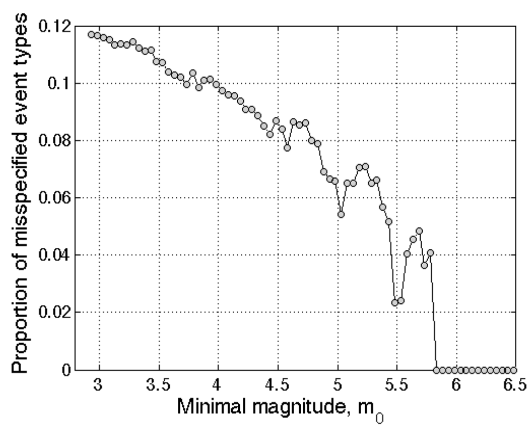

Figure D7: ETAS model – stability of cluster identification. Proportion of events with misspecified types, as a function of minimal magnitude of analysis.

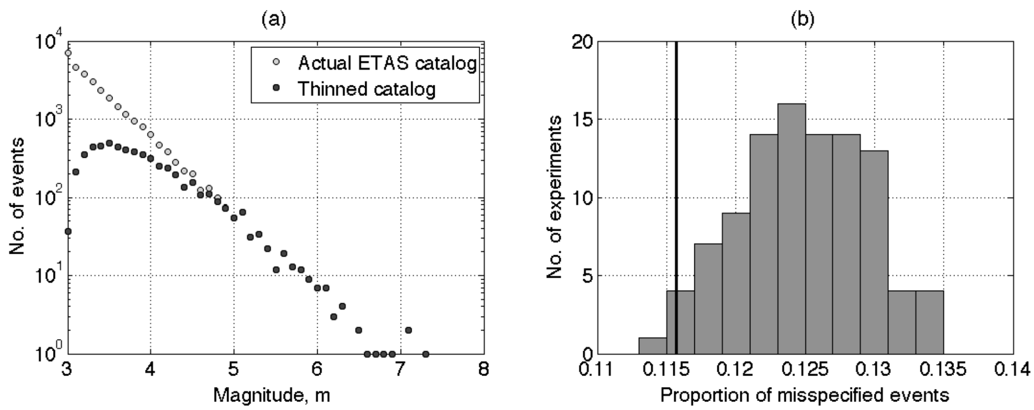

Figure D8: ETAS model – stability of cluster identification in thinning experiment. A thinned catalog is obtained from the actual catalog by removing each event with probability $P(m)$ that decrease linearly from 1 to 0 on the interval $3 \leq m \leq 5$. (a) Magnitude distribution in the actual (black 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.

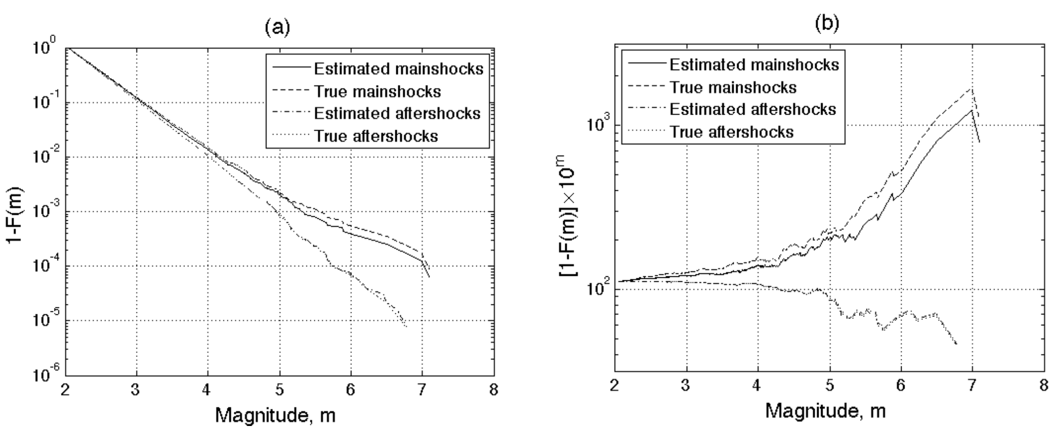

Figure D9: ETAS model – magnitude-frequency distribution. Figure refers to different event types as described in the legend. (a) Proportion 1-$F(m)$ of events with magnitude above m, where $F(m)$ is the empirical cumulative distribution function. (b) Weighted proportion of events with magnitude above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

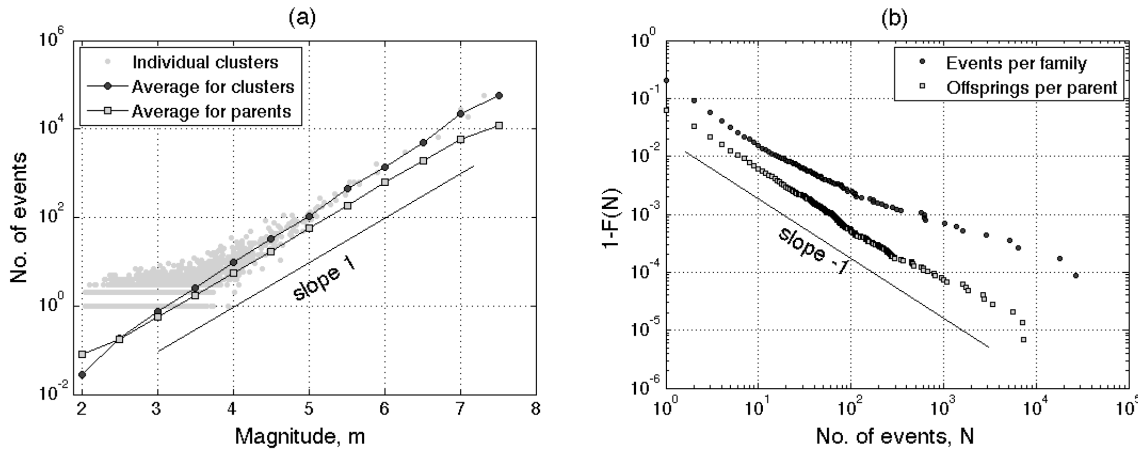

Figure D10: ETAS model – cluster productivity. (a) Number of aftershocks and foreshocks, N-1, in a cluster vs. cluster magnitude m. Black circles – average number of events in a cluster within magnitude window of length 0.5. Grey dots – individual clusters. Squares – average number of offspring per parent.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

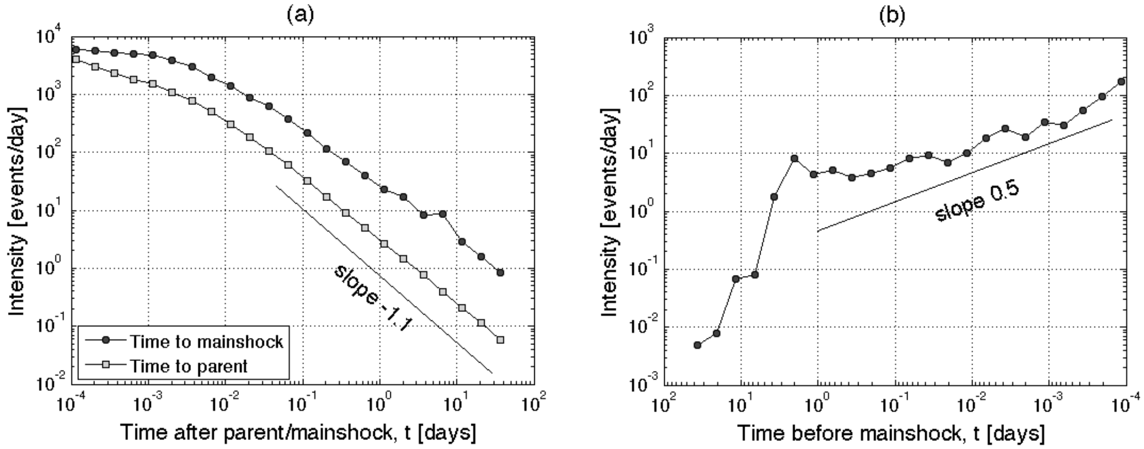

Figure D11: ETAS model – Aftershock and foreshock intensity. (a) Black dots – aftershocks within 50 days of mainshocks with magnitude $m \geq 4$. Squares – first generation offspring. (b) Foreshocks within 50 days of mainshocks with magnitude $m \geq 4$.

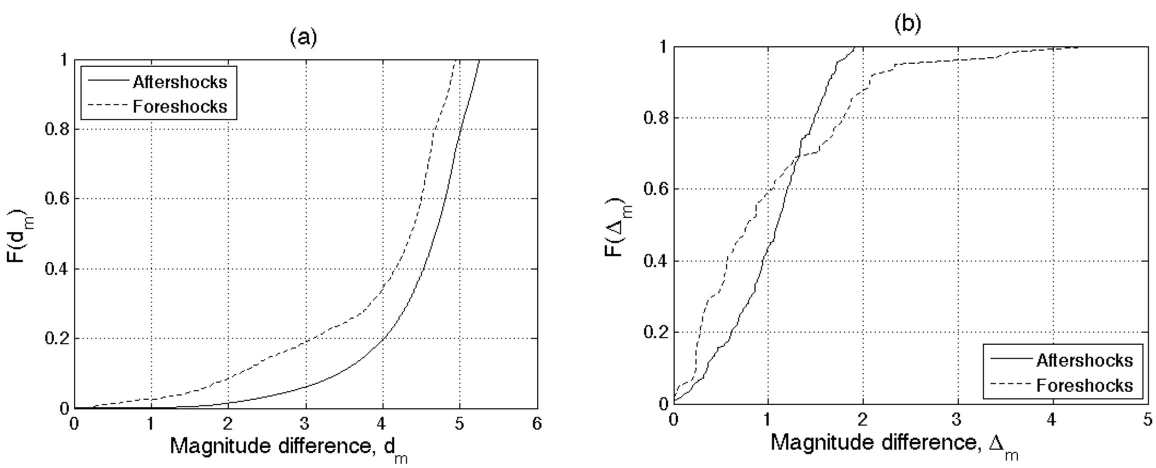

Figure D12: ETAS model – magnitude difference analysis. (a) Magnitude difference d_m between mainshock and each aftershock (solid line) and foreshock (dashed line). (b) Magnitude difference Δ_m between mainshock and the largest aftershock (solid line) and largest foreshock (dashed line). Families with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

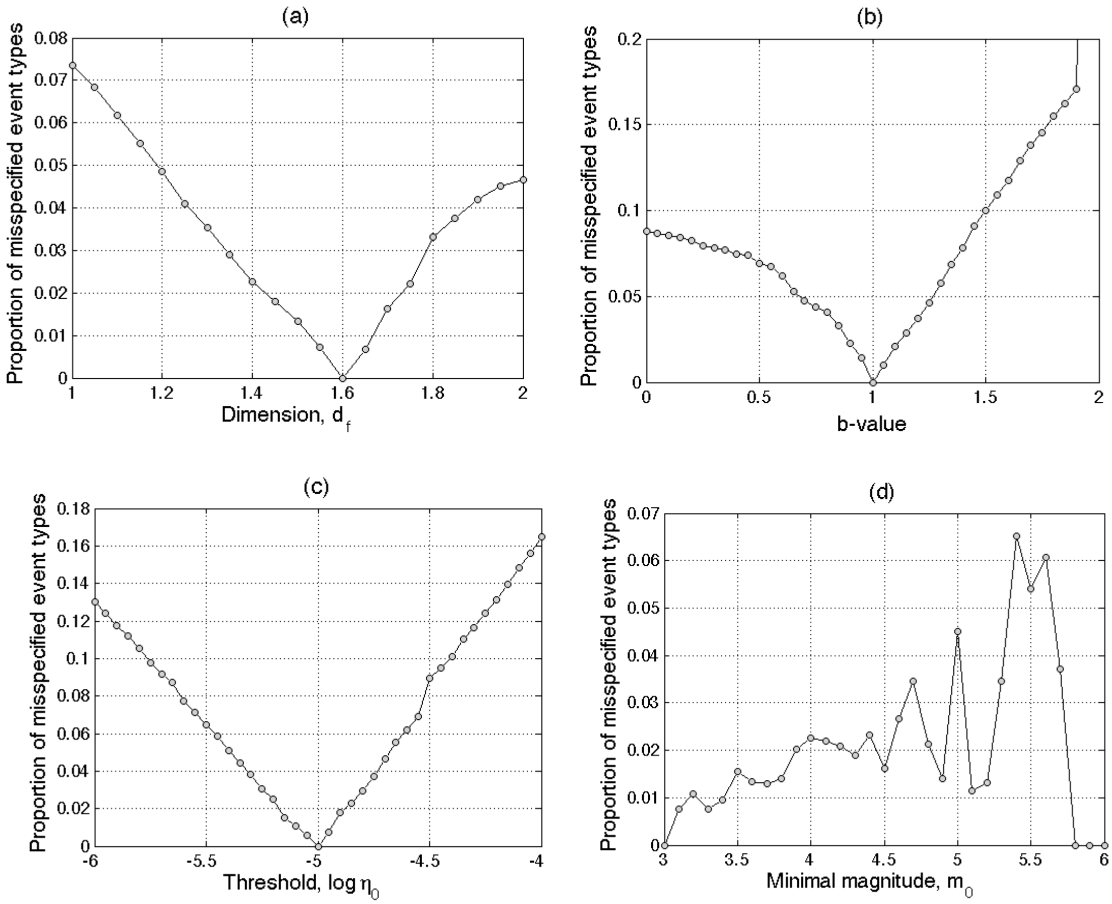

Figure E1: Stability of cluster identification in southern California. Proportion of events with event type different from that obtained in the main version of analysis as a function of algorithm parameter: (a) Fractal dimension of epicenters d_f, (b) b-value, (c) cluster threshold η_0, and (d) minimal magnitude of analysis. The main version of analysis uses $d_f = 1.6$, $b = 1$, $m_0 = 3$, and threshold η_0 estimated from the Gaussian mixture model.