A NEW ESTIMATION OF THE DECAY OF AFTERSHOCK DENSITY WITH DISTANCE TO THE MAINSHOCK

D. Marsan (1), O. Lengliné (1) (2)

(1) Laboratoire de Géophysique Interne et Tectonophysique, Université de Savoie, CNRS, 73376 Le Bourget du Lac, France

(2) now at Ecole et Observatoire des Sciences de la Terre, Université de Strasbourg, CNRS, 5 rue René Descartes, 67084 Strasbourg, France

ABSTRACT:

We investigate how aftershocks are spatially distributed relative to the mainshock. Compared to previous studies, ours focuses on earthquakes causally related to the mainshock, rather than on aftershocks of previous aftershocks. We show that this distinction can be made objectively, but becomes uncertain at long time scales and large distances. Analyzing a regional earthquake dataset, it is found that, at time t following a mainshock of magnitude m, the probability of finding an aftershock at distance r relative to the mainshock fault decays as $r^{-\gamma}$, where γ is typically between 1.7 and 2.1 for $\mathbf{3} \leq m < \mathbf{6}$, and is independent of m, for r less than 10 to 20 km and t less than 1 day. Uncertainties on this probability at larger r and t do not allow for a correct estimation of this spatial decay. We further show that a static stress model coupled with a rate-and-state friction model predicts a similar decay, with an exponent γ=2.2, in the same space and time intervals. This suggests that static stress changes could explain the repartition of aftershocks around the mainshock

even at distances larger than 10 times the rupture length.

Index terms: 7230 (Seismicity and tectonics), 7223 (Earthquake interaction, forecasting, and prediction)

1 – INTRODUCTION

The aim of this study is to measure the spatial clustering of mainshock – aftershock pairs. This clustering is different from the one characterizing any two, possibly unrelated, earthquakes, that is typically quantified by statistical correlation [*Reasenberg,* 1985]. The spatial distribution of aftershocks relative to their mainshock is a signature of the triggering process, and could therefore help us discriminating between potential candidates for this mechanism. In order to effectively measure this causal clustering, we first must isolate mainshock – aftershock pairs. This has traditionally been done using space-time window techniques, similar to declustering methods [*Gardner and Knopoff,* 1974; *Molchan and Dmitrieva,* 1992]. These methods, although very simple to implement, are known to depend on relatively arbitrary parameters. For example, *Felzer et al.* [2004], *Helmstetter et al.* [2005] and *Felzer and Brodsky* [2006] all used a selection criterion to first select mainshocks, based on their relative isolation from previous, close-by, large earthquakes, and then defined aftershocks as earthquakes occurring within a mainshock magnitude-dependent space-time window from the considered mainshock (or from previous aftershocks of this mainshock, as in *Reasenberg* [1985] and *Helmstetter et al.* [2005]). As shown by *Richards-Dinger et al.* (in preparation), in the case of the analysis conducted by *Felzer and Brodsky* [2006], this method does not discriminate efficiently the aftershocks from other, unrelated earthquakes, and can therefore lead to biases in both the results and their interpretation.

In order to avoid these shortcomings, we have developed a different method for isolating mainshock – aftershock pairs, building on *Marsan and Lengliné* [2008]. With this method, like previous ones also based on stochastic modelling [*Kagan and Knopoff*, 1976; *Zhuang et al.*, 2002, 2004], the relationship between any two earthquakes A and B is measured by the estimated probability ω_{AB} that B is an aftershock of A. This is different from usual declustering, for which ω_{AB} can only assume the two values 0 (not an aftershock) or 1 (aftershock). This probability is inverted using a

model of earthquake occurrence that relies only on a linearity and a meanfield hypotheses, and that is therefore much less sensitive to arbitrary parameterization as compared to other declustering methods.

The distribution of mainshock – aftershock distances can be compared to the distribution of distances that characterize unrelated earthquakes, or the so-called 'background' distribution. This comparison is important for understanding earthquake interactions, since the mainshock locally raises the seismicity relatively to this background level. Recently, *Powers and Jordan* (in preparation) studied the distribution of (declustered) earthquake distance r to strike-slip fault planes in California, which can be viewed as characterizing the background distribution. For all their fault segments, the probability density function of *r* is well modeled by $\left(r^2 + d^2\right)^{-a/2}$ where *d* is typically of the order of 200 m to 10 km and is interpreted as the width of the damage zone, while the exponent *a* typically ranges between 1.0 and 2.5, and could be related to the geometrical roughness of the fault. In Section 5, we will reproduce the distribution of mainshock – aftershock distances observed in Section 4 with a model that account for the background seismicity, the latter being characterized by a probability distribution very similar to the one of *Powers and Jordan* (in preparation).

2 – METHOD

We here detail the method used for finding the triggering probabilities linking each pair of earthquakes. This thorough description of the algorithm develops on the initial results discussed in *Marsan and Lengliné* [2008].

2.1 – Relationships between earthquakes:

Direct aftershocks: In the stochastic framework pertaining to this method, every earthquake can potentially be an aftershock of all the previous earthquakes. Similarly to *Zhuang et al.* [2002, 2004], we define ω_{ij} as the probability that earthquake *j* was triggered by earthquake *i*, or, equivalently, that *j* is a direct aftershock of *i*. Causality imposes that $\omega_{ij}=0$ if $t_j < t_i$ + propagation time. In the following analysis, we will simplify this condition to $t_j < t_i$, i.e., we will neglect the seismic wave propagation time.

Using the weights ω_{ij}, one can draw random realisations of a causal chain. In such a chain, each earthquake is the direct aftershock of only one mainshock. This mainshock is drawn randomly, using the weights ω_{ij}, i.e., if ω_{ij} =0.1 then *j* has a 10% chance of being a direct aftershock of *i* for any given chain.

Indirect aftershocks: If, for a given causal chain, *i* is the mainshock of *j*, and *j* is the mainshock of *k*, then *k* is said to be an indirect aftershock of *i*, i.e., it is the aftershock of a previous aftershock of *i*. If this happens 10 times out of 1000 realisations of the causal chain, then the probability ω_{ik}' that *k* is an indirect aftershock of *i* is estimated to 1%. Practically, the probabilities ω' related to indirect aftershocks can thus be computed by a Monte-Carlo method (or stochastic reconstruction, [cf. *Zhuang et al.*, 2002, 2004]).

It is important to stress that ω_{ij} and ω_{ij}' are two distinct probabilities, with two very different meanings: ω_{ij} measures the probability that *i* directly triggered *j*, via a physical mechanism (e.g., stress transfer), while ω_{ij}' is the probability that *j* is an aftershock of a previous aftershock of *i*, and therefore does not imply any direct triggering between *i* and *j*.

We further define the probability that *j* is conditionned on *i* as the sum $\omega_{ij} + \omega_{ij}'$. Its negation $1 - \omega_{ij} - \omega_{ij}'$ is the probability that *j* is unconditionned on *i*, and corresponds to the probability that *j* would

have existed had *i* not occurred. This is different from correlation: for two earthquakes to be correlated, they need a common ancestor (i.e., they both are direct or indirect aftershocks of at least one given earthquake), which does not necessarily require that one is a direct or indirect aftershock of the other.

In Figure 1, we show a simple illustrative example that is not meant to represent a realistic sequence: it involves 5 earthquakes for which we can easily calculate these probabilities ω_{ij}' . All probabilities ω_{ij} are supposed to be known a priori, with values as given in Figure 1. For the four earthquakes B, C, D and E, the probability of being conditionned on A is 1, as A initiated this sequence. We can draw causal chains based on the probabilities ω_{ij}: for example, the causal chain A→B→C→D→E, where → denotes triggering, occurs on average $\omega_{AB}\times\omega_{BC}\times\omega_{CD}\times\omega_{DE}$= 15% of times. In contrast, the chain {A→B, A→C, A→D, A→E} is less frequent, occurring on average $\omega_{AB}\times\omega_{AC}\times\omega_{AD}\times\omega_{AE}$ = 0.5% of times. In the first case, C, D and E are indirect aftershocks of A, while in the second case, they are all direct aftershocks of A. The probability that C is an indirect aftershock of A is $\omega_{AC}' = \omega_{AB}\times\omega_{BC}$ = 0.5, which is also equal to 1 - ω_{AC} since in this example C has a 100% probability of being conditionned on A. For earthquake D, we have $\omega_{AD}' = \omega_{AB}\ \omega_{BD} + \omega_{AB}\ \omega_{BC}\ \omega_{CD} + \omega_{AC}\ \omega_{CD}$ = 1 - ω_{AD} = 0.87, and so on. In general, for large numbers of earthquakes, the computation of ω_{ij}' involves listing all the causal chains such that there exists a path leading from *i* to *j*. This quickly becomes extremely tedious. For example, computing ω_{ij}' this way when 100 earthquakes occurred between *i* and *j* implies considering $1.26\ 10^{30}$ paths... This becomes untractable when studying much larger datasets, as will be done in Section 4 with a catalog of more than 70,000 earthquakes. The Monte-Carlo method described above thus provides a practical alternative, although it only yields an approximate value (sample average).

2.2 – An illustration of how the algorithm works:

We have seen that the probabilities of triggering ω_{ij} are the key ingredient in this type of stochastic

analysis. We now explain how these probabilities can be computed with the iterative algorithm of *Marsan and Lengliné* [2008]. We start by giving a simple example, with the goal of providing an intuitive understanding of how the method works. We study the case of the 5 earthquakes of Figure 1. For simplicity we will assume they all have the same magnitude. We model this sequence by assuming that each earthquake can contribute to the triggering of the subsequent earthquake, according to an intensity λ that can take two possible values λ_1 or λ_2 depending on the time delay between the two earthquakes. Value λ_1 corresponds to triggering at short time scales (duration equal to, e.g., 1 time unit), which is true for pairs AB, DC, CE and DE, while λ_2 is for long-term triggering (duration equal to, e.g., 3 time units), hence pairs AC, AD, AE, BC, BD and BE. We start with the a priori democratic choice $\lambda_1 = \lambda_2 = 1$. We will later show that this choice has no importance, i.e., other a priori choices would yield the same final results, which makes this algorithm very powerful.

The first step is to deduce the probabilities (or weights) ω from λ. Consider for example earthquake D: it receives contributions from A, B and C, with intensities λ_2 , λ_2 and λ_1, respectively. The total intensity for D is thus $\lambda_1 + 2\lambda_2 = 3$, and the weights for D are therefore $\omega_{AD} = \frac{\lambda_2}{\lambda_1 + 2\lambda_2} = \frac{1}{3}$, and similarly $\omega_{BD} = \omega_{CD} = 1/3$.

Knowing all the weights ω_{ij}, we now update λ_1 and λ_2. For example, λ_1 is the mean number of directly triggered aftershocks during the time interval covered by short-term triggering, here of duration 1. It is therefore equal to $\lambda_1 = \frac{1}{5} \times \frac{\omega_{AB} + \omega_{CD} + \omega_{CE} + \omega_{DE}}{1}$, the normalizing prefactor $\frac{1}{5}$ being due to the fact that 5 earthquakes (A to E) can a priori trigger other earthquakes at short time scale, and the normalizing 1 being the duration of the short time triggering interval. This gives $\lambda_1 = 0.367$. Similarly, we obtain $\lambda_2 = 0.144$.

These two steps are then iterated: with the updated λ_1 and λ_2 values, the new weights ω_{ij} are computed, which in turn allows to again update λ_1 and λ_2, and so on. After 9 iterations, both λ_1 and λ_2 become stable with a precision of 10^{-4}, see Figure 2. A convergence criterion can be used to automatically stop the iterations. We finally obtain that $\lambda_1 = 0.515$ and $\lambda_2 = 0.095$. Note that these are numbers of earthquakes per unit time, hence rates. The corresponding weights are those indicated on Figure 1.

This result is independent of the initial values of λ_1 and λ_2, as long as they are non-zeros, cf. Figure 2. This makes the algorithm powerful, since it always converges to the same solution whatever the arbitrarely chosen starting values. The convergence is fast even for large sets of earthquakes (e.g., several minutes on a normal desktop PC for several thousands of earthquakes). The only dependence is on the discretization of the time axis, and, in the full version of the algorithm, of the magnitude and distance axes as well.

We now detail and properly define the algorithm in the general case. A complete summary of the method is given in Annex A.

2.3 – Stochastic models of earthquake occurrence:

Seismicity is modeled as a point process in time and magnitude. Distance r_{ij} between earthquakes i and j is computed as the distance from the fault of i to the hypocenter of j. The data thus consist in a set $\{t_i, m_i\}$ of earthquakes that occur at times t_i and with magnitudes m_i, and that are separated by distances r_{ij}. This set results from a Poisson process in space and time, with expectation $\Lambda([\underline{x}, \underline{x}+\delta\underline{x}],[t,t+\delta t]) = \int_{\underline{x}}^{\underline{x}+\delta\underline{x}} d\underline{x}' \int_{t}^{t+\delta t} dt' \quad \lambda(\underline{x}',t')$, where $\lambda(\underline{x},t)$ is the (positive) rate density, or intensity, at position $\underline{x}$ and time t. Magnitudes are assumed to be independent of time and position,

and to follow an exponential (Gutenberg-Richter) law above threshold magnitude m_c:

$P(m > M) = e^{-\beta(M-m_c)}$ if $M > m_c$, $P(m > M) = 1$ otherwise.

Statistical models of seismicity aim at finding a pertinent representation of the intensity λ and its dependence on time and position. Our understanding of earthquake interactions, i.e., the fact that earthquakes can trigger one another, has prompted the construction of models in which the intensity is conditioned by past earthquakes: earthquake occurrence influence the subsequent seismic activity, generally by increasing it locally for some period of time (aftershocks). A generic linear model is then $\lambda(\underline{x},t) = \lambda_0 + \sum_{i/t_i<t} \lambda_i(\underline{x},t)$ where λ_0 is a term modelling the « background » forcing, that does not depend on past earthquakes, and $\lambda_i(\underline{x},t)$ is the contribution of earthquake *i* on the rate density at (subsequent) time *t* and position $\underline{x}$. This contribution is known as the « triggering kernel », i.e., the intensity of triggering due to a past earthquake.

Specific parameterized expressions of $\lambda_i(\underline{x},t)$ have been proposed, typically of the form $\lambda_i(\underline{x},t) = \lambda(r_i(\underline{x}), t-t_i, m_i)$ where $r_i(\underline{x})$ is the distance of $\underline{x}$ to earthquake *i*. In general, this distance is simply taken as the distance to the epi or hypocenter of *i*, but here it will represent the actual distance to the fault of *i*, i.e., the closest distance from the fault of earthquake *i* to the hypocenter of *j*. This choice has direct implications on the resulting statistics [cf. *Marsan and Lengliné*, 2008, and *Hainzl et al.*, 2008], as will be discussed in Section 5.1. Several studies have tested various forms of $\lambda_i(\underline{x},t)$ that are decoupled in space and time, for example $\lambda_i(\underline{x},t) = Ae^{\alpha m_i}(t+c-t_i)^{-p} s_i(\underline{x})$ where *A*, α, *c*, *p* are parameters, and s_i is a parameterized spatial kernel typically depending on magnitude m_i [*Zhuang et al.*, 2002, 2004; *Console et al.*, 2003; *Helmstetter et al.*, 2005]. While the temporal dependence takes the traditionnal and ubiquitous form of the Omori-Utsu's law, the main issue with such a formulation comes from the spatial dependence, which is far from being well known – to say the least. The possible dependence of this

spatial kernel on time is also generally ignored in these models. It is therefore important to leave the triggering kernel free, i.e., not to impose any priori model. This is central to the present study: since we aim to study how the distances between mainshocks and aftershocks are distributed, it is essential not to impose any a priori model for this distribution.

2.4 – Model-Independent Stochastic Declustering as an Expectation-Maximization (EM) algorithm:

In the Model-Independent Stochastic Declustering (MISD) algorithm of *Marsan and Lengliné* [2008], seismicity is described as the following: an earthquake of magnitude *m* in the magnitude interval $m \in [m_i, m_{i+1}]$ triggers aftershocks with conditional intensity

$\lambda_i(\underline{x},t) = \sum_j \sum_k \lambda_{ijk}\,\theta(t \in [t_j, t_{j+1}])\,\theta(r_i(\underline{x}) \in [r_k, r_{k+1}])$ where λ_{ijk} are the unknowns (the triple indices denote (i) magnitude (j) time (k) distance), $\theta(P) = 1$ if proposition *P* is true, $\theta(P) = 0$ otherwise, and $[t_j, t_{j+1}]$ and $[r_k, r_{k+1}]$ are the discretization intervals in time and distance. This formulation is equivalent to a simple piecewise constant kernel. Background earthquakes occur with constant and uniform rate density λ_0. MISD first requires to define the discretization intervals in magnitude, time and distance, and then amounts to finding the « best » rate densities λ_{ijk} given the data. We here show that this algorithm is in effect an EM algorithm [*Dempster et al.*, 1977]. The use of EM-algorithms for inverting ETAS parameters is described in *Veen and Schoenberg* [2008].

The algorithm works by iterating the two steps:

E-step:

given a priori intensities λ_{ijk} and λ_0, then, for all earthquakes *j*, compute the probabilities (or weights) ω_{ij} and ω_{0j} that earthquake *j* is triggered by earthquake *i* or is a background event, respectively. These probabilities are defined as

$$\omega_{ij} = \frac{\lambda_i(r_{ij}, t_j)}{\lambda_{\mathbf{0}} + \sum_{k<j} \lambda_k(r_{kj}, t_k)}$$

and

$$\omega_{\mathbf{0}j} = \frac{\lambda_{\mathbf{0}}}{\lambda_{\mathbf{0}} + \sum_{k<j} \lambda_k(r_{kj}, t_k)}$$

and sum up to 1: $\omega_{\mathbf{0}j} + \sum_{i<j} \omega_{ij} = \mathbf{1}$.

M-step:

using these weights $\underline{\underline{\omega}}$, find the best set of λ_{ijk} and λ_0. This maximization step corresponds to finding the maximum likelihood estimator (MLE) given the weights. The log-likelihood function unconditioned on the weights is

$$\ell = -\int_V d\underline{x} \int_T dt \lambda(\underline{x}, t) + \sum_i \ln \lambda_i(\underline{x}_i, t_i)$$

and becomes, knowing the weights:

$$\ell(\underline{\underline{\omega}}) = -\sum_{ijk} n_i \lambda_{ijk} \delta t_j \delta V_k - \lambda_{\mathbf{0}} T V + \sum_{ijk} n_{ijk} \ln \lambda_{ijk} + n_{\mathbf{0}} \ln \lambda_{\mathbf{0}} \quad (1)$$

(cf. *Veen and Schoenberg* [2008] for a derivation of this result in the framework of ETAS models) where n_i is the number of earthquakes with magnitude in the interval $[m_i, m_{i+1}]$, n_0 is the « number » of background earthquakes, n_{ijk} is the « number » of earthquakes triggered by a magnitude i earthquake that occurred $t \in [t_j, t_{j+\mathbf{1}}]$ before it and at a distance $r \in [r_k, r_{k+\mathbf{1}}]$, $\delta t_j = t_{j+1} - t_j$ is the duration of time interval j, δV_k is the volume of the shell related to the distance interval k, and T and V are the total duration and volume of the dataset, respectively.

The weights are here used to compute the « numbers »

$$n_{ijk} = \sum_{a \in [m_i, m_{i+\mathbf{1}}]} \sum_{b>a} \omega_{ab} \theta(t_b - t_a \in [t_j, t_{j+\mathbf{1}}]) \theta(r_{ab} \in [r_k, r_{k+\mathbf{1}}]) \quad (2)$$

and $n_{\mathbf{0}} = \sum_i \omega_{\mathbf{0}i}$. The MLE is simply given by

$$\partial_{\lambda_{ijk}} \ell(\underline{\underline{\omega}}) = \mathbf{0} \Rightarrow -n_i \delta t_j \delta V_{ik} + \frac{n_{ijk}}{\lambda_{ijk}} = \mathbf{0} \Rightarrow \lambda_{ijk} = \frac{n_{ijk}}{n_i \delta t_j \delta V_{ik}} \quad (3)$$

and

$$\partial_{\lambda_{\mathbf{0}}} \ell(\underline{\underline{\omega}}) = \mathbf{0} \Rightarrow -T\,V + \frac{n_{\mathbf{0}}}{\lambda_{\mathbf{0}}} = \mathbf{0} \Rightarrow \lambda_{\mathbf{0}} = \frac{n_{\mathbf{0}}}{T\,V} \quad (4).$$

Since we use distances to the fault (cf. Annex B for the computation of these distances), rather than hypocentral distances, δV_k is not equal to $\frac{\mathbf{4}}{\mathbf{3}}\pi(r_{k+\mathbf{1}}^{\mathbf{3}} - r_k^{\mathbf{3}})$. We will show in section 3 that the algorithm effectively depends on the values of δV_k. We will explore two alternative ways: (1) δV_k is defined as the geometrical volume of the shell with distance $r_k < r < r_{k+1}$ to the fault, averaged over all mainshocks in the same magnitude bin. We limit these volumes to *z=0* (surface) and *z=20* km (base of the schizosphere). (2) δV_k is the number of unconditioned earthquakes in the corresponding shell, averaged over all mainshocks. This 2nd way of defining δV_k requires more explanations (see section 3). At this stage, we only need to point out that δV_k depends in both cases on the magnitude bin [m_i, m_{i+1}], and is therefore denoted as δV_{ik} hereinafter.

2.5 – Advantage of an EM-approach to estimating the conditional intensity values:

This EM algorithm allows for much simpler computations than with a classical MLE method, which is known to suffer from practical drawbacks, at least in the case of ETAS models [*Veen and Schoenberg*, 2008]. In particular, the maximization step is computationally simple: the estimator of Equation (3) only requires summing up triggering weights, by using Equation (2). In contrast, a direct ML estimation would use the log-likelihood function of Equation (1), which is here:

$$\ell = -\sum_{ijk} n_i \lambda_{ijk} \delta t_j \delta V_k - \lambda_{\mathbf{0}} T\,V + \sum_a \ln(\lambda_{\mathbf{0}} + \sum_{ijk} n_{ijk}^{(a)} \lambda_{ijk})$$

where the second sum is on all earthquakes *a*, and $n_{ijk}^{(a)}$ is the number of earthquakes preceding *a*

such that their conditional intensity on a is λ_{ijk}. Searching for this minimum involves solving

$\partial_{\lambda_{ijk}} \ell = \mathbf{0} \Rightarrow -n_i \delta t_j \delta V_{ik} + \sum_a \frac{n_{ijk}^{(a)}}{\lambda_{\mathbf{0}} + \sum_{i'j'k'} n_{i'j'k'}^{(a)} \lambda_{i'j'k'}} = \mathbf{0}$, which effectively corresponds to a set of non-linear equations, a non trivial task by any means.

3 – ANALYSIS

3.1 – Data:

We analyze the earthquake catalog of *Shearer et al.* [2005], either keeping all N=72367 magnitude $m \geq \mathbf{2}$ earthquakes or only the N=6190 earthquakes with $m \geq \mathbf{3}$, depending on the analysis. The algorithm requires to define a NxN matrix $\underline{\underline{\omega}}$ for the interaction weights. A parallelized C program was thus written for this analysis, and run on 32 processors. Distances between earthquakes are computed as the shortest distance from the fault of the 1st earthquake to the hypocenter of the 2nd earthquake, see Annex B. A correction factor is introduced to account for transients in completness magnitude, cf. Annex C. The completness magnitude is m_c=2.3 for this dataset, when considered as a whole, and is therefore slightly greater than the minimum magnitude of 2.0 used in some of the analyses. The correction implemented in Annex C however allows to keep this minimum magnitude rather than cutting at m_c. A convergence criterion of 1% is used for stopping the EM-algorithm.

3.2 – Sensitivity on the choice of background model:

It can be shown that the MISD algorithm with an homogeneous background model (i.e., spatially constant background rate λ_0) is ill-defined: the inversion of the triggering kernel and probabilities depend on the total volume V (cf. equation 4) of the analyzed region, which is largely arbitrary (e.g., what is the volume of southern California?). The homogeneity property must therefore be relaxed. A simple approach is to assume the background to depend on position, but to be smooth at a

(possibly spatially variable) wavelength, as for example proposed by *Zhuang et al.* [2002, 2004]. Unfortunately, it can be shown that the resulting inversion then depends on the smoothing scale, which cannot be a priori and non-arbitrarely fixed. Another alternative was investigated by *Marsan and Lengliné* [2008], at the cost of a significantly more complex algorithm. We here prefer to keep the algorithm at its simplest. The modelling of the background seismicity is a delicate issue, that will be examined in details in a further article.

In order to measure the impact of the background seismicity on the distribution of distances between mainshocks and aftershocks, we run two versions of the method: (1) by imposing a zero background $\lambda_0 = 0$ at all time and position, and (2) by imposing an homogeneous background rate-density of $\lambda_0 = 10^{-3}$ earthquake / day / km^3. The probability of being a background earthquake is on average equal to 53% in the latter case, which can be considered as an over-estimation of the actual probability according to previous studies [*Helmstetter et al.*, 2005; *Sornette and Werner*, 2005; *Hainzl et al.*, 2006]. We therefore consider that these two versions provide conservative lower and upper bounds regarding to the influence of the background term on the distance statistics. Kernel values that are the same for the two runs are therefore insensitive to the background model, and will be considered as reliable.

3.3 – Shell volumes:

As already explained in section 2, the values of δV_{ik} are particularly important in this method. They are used to normalize the rate-densities in Equation (3), and thus directly impact the results. We define r_{ij}, the distance from earthquake *i* to earthquake *j*, as the distance from the fault of *i* to the hypocenter of *j*. We use two distinct sets of shell volumes in the following calculations.

The first set simply corresponds to the mean geometrical (Euclidean) volumes of the shells surrounding the faults of $m_i \leq m < m_{i+1}$ earthquakes. They increase with magnitude since the faults

then become bigger. We limit the depth of the shells to 20 km, corresponding to the base of the schizosphere. Table 1 details these volumes.

The second set of δV_{ik} values is dictated by the fact that the damage zone and fault distribution surrounding the mainshock faults are not densely populating the shell volumes. In fact, a fractal distribution would imply that δV_{ik} should grow with distance interval $[r_k, r_{k+1}]$ more slowly than $(r_{k+1}^3 - r_k^3)$. The potential for a mainshock to trigger aftershocks in such a distance interval is thus not directly proportionnal to the geometrical volume of this shell, but rather to the number of faults that are located in it. In order to « count » this number, we use the probabilities of being unconditioned to the mainshock: the number of faults within distance $[r_k, r_{k+1}]$ on which earthquake i can act upon is equal to the sum of the earthquakes within this shell that are unconditioned to the occurrence of i. We thus define this second set ofδV_{ik} values in the following way:

- for all n_i earthquakes (with index a) in the magnitude bin $[m_i, m_{i+1}]$, we search for the earthquakes located within r_k and r_{k+1} of the fault.
- We then sum up their probabilities of being unconditioned on earthquakes a. For any earthquake b, this corresponds to $1 - \omega_{ab} - \omega_{ab}'$, cf. section 1.
- These probabilities are then averaged over all n_i earthquakes a, and this final number is used as δV_{ik}:

$$\delta V_{ik} = \frac{1}{n_i} \sum_{a / m_a \in [m_i, m_{i+1}]} \sum_{b / r_{ab} \in [r_k, r_{k+1}]} 1 - \omega_{ab} - \omega_{ab}'$$

Note that δV_{ik} is defined up to an arbitrary multiplicative constant.

To compute δV_{ik} this way, we first need to know the probabilities ω_{ij}, which themselves require the knowledge of δV_{ik}. We thus proceed by initializing δV_{ik} to non-zero values, run the algorithm until convergence is reached, and use the obtained ω_{ij} to update te values of δV_{ik}. This is reiterated until these values do not change any longer. We have tested several initial conditions for δV_{ik}: (1) the

values of Table 1, (2) the spherical shell volumes $\delta V_{ik} = \frac{4}{3}\pi(r_{k+1}^3 - r_k^3)$, and (3) $\delta V_{ik} = 1$. We end up in each case with the same final values of δV_{ik}, as shown in Figure 3 and in Table 2. The δV_{ik} values computed this way are relatively similar to the values that would be obtained by counting the preceding, rather than unconditioned, earthquakes. The preceding earthquakes are of course unconditioned.

3.4 – The example of the Landers earthquake

We illustrate the results of this analysis by examining the relationship of $m \geq 3$ earthquakes to the 1992, Mw7.3 Landers earthquake. The δV_{ik} values are those estimated by counting unconditioned earthquakes. A zero background is assumed: $\lambda_0 = 0$. Figure 4a shows the time evolution of the direct aftershocks, of all (i.e., direct and indirect) aftershocks, and of the earthquakes unconditioned to Landers (i.e., earthquakes that are neither direct nor indirect aftershocks). The cumulative numbers are obtained by adding the probabilities $\omega_{L,j}$ (for direct aftershocks) and $\omega_{L,j}'$ (for indirect aftershocks), where *L* refers to Landers. The black line (unconditioned earthquakes) therefore corresponds to the seismicity 'declustered' from the influence of Landers. It is observed to undergo a slow-down at about ~ 2000 days, hence ~ 3 years prior to Landers, and then to keep a relatively stationnary trend. We obtain $\sum_j \omega_{L,j}$ = 569.48 direct aftershocks, which are spatially distributed as shown in Figure 4b, and $\sum_j \omega_{L,j} + \omega_{L,j}'$ = 2677.89 direct and indirect aftershocks, as shown in Figure 4c. The direct aftershocks only last for about a year: 94% of them occur during the first year. They are concentrated along the main rupture, and extend very little outside this zone, except towards the North, which suggests dynamic triggering [*Kilb et al.*, 2000; *Gomberg et al.*, 2003]. In comparison, the indirect aftershocks last much longer: the sequence is not over yet by the end of the dataset, more than 10 years after the mainshock. They also act to extend the influence of the mainshock, up to a much larger zone, in agreement with similar findings by *Ziv* [2006]. For

example, the earthquakes located in the rupture zones of the Big Bear and Hector Mine earthquakes are significantly conditionned on Landers (with an average probability greater than 95% for Big Bear, and between 40% and 60% for Hector Mine). Landers promoted these two large shocks, and their aftershocks are therefore likely to be aftershocks of previous aftershocks of Landers. This is extremely likely for the Big Bear aftershocks, as could be naturally expected. The probability of being conditionned on Landers is lower for the Hector Mine aftershocks, but is far from negligible: this sequence is therefore possibly promoted by the Landers mainshock and / or its aftershocks, as already hypothesized by *Felzer et al.* [2002] and *Ziv* [2006].

4 – DISTANCES FROM MAIN FAULT TO DIRECT AFTERSHOCKS

4.1 – Estimating the linear density

The kernel values λ_{ijk} give the rate – densities of direct aftershocks following mainshocks of magnitude $m_i \leq m < m_{i+1}$, after a delay $t_j \leq t < t_{j+1}$ and at a distance $r_k \leq r < r_{k+1}$. To compare our results to those of *Felzer and Brodsky* [2006] and *Richards-Dinger et al.* (in preparation), we compute the linear density, still for the same intervals of mainshock magnitude and time delays, as:

$$f_{ij}(r_k \leq r < r_{k+1}) = \lambda_{ijk} \frac{\delta V_{ik}}{r_{k+1} - r_k}$$

This is done for the four runs of our MISD algorithm, corresponding to the four combinations of (1) δV_{ik} computed as geometrical Euclidean shell volumes, or as the number of unconditioned earthquakes in these volumes, and (2) $\lambda_0 = 0$ or 10^{-3} per km^3 and day, which for both δV_{ik} sets gives a 53% proportion of background earthquakes, see section 3. For each $f_{ij}(r_k \leq r < r_{k+1})$ density value, we compute from these four values the mean and the standard deviation, and only consider as robust estimates the mean values greater than the corresponding standard deviation, hence with less than 100% relative uncertainty. Power-law best fits are then estimated, only using the robust values.

We also only display the robust values on the corresponding graphs (Figures 5 and 6). This implies that the power-law trends could possibly extend to greater distances, although we cannot validate this.

Figures 5 and 6 show the linear density for $\mathbf{3} \le m < \mathbf{4}$ and $\mathbf{5} \le m < \mathbf{6}$ respectively, and for two time intervals (first 15 minutes, and 12 to 24 hours after the mainshock). Robust estimates are found for distances up to 16 or 32 km, which is more than 20 times the rupture length of $\mathbf{3} \le m < \mathbf{4}$ earthquakes, and a few times this of $\mathbf{5} \le m < \mathbf{6}$ earthquakes. The dispersion in the four runs is, by far, mostly caused by changing the background term λ_0. This dispersion increases with distance to the main fault, as linking aftershocks to mainshocks become more and more sensitive to the parameterization of the algorithm.

In order to explore a greater range of mainshock magnitudes, we also compute the linear density for the Landers earthquake (m=7.3), for the 0 – 100 days interval, see Figure 7. The linear density for the direct aftershocks is compared to the linear density of all the earthquakes following Landers in the first 100 days. Clearly, and as expected, the direct aftershocks are more clustered close to the mainfault. We obtain $\gamma = \mathbf{1.66} \pm \mathbf{0.08}$, when fitting over 1 – 100 km. At distances greater than 10 km, roughly 10% only of the earthquakes occurring in the first 100 days can be considered as direct aftershocks of Landers, the remaining 90% being mostly aftershocks of previous aftershocks, i.e., indirect aftershocks. The importance of multiple interactions at even larger distances was already advocated by *Ziv* [2006].

4.2 – Comparison with the analysis of *Felzer and Brodsky* [2006]

In this analysis, the uncertainty on the linear density is estimated by exploring several different ways for computing the triggering kernels. Similarly, we test and compare the results of *Felzer and Brodsky* [2006] to ours by performing their analysis with different sets of (declustering) parameters.

We recall that, with their approach, an earthquake is not considered as a mainshock if there exists a larger earthquake within a radius L that occurred less than T_1 before or T_2 after it. For each of the two time intervals (0 – 15 minutes and 12 – 24 hours after the mainshock), we define 5 sets of {L, T_1, T_2} parameters, see Table 3, paying attention to the requirement that $T_1 >> T_2 > 15$ minutes or 24 hours. For each of these sets, we count the number n of earthquakes occurring within the considered time window (i.e., first 15 minutes, and 12 to 24 hours after the mainshock) and sort them by distance intervals. We further modify the analysis of *Felzer and Brodsky* [2006] by accounting for background seismicity: we count the number of earthquakes occurring in time intervals with similar durations (15 minutes or 12 hours) but taken at random, excluding the 100 days before and 100 days after the mainshock. We then estimate the rate change $\delta\lambda$ corresponding to the rate of aftershocks as the mean of the positively-defined random variable X with probability density $f_X(x) = A\int_0^\infty d\mu \; f_0(\mu) \; f(\mu + x)$ where $f_0(\mu)$ is the probability density of the background rate μ, $f(\mu+x)$ is the probability density of the rate of posterior earthquakes, equal to the background rate μ plus the aftershock rate x, and A is a normalizing constant such that $\int_0^\infty dx \; f_X(x) = 1$. The densities f_0 and f are constructed using a Poisson law [*Marsan and Nalbant*, 2005]. The linear density for the distance interval $r_k \le r < r_{k+1}$ is then estimated as $\frac{\overline{X}}{r_{k+1} - r_k}$. The computation of the background density is performed independently for each {L, T_1, T_2} set. We finally compute the mean and the standard deviation from the five estimates, and plot them on Figures 5 and 6. As previously, we only use those values for which the standard deviation is less than the mean (relative error on the true value – not its logarithm – less than 100%).

For the first time window (less than 15 minutes after the mainshock), the two methods give similar results, see Figure 5 (top graph). This means that the two methods agree on which pairs can be considered as mainshock – aftershock. At this short time scale, the causal relationship is relatively

obvious. We find that the linear density can be fitted with a power-law $r^{-\gamma}$ with exponent $\gamma = \mathbf{1.76 \pm 0.35}$ (our method, fitted over 0 – 16 km) and $\gamma = \mathbf{1.62 \pm 0.17}$ (method modified from *Felzer and Brodksy* [2006], fitted over 0 – 32 km). We will however show in Section 5.2 that this decay can be reproduced by a static stress model, in opposition to the dynamic stress model of *Felzer and Brodsky* [2006].

Looking further ahead in time (12 to 24 hours after the mainshock), the two methods give distinct results, and the uncertainties increase significantly (Figure 5, bottom graph). The method modified from *Felzer and Brodsky* [2006] over-estimates the linear density at large distances, as compared to our method: too many mainshock – aftershock pairs are found at long time intervals and large distances to the main fault. A similar conclusion regarding to this over-estimation was reached by *Richards-Dinger et al.* (in preparation). It can be either due to two effects: (a) 'mainshock' A and 'aftershock' B are actually both aftershocks of a common, previous mainshock C, i.e., C→A and C→B; or (b) 'aftershock' B is actually triggered by a previous aftershock C of mainshock A, i.e., A→C→B. Our method accounts for these two possibilities, and hence results in a lower estimate. We find that $\gamma = \mathbf{1.97 \pm 1.11}$ (fitted over 0 – 16 km), while $\gamma = \mathbf{1.43 \pm 0.33}$ as given using the method modified from *Felzer and Brodksy* [2006], fitted over 0 – 32 km. The large uncertainty (caused by the large error bars on the linear density) shows that deciphering the causal chain of triggering is not straightforward at these time and spatial scales.

Similar conclusions are reached when analyzing larger mainshocks, as shown in Figure 6 for $\mathbf{5} \leq m < \mathbf{6}$ mainshocks. The agreement between our method and the analysis modified from *Felzer and Brodsky* [2006] is good at short time scales (first 15 minutes) when there is a greater chance to find a causal connection between the two earthquakes. We obtain $\gamma = \mathbf{1.94 \pm 0.31}$ with our method, while the second analysis gives $\gamma = \mathbf{1.67 \pm 0.04}$, both fitted over 0 – 32 km.

For these mainshocks also, the two methods disagree at longer times (12 to 24 hours), the second method again over-estimating the number of aftershocks. We obtain $\gamma = \mathbf{1.87 \pm 0.61}$ with our method, fitted over 0 – 16 km, while the second analysis gives $\gamma = \mathbf{1.61 \pm 0.06}$, fitted over 0 – 32 km.

5 – DISCUSSION

5.1 – Limits of the method and uncertainties

The use of a stochastic model and inversion to estimate the probability that an earthquake is an aftershock of another earthquake is required to avoid arbitrary selection rules, that will always impact the results. The method proposed here is, so far, the less prone to arbitrary choices; it is thus an improvement of previous studies on mainshock – aftershock distances distribution. However, two thorny issues can still be identified.

First, it is important to use distances to the fault rather than hypocentral distances when optimizing stochastic models of seismicity [*Hainzl et al.*, 2008]. In particular, *Marsan and Lengliné* [2008] showed that the triggering caused by large mainshocks is underestimated if considering hypocentral distances. Indeed, the distances between mainshock and aftershocks can only increase if using hypocentral distances instead of distances to the rupture; more importantly, the stronger this effect for large mainshocks, hence a 30% decrease in the productivity exponent for direct triggering, as evidenced by *Marsan and Lengliné* [2008] using the same dataset as in the present study.

As a result, we face the problem of defining the shell volumes (δV, see section 3). Using geometrical Euclidean volumes implies that the biggest shocks have relatively little impact on the seismicity dynamics, to an extent that seems unrealistic. We have therefore proposed an alternative

way of computing the δV values that gives seemingly acceptable results in terms of the triggering kernels. We emphasize that this issue is not stictly related to our method: the inversion of ETAS model parameters would also suffer the same problem if using distance to faults. This is therefore an issue that is general to all stochastic declustering methods.

Second, the choice of a background model also strongly impacts the inversion. The smoothing method by *Zhuang et al*. [2002, 2004] depends on two arbitrary parameters, which possibly can be reduced to just one (smoothing wavelength). However, the latter is still arbitrary. A simple argument would be to set it to the location error estimate. However, this is not a robust choice: if, in the future, a better localized earthquake catalogue is analyzed, then this wavelength would need to be made shorter, automatically resulting in a greater proportion of background events. Alternative solutions must then be thought of, that would reduce this lack of robustness to an acceptable level. Being aware of this problem, we have here adopted the strategy of upper and lower bounding our statistics by running two versions of the algorithm, that over and under-estimates the importance of background seismicity.

As evidenced by the error bars in Figures 5 and 6, this strategy leads to large uncertainties. In this analysis, the number of earthquakes is very large (>70,000) and the uncertainties are therefore not due to limited sampling, but rather to the limits discussed above. Despite the recent development of new methods, we are still struggling with this difficult inverse problem.

5.2 – Mainshock – aftershock distances are coherent with static stress changes

The decay of the distribution with distance is relatively steep, cf. Figures 5 and 6. The exponent γ as estimated by our method typically ranges from 1.70 to 2.10, accounting for the large uncertainties. The estimates using the method ~~by~~ modified from *Felzer and Brodsky* [2006], i.e., after removing the background earthquakes, yield a more restricted interval of about 1.60 to 1.70. As already

explained, the difference between the two methods is likely to be due to an over-estimation in the latter analysis of the number of aftershocks, some of them actually being indirect aftershocks, especially when looking at long time scales. The null hypothesis that the exponent γ is independent of the magnitude of the mainshock cannot be rejected given our results.

We model the triggering of aftershocks with a static stress model coupled with rate and state friction. We focus on the case of a generic m=3 earthquake, in order to compare the results of the model with the estimated density shown in Figure 5. The rupture is a square dislocation of length L = 400 m, and u = 1 cm slip. The stress tensor is computed using the Fortran program of *Gomberg and Ellis* [1994], that uses *Okada* [1992] 's equations. The rupture is modelled at great depth, in order to suppress the effects due to the free surface. Young's modulus is equal to 70 GPa, Poisson's coefficient is 0.25, and the static friction coefficient is 0.4. We compute the Coulomb static stress distribution for target faults located at distance $R_1<r<R_2$ from the rupture, keeping their focal mechanisms identical to the one of the mainshock (same strike, dip and rake angles). We show in Figure 8 the resulting means and standard deviations of this stress change, vs. distance. The variability of the stress values within a given shell is very high; the standard deviation is typically 7 times the absolute value of the mean, on average. This is due to the random locations of the target faults within the shell. The near-field (r<400m) corresponds to a stress shadow (i.e., the stress change is negative on average), as expected since the target faults are parallel to the causative fault. However, since the stress is very heterogeneous, concentrations of positive stress changes will control the evolution of seismicity at short time scales, resulting in strong aftershock triggering in this stress shadow [*Helmstetter and Shaw*, 2006; *Marsan*, 2006], at least for the time windows we are interested in (typically less than one day). The stress change decaying as r^{-3}, it becomes quickly very small, i.e., less than 100 Pa for r>3 km. However, as we will see, even this very small stress change, because it affects large volumes, can trigger aftershocks, in coherence with what we observe in our analysis.

The rate-and-state friction model [*Dieterich*, 1994] is then used to translate the stress changes into aftershock density. To do so, we compute for a given shell $R_1<r<R_2$ the mean number of aftershock occurrences in a time interval of duration t after the mainshock, as $n_{trig} = \mu(R_1, R_2) \times t_a \left[\ln(e^{t/t_a} + e^{-\tau/A\sigma} - \mathbf{1}) + \tau/A\sigma - t/t_a \right]$, where $\mu(R_1, R_2)$ is the background rate for this shell, t_a is a parameter related to the nucleation cycle duration, τ are the stress changes for this shell, and $A\sigma$ a parameter typically in the range of 0.01 to 0.1 MPa. Since only the t/t_a ratio matters, we take $t/t_a = 10^{-5}$, which, for, e.g., t_a = 10 years, corresponds to $t \sim 1$ hour . As already discussed in section 3.3, we use for the background rate the rate of unconditionned earthquakes, cf. Table 2. This rate grows as $r^{1.65}$ at short distances, for $\mathbf{3} \le m < \mathbf{4}$ mainshocks. We therefore take $\mu(R_1, R_2)$ proportionnal to $R_2^{\mathbf{1.65}} - R_1^{\mathbf{1.65}}$. The density for $R_1<r<R_2$ is then computed as the averaged n_{trig}, normalized by R_2-R_1 (linear density); it is defined up to a multiplicative factor, which does not depend on distance. Therefore, only its relative change with distance is investigated. Figure 9 displays this linear density, for 3 values of $A\sigma$ in the 0.01 to 0.1 MPa interval. In the far field (0.7 to 15 km away from the fault), a $r^{-2.2}$ power-law decay is observed. This distance interval is similar to the one of Figure 5a. Its lower cut-off is conditionned by the rupture length (400 m). The exponent predicted by the model is slightly greater than the $\gamma = \mathbf{1.76} \pm \mathbf{0.35}$ value of Figure 5a, although it is not far off the 1.70 – 2.10 interval typically observed using our method. The choice of $A\sigma$ influences this exponent, which is therefore not well constrained by the model, at least in the near-field (r < 0.7 km).

We conclude that static stress triggering, even at distances equal to several times the rupture length of a $\mathbf{3} \le m < \mathbf{4}$ mainshock, can explain the observed triggering. Based on a similar model, the same conclusion is reached by *Hainzl et al.* [2009; in preparation], in the particular case of the Landers earthquake (see also *Richards-Dinger et al.*, in preparation): the distance distribution for the direct aftershocks of Landers is well fitted by a static stress and rate-and-state friction model, assuming

588 that the receiver faults have, at any location, all possible orientations [*Hainzl et al.*, in press].
589
590 The question remains as whether tiny stress changes (<100 Pa) can produce a noticeable change in
591 seismicity at distances greater than ~ 3 km. In comparison, tidal stress changes are larger, but do not
592 act permanently as their periodicity implies both loading and unloading of faults. It is therefore
593 difficult to compare these two phenomena. Analyzing a small set of intermediate-sized shocks, *Ziv*
594 *and Rubin* [2000] argued that static stress changes less than 1 kPa could affect the occurrence of
595 earthquakes. Unfortunately, similar conclusions cannot be met when studying aftershock sequences
596 (e.g., *Hardebeck et al.*, 1998), mainly owing to the difficulty of separating aftershocks from
597 background seismicity. Our study however suggests that this separation can be done objectively,
598 and that the power-law decay of earthquake triggering with distance does not have any apparent
599 cut-off scale. Therefore, static stress triggering could indeed exist even for very small stress
600 changes, the very large volumes affected by these changes partly counter-balancing their weakness.
601
602 **6 – CONCLUSIONS**
603
604 The spatial pattern of aftershock locations relative to the main fault is likely to yield important
605 information regarding to the processes involved in earthquake triggering. One way of studying this
606 spatial pattern is by analyzing the distribution of distances between mainshocks and aftershocks,
607 and to investigate which triggering model is able to reproduce this distribution. However, this
608 requires to select only actual aftershocks, rather than indirect aftershocks or even uncorrelated
609 earthquakes.
610
611 In the last few years, new methods have been proposed that can statistically estimate the
612 relationships between any two earthquakes A and B, in particular by computing the probability that
613 A could have triggered B. Alternatively, this probability can be seen as the influence that A had on

614 the occurrence of B, assuming that B results from the whole seismicity history rather than just one
615 single triggering earthquake.
616
617 These methods rely on various hypotheses (linearity, in particular), which, given our present
618 knowledge on earthquake processes, are difficult to validate or refute. The MISD algorithm
619 described here is based on a minimal set of such hypotheses. Despite this minimal a priori belief,
620 we show that its results significantly depend on the choice of normalizing shell volumes. Testing
621 several reasonable choices for these volume values, as well as different estimates for the
622 background seismicity, we end up with uncertainties that are significant at large distances
623 (typically greater than 20 km away from the main fault) and long time scales (typically greater than
624 1 day past the mainshock occurrence) for small to intermediate mainshocks, forcing us to only
625 investigate small temporal and spatial scales.
626
627 We find that the distance distribution decays according to a power-law of the distance, with an
628 exponent that typically lies in the 1.70 – 2.10 interval. This is true for small ($\mathbf{3} \leq m < \mathbf{4}$) and
629 intermediate ($\mathbf{5} \leq m < \mathbf{6}$) mainshocks. For the specific case of the m = 7.3 Landers earthquake we
630 obtain a decay best described by an exponent of $\mathbf{1.66 \pm 0.08}$ for the 0 – 100 days time interval,
631 hence coherent with this 1.70 – 2.10 range, cf. Figure 7. These results suggest that the mainshock
632 magnitude has little influence on the distribution, as long as distances to the main fault rather than
633 hypo- or epi-center distances are considered.
634
635 Although the observed triggering is significant even at many (up to more than 10, for small
636 mainshocks) rupture lengths, a static stress and rate-and-state friction model yields a linear density
637 that is coherent with the observation. Triggering by static stress cannot therefore be discounted on
638 the basis of these observations: the distribution of mainshock – aftershock distances cannot help
639 discriminating between static and dynamic stress triggering, at least with our present know-how

concerning the distinction between direct and indirect aftershocks.

Acknowledgments

This work benefited from financial support by ANR project ASEISMIC. O. L. was supported by European Commission FP6 project VOLUME (contract 18471). We thank Sebastian Hainzl for stimulating discussions, as well as Ross Stein, an anonymous reviewer and the associate editor for constructive reviews.

REFERENCES

Console R., M. Murru, and A. M. Lombardi (2003), Refining earthquake clustering models, *J. Geophys. Res.*, 108, B10, 2468.

Dempster A. P., Laird N. M., and Rubin D. B. (1977), Maximum likelihood from incomplete data via the EM algorithm, *J. Royal Statistical Soc.*, series B, 39, 1, 1-38.

Dieterich J. H. (1994), A constitutive law for rate of earthquake production and its application to earthquake clustering, *J. Geophys. Res.*, 99, 2601-2618.

Felzer K. R., T. W. Becker, R. E. Abercrombie, G. Ekström, and J. R. Rice (2002), Triggering of the 1999 Mw7.1 Hector Mine earthquake by aftershocks of the 1992 Mw7.3 Landers earthquake, *J. Geophys. Res.*, 107, B9, 2190.

Felzer K. R., R. E. Abercrombie, and G. Ekstrom (2004), A common origin for aftershocks, foreshocks, and multiplets, *Bull. Seismol. Soc. Am.*, 94, 88 – 99.

666 Felzer K. R., and E. E. Brodsky (2006), Decay of aftershock density with distance indicates
667 triggering by dynamic stress, *Nature*, 441, 735 – 737.
668
669 Gardner J., and L. Knopoff (1974), Is the sequence of earthquakes in southern California, with
670 aftershock removed, Poissonian?, *Bull. Seismol. Soc. Am.*, 64, 1363-1367.
671
672 Gomberg J., and M. Ellis (1994), Topography and tectonics of the central New Madrid seismic zone
673 – results of numerical experiments using a 3-dimensional boundary-element program, *J. Geophys.*
674 *Res.*, 99, 20299-20310.
675
676 Gomberg J., P. Bodin, and P. A. Reasenberg (2003), Observing earthquakes triggered in the near
677 field by dynamic deformations, *Bull. Seismol. Soc. Am.*, 93, 118-138.
678
679 Hainzl S., F. Scherbaum, and C. Beauval (2006), Estimating background activity based on
680 interevent-time distribution, *Bull. Seismol. Soc. Am.*, 96, 313-320.
681
682 Hainzl S., Christophersen A., and Enescu B. (2008), Impact of earthquake rupture extensions on
683 parameter estimations of point-process models, *Bull. Seismol. Soc. Am.*, 98, 2066-2072.
684
685 Hainzl S., B. Enescu, M. Cocco, J. Woessner, F. Catalli, R. Wang, and F. Roth (2009), Aftershock
686 modeling based on uncertain stress calculations, *J. Geophys. Res.*, 114, B05309.
687
688 Hainzl S., G. Zöller, and R. Wang, Impact of the receiver fault distribution on aftershock activity, *J.*
689 *Geophys. Res.*, in press.
690
691 Hardebeck J. L., J. J. Nazareth, and E. Hauksson (1998), The static stress change triggering model:

692 Constraints from two southern California aftershock sequences, *J. Geophys. Res.*, 103, 24427-
693 24437.
694
695 Hardebeck J. L., and P.M. Shearer (2003), Using S/P amplitudes ratio to constrain the focal
696 mechanisms of small earthquakes, *Bull. Seismol. Soc. Am.*, 93, 2434-2444.
697
698 Helmstetter A., Y. Y. Kagan, and D. D. Jackson (2005), Importance of small earthquakes for stress
699 transfers and earthquake triggering, *J. Geophys. Res.*, 110, B05S08.
700
701 Helmstetter A., and B. E. Shaw (2006), Relation between stress heterogeneity and aftershock rate in
702 the rate-and-state model, *J. Geophys. Res.*, 111, B07304.
703
704 Kagan Y., and L. Knopoff (1976), Statistical search for non-random feature of the seismicity of
705 strong earthquakes, *Phys. Earth Planet. Int.*, 12, 291-318.
706
707 Kilb D., J. Gomberg, and P. Bodin (2000), Triggering of earthquake aftershocks by dynamic
708 stresses, *Nature*, 408, 570-574.
709
710 Marsan D., and S. S. Nalbant (2005), Methods for measuring seismicity rate changes: A review and
711 a study of how the M-w 7.3 Landers earthquake affected the aftershock sequence of the M-w 6.1
712 Joshua Tree earthquake, *PAGEOPH*, 162, 1151-1185.
713
714 Marsan D. (2006), Can coseismic stress variability suppress seismicity shadows? Insights from a
715 rate-and-state friction model, *J. Geophys. Res.*, 111, B06305.
716
717 Marsan D., and Lengliné O. (2008), Extending earthquakes' reach through cascading, *Science*, 319,

718 5866, 1076-1079.
719
720 Molchan G. M., and O. E. Dmitrieva (1992), Aftershock identification - methods and new
721 approaches, *Geophys. J. Int.*, 109, 501.
722
723 Okada Y. (1992), Internal deformation due to shear and tensile faults in a half-space, *Bull. Seismol.*
724 *Soc. Am.*, 82, 1018-1040.
725
726 Peng Z., J. E. Vidale, M. Ishii, and A. Helmstetter (2007), Seismicity rate immediately before and
727 after main shock rupture from high-frequency waveforms in Japan, *J. Geophys. Res.*, 112, B03306,
728 doi:10.1029/2006JB004386.
729
730 Powers P. M., and T. H. Jordan, Distribution of seismicity across strike-slip faults in California,
731 preprint.
732
733 Reasenberg P. (1985), 2nd-order moment of central california seismicity, 1969-1982, *J. Geophys.*
734 *Res.*, 90, 5479-5495.
735
736 Richards-Dinger K., R. S. Stein, and S. Toda, Test of the hypothesis that all aftershocks are
737 triggered by dynamic stress, preprint.
738
739 Shearer P., E. Hauksson, and G. Lin (2005), Southern california hypocenter relocation with
740 waveform cross-correlation, part 2: Results using source-specific station terms and cluster analysis,
741 *Bull. Seismol. Soc. Am.*, 95, 904-915.
742
743 Sornette D., M. J. Werner (2005), Apparent clustering and apparent background earthquakes biased

744 by undetected seismicity, *J. Geophys. Res.*, 110, B09303, doi:10.1029/2005JB003621.
745
746 Veen A., and Schoenberg F. P. (2008), Estimation of space-time branching process models in
747 seismology using and EM-type algorithm, *J. Am. Stat. Assoc.*, 103, 614-624.
748
749 Zhuang J., Y. Ogata, and D. Vere-Jones (2002), Stochastic declustering of space-time earthquake
750 occurrences, *J. Am. Stat. Assoc.*, 97, 458, 369- 380.
751
752 Zhuang J., Y. Ogata, and D. Vere-Jones (2004), Analyzing earthquake clustering features by using
753 stochastic reconstruction, *J. Geophys. Res.*, 109, B5, B05301.
754
755 Ziv A., and A. M. Rubin (2000), Static stress transfer and earthquake triggering: No lower threshold
756 in sight?, *J. Geophys. Res.*, 105, 13631-13642.
757
758 Ziv A. (2006), On the role of multiple interactions in remote aftershock triggering: The Landers and
759 the Hector Mine case studies, *Bull. Seismol. Soc. Am.*, 96, 80-89.

Annex A – The MISD algorithm

We recap in this annex how the MISD algorithm works. We analyze a set of N earthquakes, with occurrence times t_i, magnitudes m_i, and hypocenter positions $\underline{x}_i$. The relative distance r_{ij} between two earthquakes i and j needs not to be the hypocentral distance. It can for example be the distance of the 2nd earthquake hypocenter to the fault of the 1st earthquake. The total duration of the dataset is T, and it covers a volume V.

Discretization: prior to the analysis, a set of discrete intervals in time $[t_j, t_{j+1}]$ (duration $\delta t_j = t_{j+1} - t_j$), distance $[r_k, r_{k+1}]$ and magnitude $[m_i, m_{i+1}]$ must be defined. The distance bins are related to shell volumes δV_{ik}, see section 3.3. A trade-off must be found between too few intervals, hence too great a smoothing of the kernel values, and too many intervals, hence too few data points falling in these intervals, resulting in too large an error on the kernel estimates.

Algorithm:

[1] – define a probability NxN matrix $\underline{\underline{\omega}}$ such that $\omega_{ij} \geq \mathbf{0}$ for all $i<j$, $\omega_{ij} = \mathbf{0}$ for $i \geq j$, as well as a background probability vector $\omega_{\mathbf{0}j} \geq \mathbf{0}$ for all j, such that they sum up to 1: $\omega_{\mathbf{0}j} + \sum_{i<j} \omega_{ij} = \mathbf{1}$, for all j. The starting values can be arbitrary, as long as the choice is not cumbersome (e.g., filled with too many zeros). In this work, we chose $\omega_{ij}(i<j) = \omega_{\mathbf{0}j} = \frac{\mathbf{1}}{N + j - \mathbf{1}}$.

[2] – M – Step: compute $n_{ijk} = \sum_{a \in [m_i, m_{i+\mathbf{1}}]} \sum_{b>a} \omega_{ab} \theta(t_b - t_a \in [t_j, t_{j+\mathbf{1}}]) \theta(r_{ab} \in [r_k, r_{k+\mathbf{1}}])$,

$n_i = \sum_a \theta(m_a \in [m_i, m_{i+\mathbf{1}}])$ the number of earthquakes with magnitude in magnitude interval

$[m_i, m_{i+1}]$, and $n_0 = \sum_j \omega_{0j}$. Then, compute $\lambda_{ijk} = \frac{n_{ijk}}{n_i \delta t_j \delta V_{ik}}$ and $\lambda_0 = \frac{n_0}{TV}$ the kernel and background potential MLE given the probabilities ω_{ij} and ω_{0j}.

[3] – E – Step: recompute the probabilities using the kernel values just estimated in [2], with $\omega_{ij} = \frac{\lambda_i(r_{ij}, t_j)}{\lambda_0 + \sum_{k<j} \lambda_k(r_{kj}, t_k)}$ and $\omega_{0j} = \frac{\lambda_0}{\lambda_0 + \sum_{k<j} \lambda_k(r_{kj}, t_k)}$, and normalize them so that $\omega_{0j} + \sum_{i<j} \omega_{ij} = 1$ for all *j*. The kernel values $\lambda_i(r_{ij}, t_j)$ entering the above equations correspond to λ_{abc} of step [2], with *a* the magnitude interval corresponding to m_i, *b* the time interval corresponding to t_j-t_i, and *c* the distance interval corresponding to r_{ij}.

[4] – Iterate steps [2] and [3] until convergence is reached. A convergence criterion must be defined. In this study, we stop the algorithm when an iteration of steps [2] and [3] does not modify any of the logarithm of the kernel values by more than a given threshold, taken to be either 1% or 0.1% depending on the run.

In the present study, we specifically imposed the background rate-density λ_0, which is therefore not updated in step [2].

Annex B – Computing closest distances to the fault

The distance r_{ij} between earthquake *i* and earthquake *j* is taken to be the shortest distance from the fault plane of earthquake *i* to the hypocenter of earthquake *j*, or equal to the inter-hypocentral distance if the latter is shorter than the former. For all $m \geq 6$ earthquakes (8 earthquakes), the fault plane geometry is taken from rupture models deduced from inversion of seismic and geodetic data. All source models are provided by Martin Mai's finite source rupture model database

(www.seismo.ethz.ch/srcmod). We assumed the same fault plane for the Hector Mine mainshock and its biggest (m=6.7) recorded aftershock. For $4 \leq m < 6$ earthquakes, the fault plane is computed in two different ways depending on the existence of an available focal mechanism:

- We test whether a focal mechanism is available in the earthquake focal mechanisms database available at SCEC [*Hardebeck and Shearer*, 2003]. If the considered mainshock has a reported mechanism in the database, we first selected earthquakes occurring within 100 days following the mainshock and with hypocentral distances less than twice the length $d = 10^{0.5(m-4)}$ in km. The fault plane is then defined as the nodal plane passing through the mainshock and minimizing the distances, in the least-square sense, to all the selected earthquakes. Because rupture on the fault plane can be unilateral or bilateral, the center of the fault plane is chosen as the mean position of all the selected earthquakes, instead of the hypocentre of the mainshock. The size of the (square) fault plane is defined by its half-length d as defined above.
- If no mechanism exists for the considered mainshock we repeat the procedure described above but instead of testing only two planes (the two nodal planes), we test all possible planes discretized every 1° in dip and azimuth.

For all the other earthquakes (m<4), we simply used hypocentral distances as the fault plane length becomes small compared to the first distance bin used.

Annex C – Correcting transient changes in completness magnitude

It is well known that the completness magnitude m_c systematically goes up by several units immediately after a very large mainshock. This is due to the mainshock rupture itself, its coda waves, and the very high rate of aftershocks in the minutes to hours after the end of the coda, which all swamp the seismic signal and mask intermediate-size aftershocks which would normally have been detected by the network. In order to account and correct for this effect, which will mostly

affect the triggering kernel for large magnitude mainshocks and at short time scales, we adopt the approach of *Peng et al.* [2007]:

- a completness magnitude m_c and b-value b are estimated globally for the whole catalogue.
- Assuming that the b-value stays constant through time, the transients fluctuations $m_c(t)$ of m_c can be computed by $m_c(t) = \overline{m}(t) - \frac{\mathbf{1}}{b \ln \mathbf{10}}$, where $\overline{m}(t)$ is the mean magnitude computed over the N_e earthquakes closest to time t.
- An earthquake that occurs at time t thus « sees » the completness magnitude $m_c(t)$, and therefore counts as $n(t) = \mathbf{10}^{b[m_c(t) - m_c]}$ earthquakes instead of just 1. Practically, this means that the kernel MLE computed during the M-Step uses the revised « numbers »

$$n_{ijk} = \sum_{a \in [m_i, m_{i+\mathbf{1}}]} \sum_{b>a} \omega_{ab}\ n(t_b)\ \theta(t_b - t_a \in [t_j, t_{j+\mathbf{1}}])\theta(r_{ab} \in [r_k, r_{k+\mathbf{1}}]) \text{ and } n_{\mathbf{0}} = \sum_j \omega_{\mathbf{0}j}\ n(t_j).$$

This method unfortunately depends on the smoothing parameter N_e. In this study, we used N_e=10. Larger values (up to 100) of this parameter were tested, but resulted in a greater truncation of the kernel at large magnitude and short time scales.

	1 km	2 km	4 km	8 km	16 km	32 km	64 km	128 km	256 km	512 km
$2 \le m < 3$	5.22	31.5	233	1.67e+03	1.05e+04	4.82e+04	1.93e+05	7.73e+05	3.09e+06	1.24e+07
$3 \le m < 4$	9.16	40.8	263	1.74e+03	1.05e+04	4.84e+04	1.94e+05	7.74e+05	3.09e+06	1.24e+07
$4 \le m < 5$	38.1	93.4	429	2.25e+03	1.17e+04	5.01e+04	1.97e+05	7.80e+05	3.10e+06	1.24e+07
$5 \le m < 6$	225	337	1.06e+03	3.87e+03	1.48e+04	5.46e+04	2.06e+05	7.97e+05	3.14e+06	1.24e+07
$26 \le m < 7$	7.71e+02	1.49e+03	3.13e+03	8.04e+03	2.31e+04	7.03e+04	2.37e+05	8.60e+05	3.26e+06	1.27e+07
$m \ge 7$	1.49e+03	2.84e+03	5.41e+03	1.25e+04	3.13e+04	8.61e+04	2.70e+05	9.27e+05	3.40e+06	1.27e+07

Table 1: volumes δV_{ik} (in km^3) of the shells $[r_k, r_{k+1}]$, function of the magnitude m and outer radius r_{k+1}. The first shell is from 0 to 1 km, the second from 1 to 2 km, and so on.

	1 km	2 km	4 km	8 km	16 km	32 km	64 km	128 km	256 km	512 km
$2 \le m < 3$	17.6	45.8	161	493	1.04e+03	2.93e+03	7.75e+03	7.85e+03	7.80e+03	7.91e+03
$3 \le m < 4$	16.6	46.2	163	503	1.08e+03	2.85e+03	7.35e+03	1.30e+04	3.11e+04	1.59e+04
$4 \le m < 5$	64.9	84.6	221	608	1.20e+03	3.30e+03	8.18e+03	1.31e+04	3.15e+04	1.35e+04
$5 \le m < 6$	144	141	312	645	1.32e+03	2.96e+03	7.54e+03	1.08e+04	3.46e+04	1.22e+04
$26 \le m < 7$	243	383	466	631	816	4.30e+03	7.85e+03	1.23e+04	2.78e+04	8.25e+03
$m \ge 7$	155	107	169	237	900	7.36e+03	8.53e+03	9.69e+03	2.19e+04	1.81e+03

Table 2: values δV_{ik} (in number of unconditioned earthquakes) of the shells$[r_k, r_{k+1}]$, function of the magnitude m and outer radius r_{k+1}.

Aftershocks within 0 – 15 minutes			Aftershocks within 12 – 24 hours		
L (km)	T_1 (days)	T_2 (days)	L (km)	T_1 (days)	T_2 (days)
100	3	0.5	100	30	2
200	6	1	200	60	4
50	1.5	0.25	50	15	1
200	60	1	200	200	6
500	6	1	500	30	2

Table 3: parameters used for selecting mainshocks, following the treatment of Felzer and Brodksy [2006].

FIGURE CAPTIONS:

Figure 1: example of the relations between 5 earthquakes A, B, C, D and E occurring in that order. The numbers attached to the arrows refer to the probabilities ω_{ij} that the 2nd earthquake (j) is a direct aftershock of the 1st (i). Note that, for all j except A, $\sum_{i=1}^{j-1} \omega_{ij} = \mathbf{1}$, i.e., there is a 100% probability that its mainshock is listed in this sequence (it is either A, or B, ...).

Figure 2: values of λ_1 (top curves) and λ_2 (bottom curves) for the example of Figure 1. The three sets of curves correspond to different starting values, as indicated on the graph. All starting values converge after a few iterations to the same solution.

Figure 3: cumulative shell volumes δV, for the two sets as explained in the text. Only the two most extreme magnitude bins for the initial earthquake are shown: $2 \leq m < 3$ (triangles) and $m \geq 7$ (squares). Dashed lines: Euclidean geometrical volumes. Continuous lines: counting the number of uncorrelated earthquakes in the shell.

Figure 4: (top) time series of the direct, and of all (direct and indirect) $m \geq 3$ aftershocks of Landers, which occurs on day 3101 (vertical black line). (Center) map showing the probability $\omega_{L,j}$ of being a direct aftershock of Landers. (Bottom) same as center graph, but for the probability $\omega_{L,j} + \omega_{L,j}'$ of being conditionned on Landers. The color corresponds to the probability, as given by the color scale bar.

Figure 5: linear density for $3 \leq m < 4$ mainshocks and $m \geq 2$ aftershocks, for two time intervals as indicated on the graphs. Thick crosses: our method. Thin circles: using the approach modified from *Felzer and Brodsky* [2006]. The plotted values and the error bars are the means and standard deviations of the logarithm of the linear density values. We offset the circles for visual purposes.

Figure 6: same as in Figure 5, but for $5 \leq m < 6$ mainshocks. The error bar for the first distance interval (0 – 1 km) is under estimated: the distance to a $5 \leq m < 6$ rupture is likely to be badly resolved in this interval.

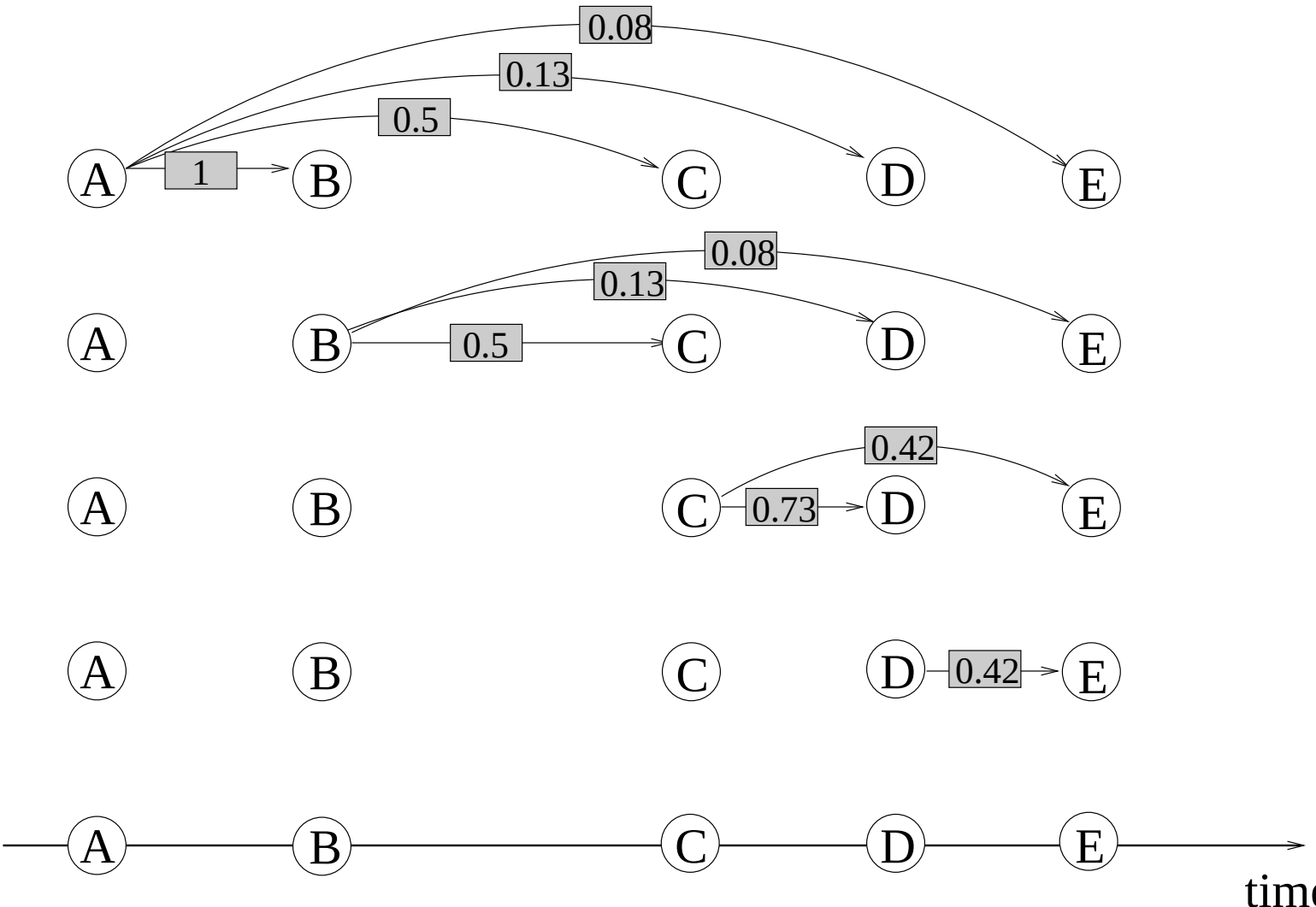
Figure 7: linear densities of the direct aftershocks of the Landers earthquake and all earthquakes in the first 100 days after Landers. The dashed line gives a power-law exponent of -1.66, when fitted in the 1 – 100 km range.

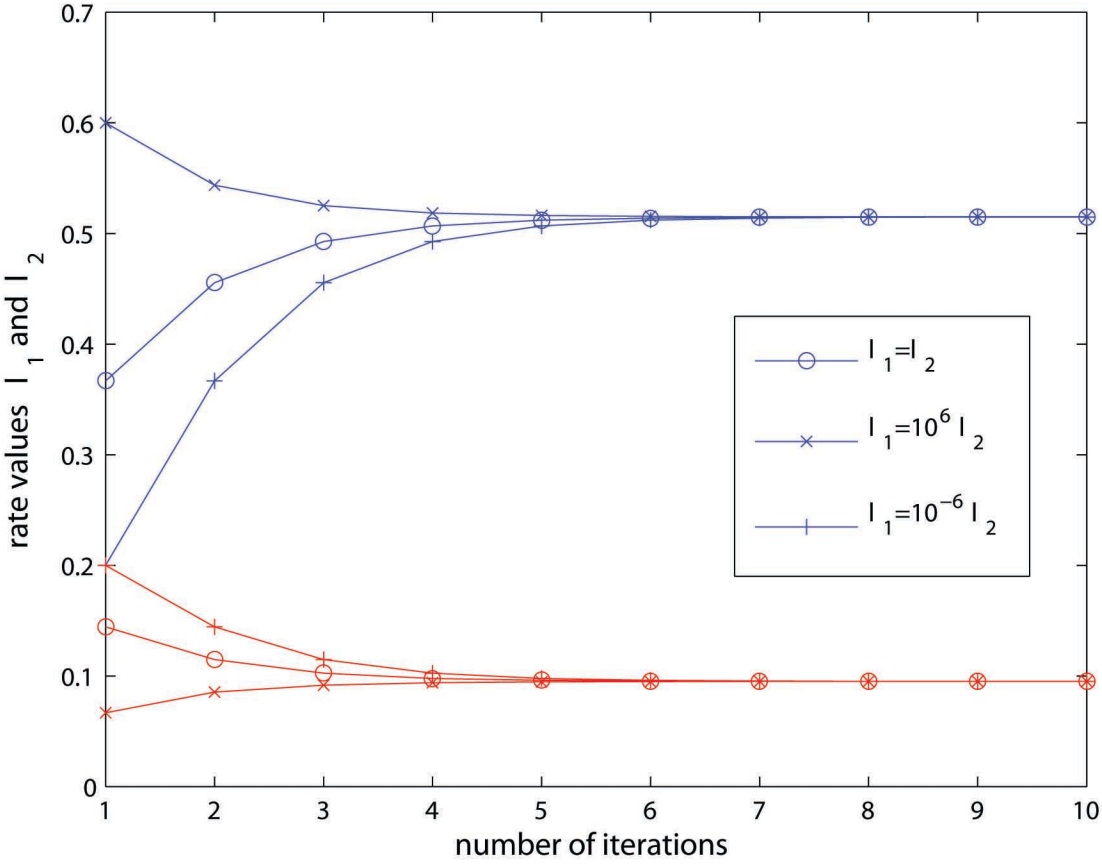
Figure 8: mean and standard deviation of the Coulomb static stress change distribution for shell volumes $R_1 < r < R_2$ away from a $m=3$ mainshock. The vertical lines show the $[R_1, R_2]$ intervals. We distinguish between the positive (□) and the negative (▽) mean stress changes.

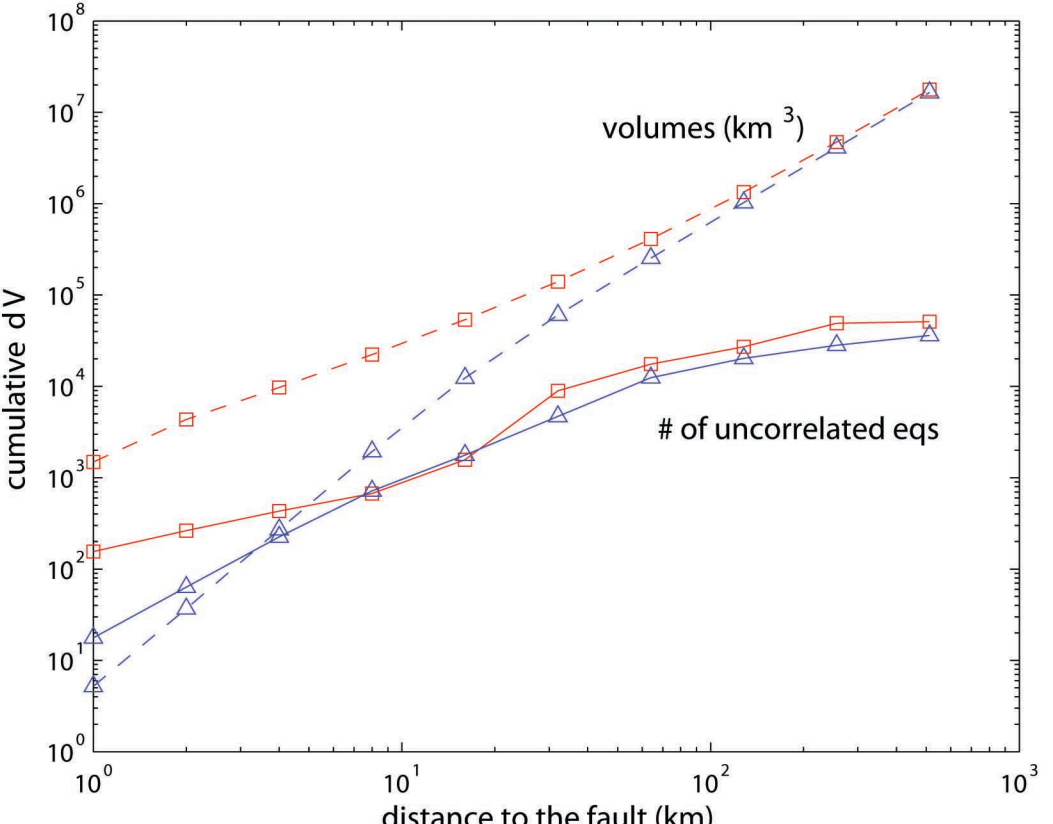
Figure 9: aftershock linear density following a $m=3$ mainshock, inferred from the stress distribution of Figure 7, using the rate-and-state friction model and a time interval of about 1 hour after the mainshock. The density is defined up to an arbitrary multiplicative factor. Three values for the $A\sigma$ parameter were tested. The straight line shows a $r^{-2.2}$ decay.

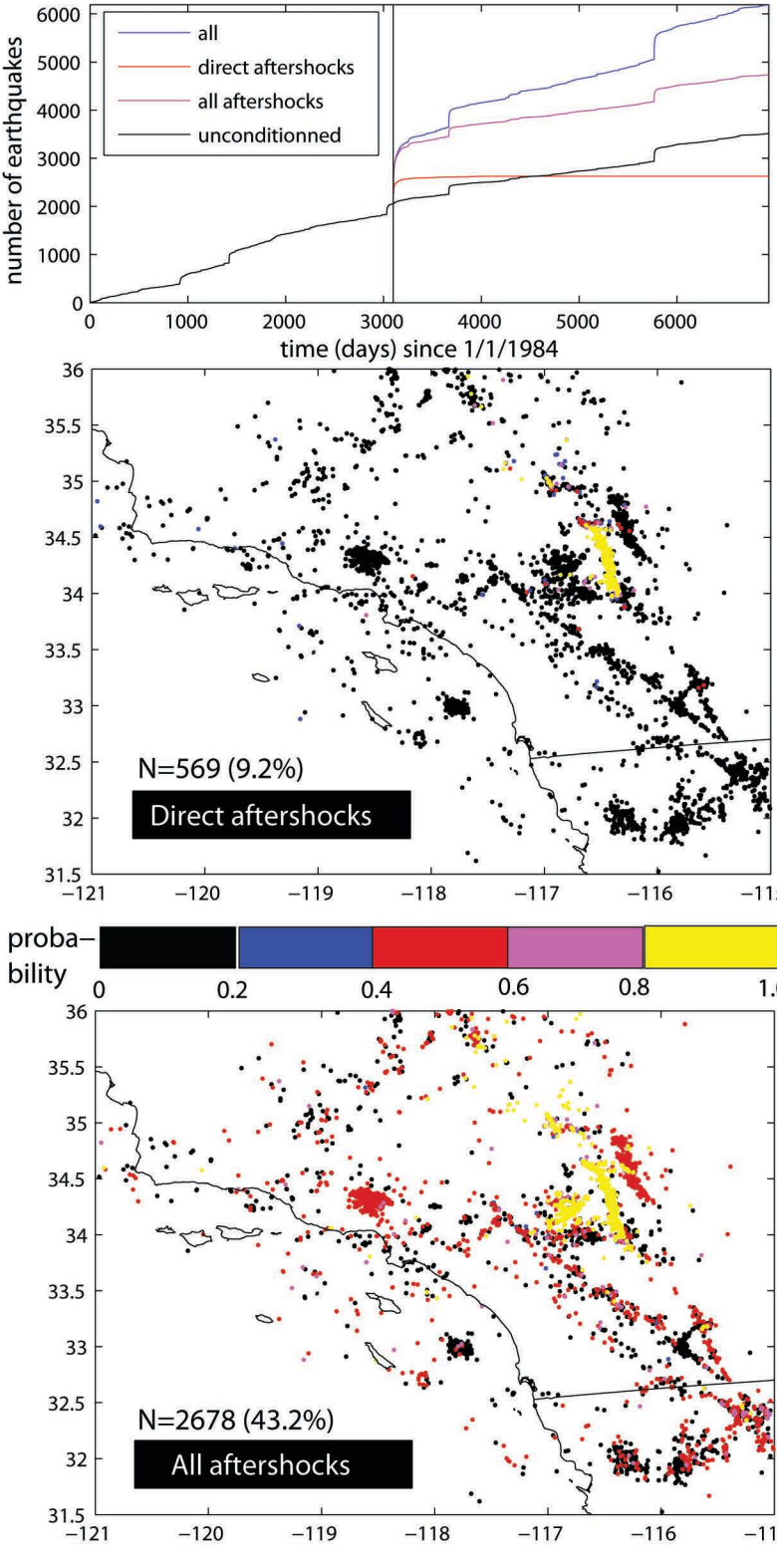
852

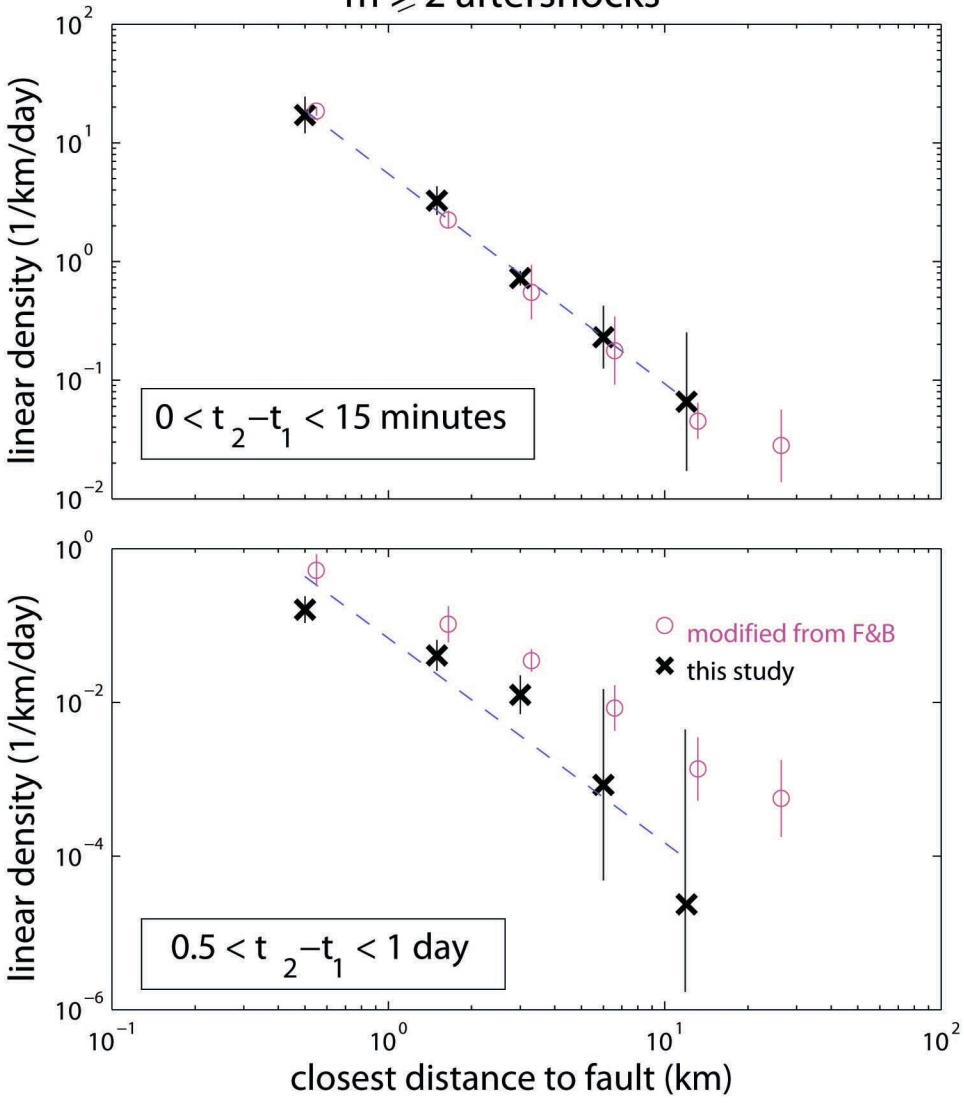
0.08 : probability of triggering

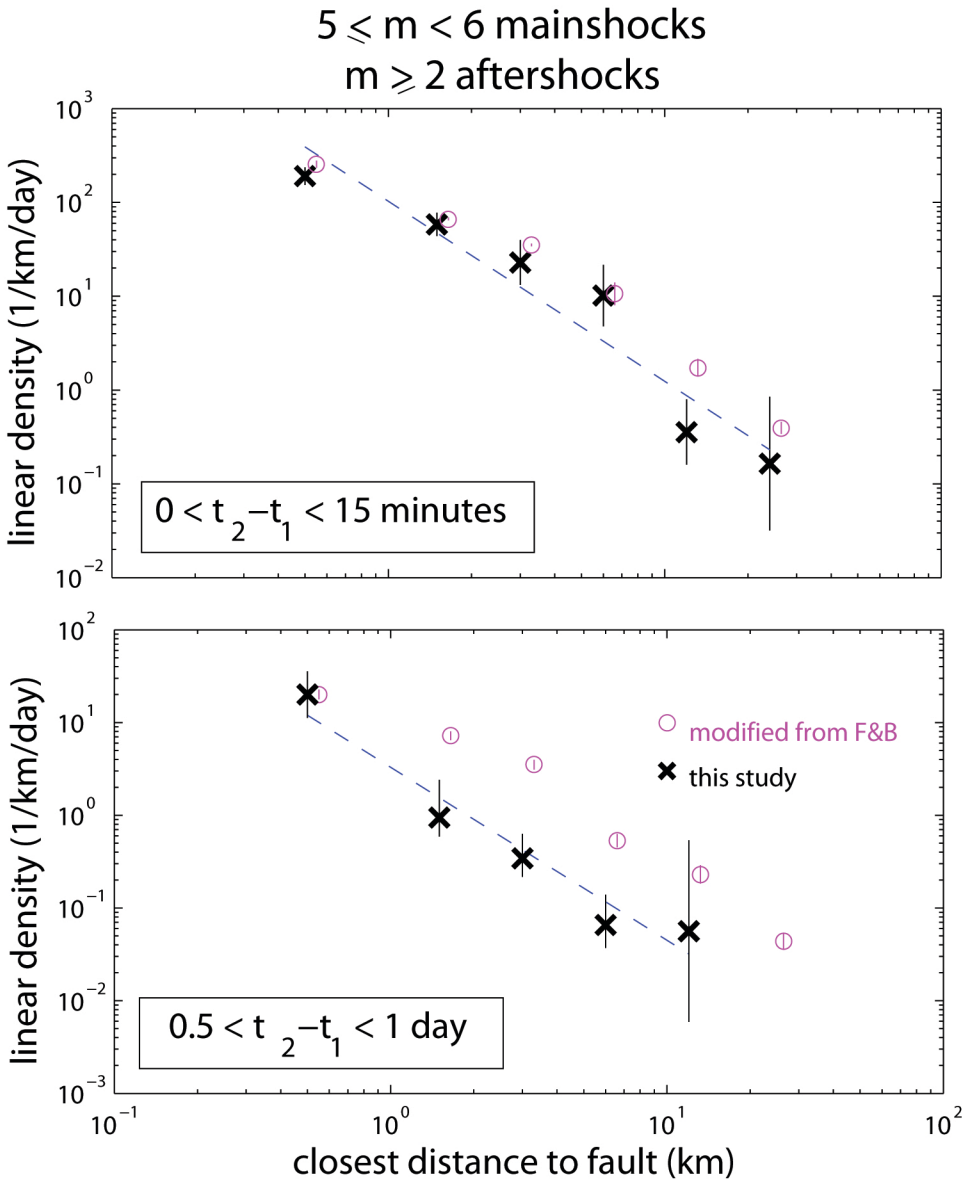


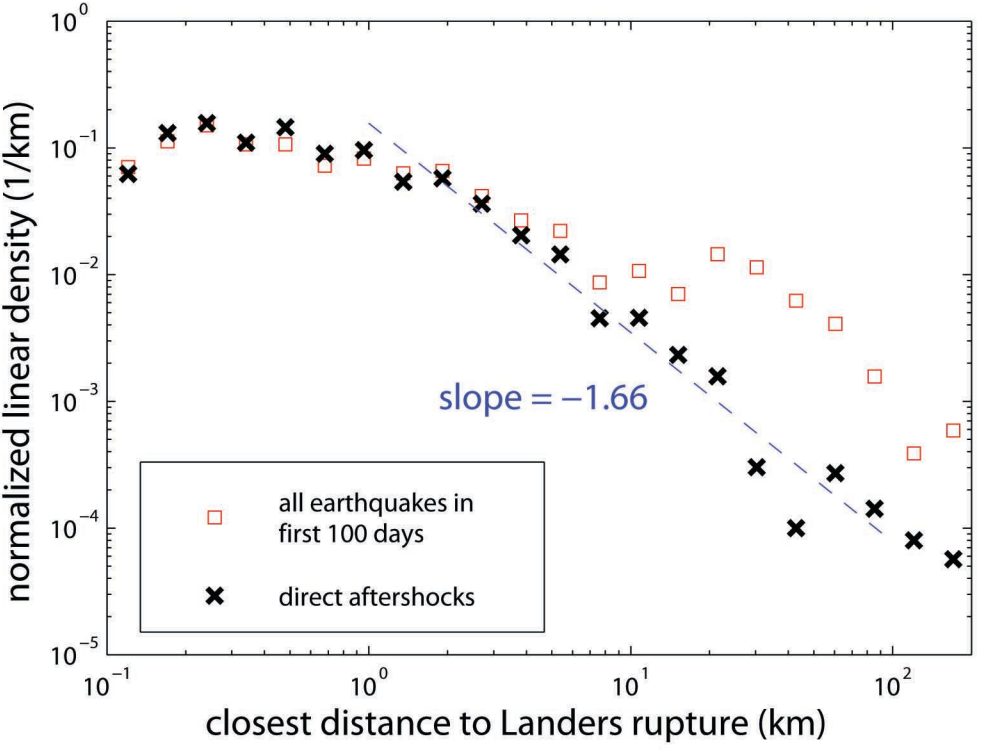

0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
rate values I 1 and I 2
1
2
3
4
5
6
7
8
9
10
number of iterations
I 1=I 2
I 1=10^6 I 2
I 1=10^−6 I 2

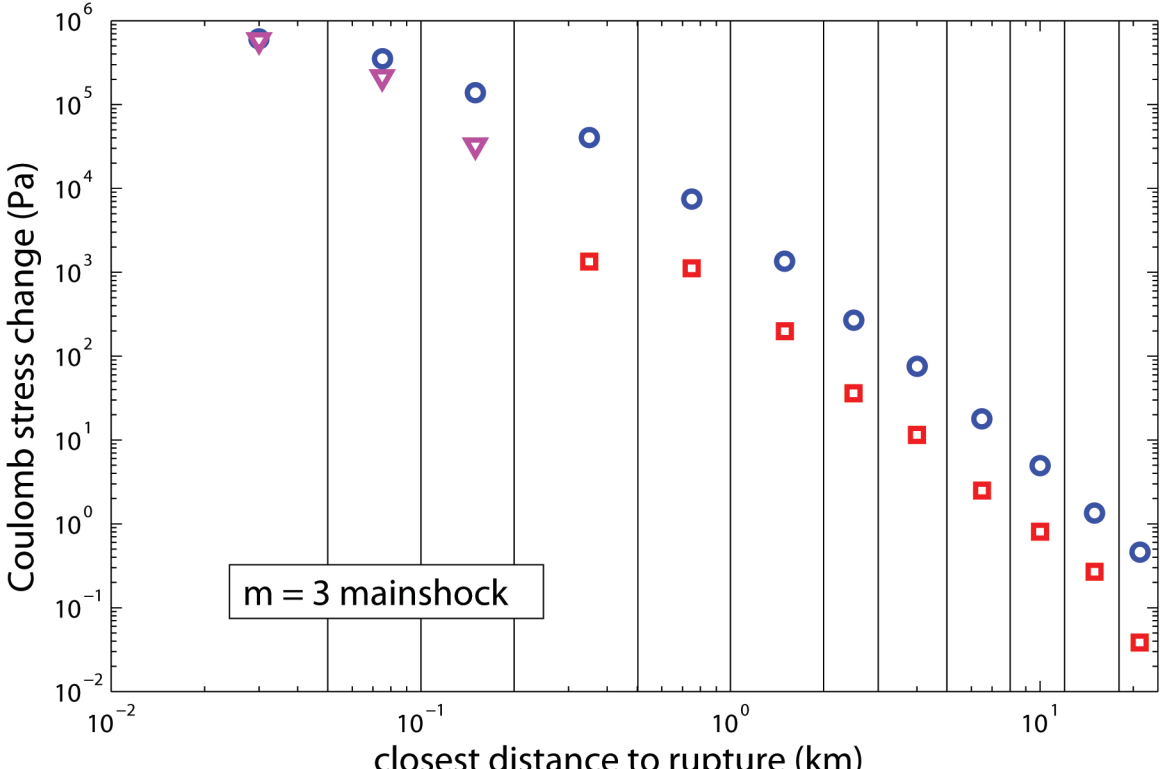

volumes (km 3)
# of uncorrelated eqs
cumulative dV
distance to the fault (km)
10^0
10^1
10^2
10^3
10^4
10^5
10^6
10^7
10^8

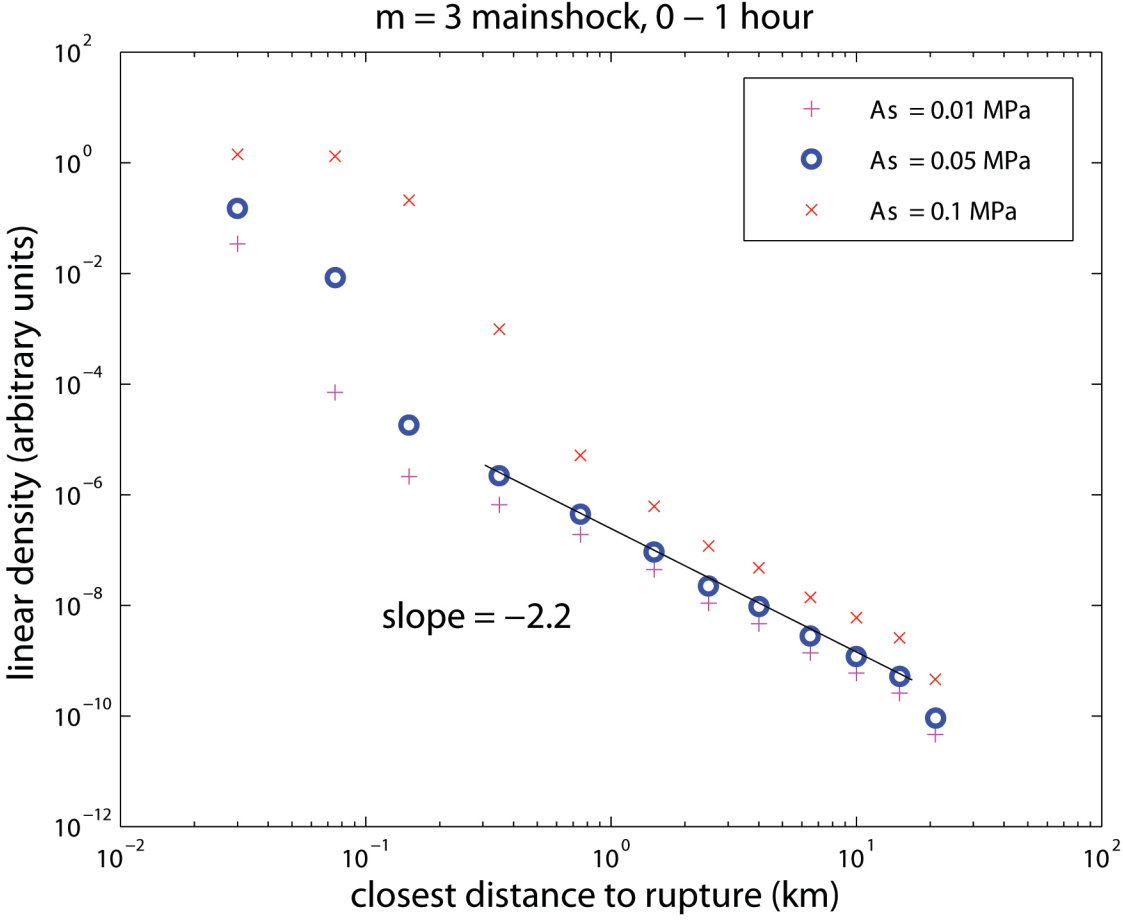

all
direct aftershocks
all aftershocks
unconditionned
number of earthquakes
time (days) since 1/1/1984
N=569 (9.2%)
Direct aftershocks
proba–
bility
N=2678 (43.2%)
All aftershocks


3 ≤ m < 4 mainshocks
m ≥ 2 aftershocks
linear density (1/km/day)
0 < t 2 −t 1 < 15 minutes
linear density (1/km/day)
modified from F&B
this study
0.5 < t 2 −t 1 < 1 day
closest distance to fault (km)


5 ≤ m < 6 mainshocks
m ≥ 2 aftershocks
linear density (1/km/day)
0 < t 2 −t 1 < 15 minutes
linear density (1/km/day)
0.5 < t 2 −t 1 < 1 day
modified from F&B
this study
closest distance to fault (km)


normalized linear density (1/km)
closest distance to Landers rupture (km)
slope = −1.66
all earthquakes in first 100 days
direct aftershocks
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-1}
10^{0}
10^{1}
10^{2}


Coulomb stress change (Pa)
closest distance to rupture (km)
m = 3 mainshock
10^{6}
10^{5}
10^{4}
10^{3}
10^{2}
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-2}
10^{-1}
10^{0}
10^{1}


m = 3 mainshock, 0 – 1 hour
As = 0.01 MPa
As = 0.05 MPa
As = 0.1 MPa
slope = −2.2
linear density (arbitrary units)
closest distance to rupture (km)
10^2
10^0
10^−2
10^−4
10^−6
10^−8
10^−10
10^−12
10^−2
10^−1
10^0
10^1
10^2