12. P. Stallknecht, H. Kohl, *Ultramicroscopy* **66**, 261 (1996).
13. E. C. Cosgriff, M. P. Oxley, L. J. Allen, S. J. Pennycook, *Ultramicroscopy* **102**, 317 (2005).
14. H. Kohl, H. Rose, *Adv. Imaging. Electron Phys.* **65**, 173 (1985).
15. D. A. Muller, J. Silcox, *Ultramicroscopy* **59**, 195 (1995).
16. R. F. Egerton, *Electron Energy Loss Spectroscopy in the Electron Microscope* (Plenum, New York, ed. 2, 1996).
17. P. Trebbia, N. Bonnet, *Ultramicroscopy* **34**, 165 (1990).
18. J. G. Chen, *Surf. Sci. Rep.* **30**, 1 (1997).
19. D. A. Muller, N. Nakagawa, A. Ohtomo, J. L. Grazul, H. Y. Hwang, *Nature* **430**, 657 (2004).
20. K. Dorr *et al.*, *J. Appl. Phys.* **89**, 6973 (2001).
21. X. Hong, A. Posadas, C. H. Ahn, *Appl. Phys. Lett.* **86**, 142501 (2005).
22. M. Izumi *et al.*, *Phys. Rev. B* **64**, 064429 (2001).
23. The instrument was developed under NSF grant DMR-9977547 with an additional contribution from Cornell University. In-kind contributions were made by Gatan Inc., Sandia National Laboratory, Pacific Northwest Laboratory, and Bell Laboratories. Supported by NSF Nanoscale Science and Engineering Initiative grant EEC-0117770 (D.A.M. and J.S.), NSF Materials Research Science and Engineering Center grant DMR-0520404 (L.F.K.), the TEPCO Research Foundation (H.Y.H.), and the Japan Society for the Promotion of Science (J.H.S.).

# Extending Earthquakes' Reach Through Cascading

David Marsan* and Olivier Lengliné

Earthquakes, whatever their size, can trigger other earthquakes. Mainshocks cause aftershocks to occur, which in turn activate their own local aftershock sequences, resulting in a cascade of triggering that extends the reach of the initial mainshock. A long-lasting difficulty is to determine which earthquakes are connected, either directly or indirectly. Here we show that this causal structure can be found probabilistically, with no a priori model nor parameterization. Large regional earthquakes are found to have a short direct influence in comparison to the overall aftershock sequence duration. Relative to these large mainshocks, small earthquakes collectively have a greater effect on triggering. Hence, cascade triggering is a key component in earthquake interactions.

E arthquakes of all sizes, including after-shocks, are able to trigger their own aftershocks. The cascade of earthquake triggering causes the seismicity to develop complex, scale-invariant patterns. The causality of "mainshock A triggered aftershock B," which appears so obvious if mainshock A happens to be large, must then be modified into a more subtle "mainshock A triggered C1, which triggered C2, ..., which triggered B." This has paramount consequences: The physical mechanism that causes direct triggering (static or dynamic stress changes, fluid flow, afterslip, etc.) cannot be studied by looking at aftershocks that were not directly triggered by the mainshock. Moreover, if indirect triggering is important in the overall aftershock budget (*1–3*), then direct triggering must be confined to spatial ranges and times shorter than the size of the total

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

*To whom correspondence should be addressed. E-mail: david.marsan@univ-savoie.fr

**Fig. 1.** Estimated rates and densities for California. (**A** and **B**) Bare kernels; (**C** and **D**) dressed kernels. The best power laws for the temporal rates $\lambda_t(t, m)$ and the best $[1 + (r/L)]^{-3}$ laws for the densities $\lambda_s(x, y, m)$ are shown as black dashed lines. The background temporal rate $\lambda_{0,t}$ [black horizontal line in (A) and (C)] is computed as $\sum_{i=1}^{N} w_{0,i}/T$. In (C) and (D), the dressed kernels (continuous lines) are compared to the bare ones (color dashed lines). The densities $\lambda_s$ have been vertically shifted for clarity.
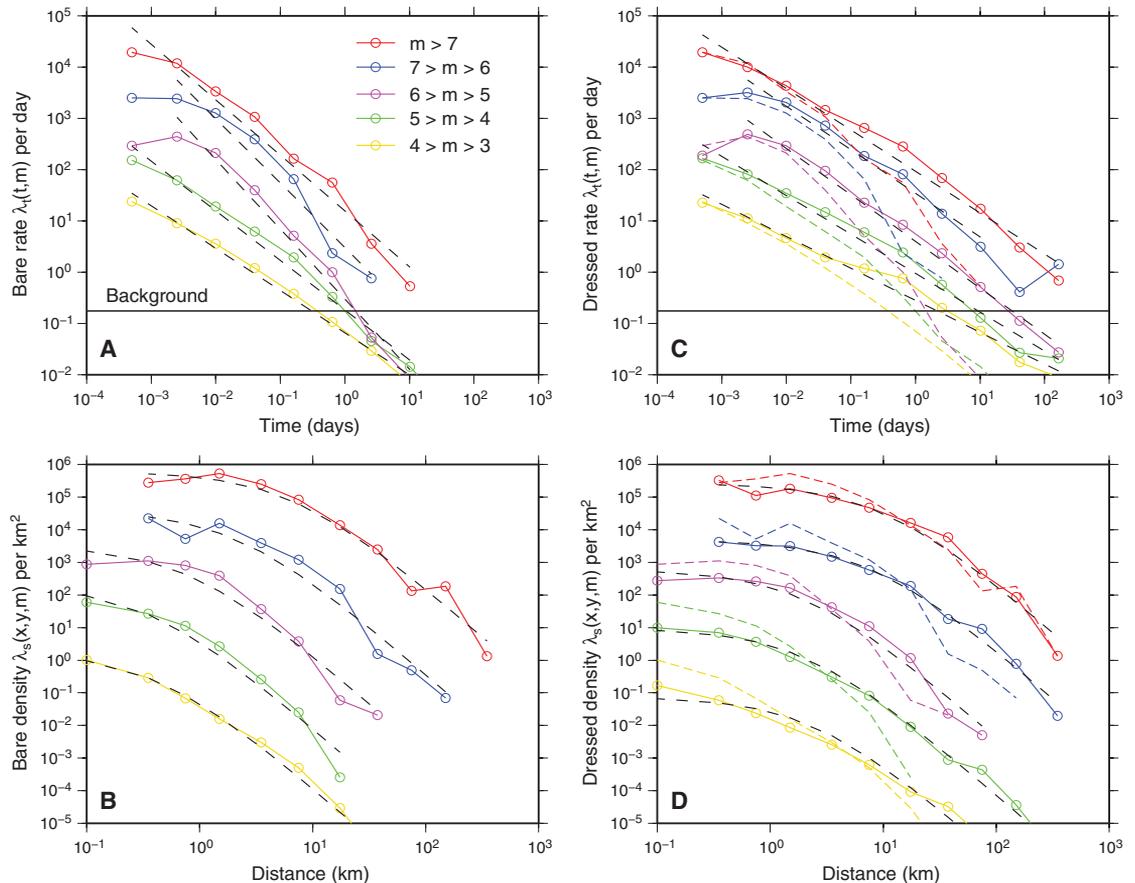
aftershock sequence. How much shorter is still an open question, in the absence of any simple or standard way to efficiently discriminate direct and indirect triggering in the data.

Seismologists mostly resort to declustering algorithms to separate earthquakes between mainshocks and aftershocks (4–7). These methods have arbitrary rules and are heavily parameter-dependent. Recently, more sophisticated methods were proposed to perform stochastic declustering—that is, determining the probability that earthquake A triggered earthquake B (8). This overseeds the usual approach of binary linking of one aftershock to one single mainshock: An earthquake $j$ is then influenced by all preceding earthquakes $i$, according to influence weights $w_{i,j}$. These methods, however, are model-dependent, as the influence of a trigger earthquake is constrained to follow a specific law, whose parameters must be inverted.

Here we show that the probability of directly and indirectly triggering aftershocks can be estimated with no a priori model. A rapidly converging algorithm with a small number of hypotheses (linearity, mean-field) can decipher the complex seismicity time series to reveal the underlying triggering influences exerted by earthquakes of all sizes. A notable result is that even large earthquakes causally trigger aftershocks only during a relatively short time span. However, they condition regional seismicity for a much longer time period, and over larger, time-increasing distances, through the local triggering caused by their aftershocks. This cascading effect, dominated by small shocks, thus appears to be a crucial component in earthquake interactions.

Seismicity is considered as a point process in time, space, and magnitude. The observed (dressed) seismicity rate density $\lambda(x, t)$, defined as the number of earthquakes per unit time and unit area at position $x$ and time $t$, is modeled as

$$\lambda(x, t) = \lambda_0 + \sum_{t_i < t} \lambda_i(x, t) \qquad (1)$$

where $\lambda_0$ is the uniform background rate density, and $\lambda_i(x, t)$ is the (bare) contribution of earthquake $i$ that occurred at $\{x_i, t_i\}$, representing the aftershocks directly caused by this earthquake. We assume only that (i) the triggering process is linear [i.e., the bare contributions $\lambda_i(x, t)$ sum up], and (ii) a mean-field response to the occurrence of an earthquake can be estimated that depends only on its magnitude, $\lambda_i(x, t) = \lambda(|x - x_i|, t - t_i, m_i)$, hence two earthquakes of equal magnitude are modeled similarly.

The algorithm works as follows:

1. Knowing an a priori bare kernel $\lambda(|\Delta x|, \Delta t, m)$ and $\lambda_0$, we compute the triggering weights $w_{i,j} = \alpha_j \lambda(|x_j - x_i|, t_j - t_i, m_i)$ if $t_i < t_j$, $w_{i,j} = 0$ otherwise, and the background weights $w_{0,j} = \alpha_j \lambda_0$. The normalization coefficients $\alpha_j$ are such that

$$\sum_{i=1}^{j-1} w_{i,j} + w_{0,j} = 1 \qquad (2)$$

2. The updated bare rates are then computed as

$$\lambda(|\Delta x|, \Delta t, m)$$
$$= \frac{1}{N_m \times \delta t \times S(|\Delta x|, \delta r)} \sum_{i,j \in A} w_{i,j} \qquad (3)$$

where $A$ is the set of pairs such that $|x_i - x_j| = |\Delta x| \pm \delta r$, $m_i = m \pm \delta m$, and $t_j - t_i = t \pm \delta t$ ($\delta r$, $\delta t$, and $\delta m$ are discretization parameters), $N_m$ is the number of earthquakes such that $m_i = m \pm \delta m$, and $S(|\Delta x|, \delta r)$ is the surface covered by the disk with radii $|\Delta x| \pm \delta r$. The a posteriori background rate is

$$\lambda_0 = \frac{1}{T \times S} \sum_{j=1}^{N} w_{0,j} \qquad (4)$$

where $T$ is the duration of the time series (containing $N$ earthquakes) and $S$ is the surface analyzed. This corresponds to stacking all the aftershocks following mainshocks $i$ of similar magnitudes, but counting an aftershock $j$ according to its weight $w_{i,j}$.

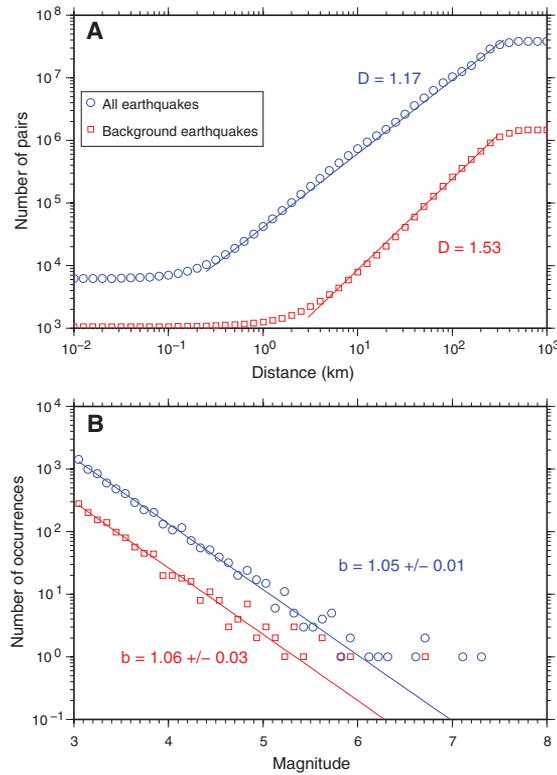Starting with an initial guess for $\lambda(|\Delta x|, \Delta t, m)$, these two steps are iterated until
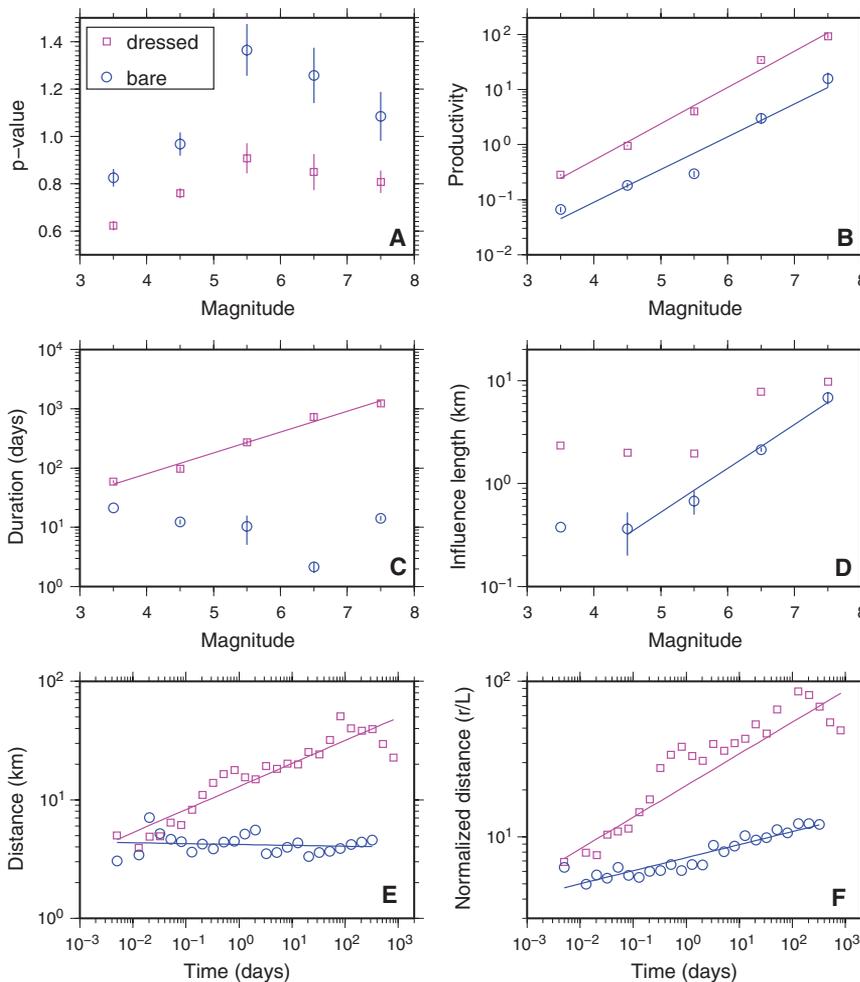


**Fig. 2.** Characteristics of bare (blue) and dressed (purple) aftershock sequences. (**A** and **B**) $p$ values and productivity parameters $\chi$ from the fits $\lambda_f(t, m) = \chi t^{-p}$ of Fig. 1. Productivity grows as $10^{0.60 \pm 0.07 m}$ (bare) and $10^{0.66 \pm 0.04 m}$ (dressed). (**C**) Durations of the sequences. The dressed duration follows $10^{0.35 m}$. (**D**) Influence lengths $L$ estimated from the $[1 + (r/L)]^{-3}$ fits of the spatial densities. (**E**) Mean epicentral distance between mainshocks and aftershocks versus time following the mainshock. (**F**) Same as (E), for epicentral distance normalized by the bare influence length $L$ of the mainshock. The best power laws $r \sim \delta t^H$ give (E) H = −0.01 ± 0.03 (bare), H = 0.19 ± 0.04 (dressed), and (F) H = 0.08 ± 0.02 (bare), H = 0.21 ± 0.06 (dressed).
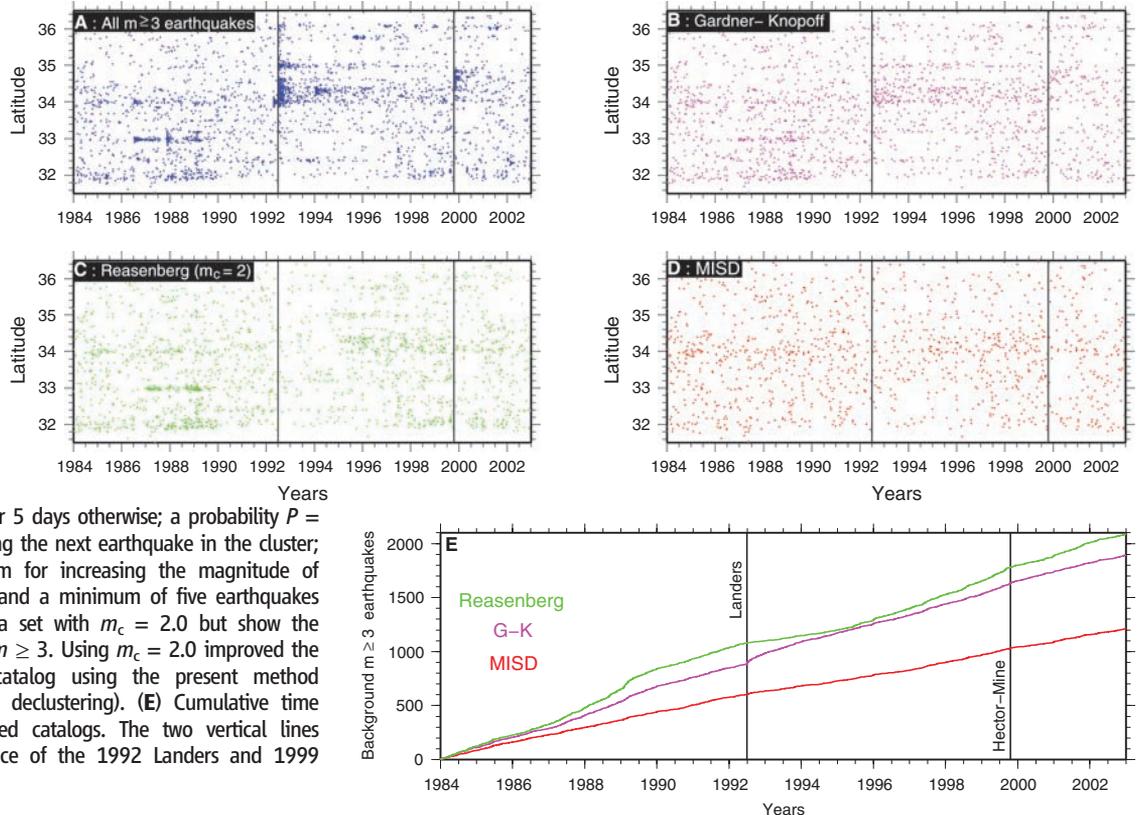
convergence is reached—that is, until the weights (or the rates) do not change substantially during an iteration. Tests were conducted on synthetic catalogs, showing the ability of the algorithm to correctly estimate the cascade structure (9).

We analyzed seismicity in southern California (10) from 1 January 1984 to 31 December 2002. We considered only the $N = 6190$ $m \geq 3$ earthquakes in the catalog, both for completeness reasons and because the method uses a $N \times N$ weight matrix, preventing the analysis of too large a data set on a standard desktop computer. The completeness magnitude is estimated to be 2.2 for this catalog. The method was slightly modified to account for the fact that the background earthquakes are nonuniformly distributed in space (9).

The rates had roughly an Omori-Utsu decay $\lambda_r(t, m) = \chi t^{-p}$ (Fig. 1), where $\chi$ is the productivity. The $p$ value increased with $m$ (Fig. 2A), as observed previously (11) for dressed aftershock sequences, although with a saturation at $m \geq 6$. The rates decayed significantly more slowly when considering the full cascade including indirectly triggered aftershocks. The dressed $p$ values were 0.2 to 0.4 units smaller than the bare $p$ values (Fig. 2A). The productivity parameter grew as $= \chi \sim 10^{am}$, with $a = 0.60 \pm 0.07$ and $a = 0.66 \pm 0.04$ for the bare and dressed kernels, respectively (Fig. 2B). This yielded a significantly lower scaling exponent than previous estimates (3, 12–14) for dressed sequences.

The densities were well fitted with a $\lambda_s(x, y, m) \sim [1 + (r/L)]^{-3}$ law, hence a $r^{-2}$ decay of the number of aftershocks (Fig. 1). The bare

**Fig. 3.** Comparison of total (blue) and background (red) seismicity. (**A**) Number of pairs $N(r)$ with epicentral distance less than distance $r$, along with estimates of the fractal correlation dimension $D$ such that $N(r) \sim r^D$. (**B**) Number of occurrences versus magnitude, scaling as $10^{-b \times m}$.



**Fig. 4.** Comparison of declustering algorithms. (**A** to **D**) Latitude versus time of occurrence of $m \geq 3$ earthquakes. (A) All earthquakes in the catalog. (B) Declustered catalog using the method of (4). (C) Declustered catalog using the algorithm of (5). Parameters are $r_{fact} = 8$, which gives the number of mainshock rupture lengths up to which a mainshock is supposed to trigger aftershocks; the maximum ahead-time for linking an aftershock to a cluster, which is 1 day if only one event belongs to the cluster, or 5 days otherwise; a probability $P = 95\%$ of confidence of observing the next earthquake in the cluster; $\chi_k = 0.5$, the correction term for increasing the magnitude of completeness during clusters; and a minimum of five earthquakes per cluster. We used the data set with $m_c = 2.0$ but show the declustered earthquakes with $m \geq 3$. Using $m_c = 2.0$ improved the algorithm. (D) Declustered catalog using the present method (model-independent stochastic declustering). (**E**) Cumulative time series of the three declustered catalogs. The two vertical lines indicate the time of occurrence of the 1992 Landers and 1999 Hector Mine earthquakes.

influence length $L$ was remarkably small, ranging from 0.35 km at $4 \leq m < 5$ to 6.6 km at $m \geq 7$. It grew as $10^{0.43m}$ (Fig. 2D), which is close to the $10^{0.5m}$ dependence expected for the rupture length of small to intermediate-size earthquakes (15). The bare length was 0.35 km for $3 \leq m < 5$, which is due to the limited resolution on the relative hypocenter positions, as confirmed by the break in scaling at ~400 m for the correlation integral (Fig. 3A). The dressed influence lengths were about 5 times the bare ones. These influence lengths were not maximum triggering distances: Many aftershocks were triggered past $L$. Relaxing the point-like earthquake hypothesis affected these results for the large mainshocks. Using distances to the rupture plane rather than epicentral distances, the bare kernel of $m \geq 6$ mainshocks was moved toward longer-range triggering, greater aftershock productivity, and longer durations (9).

The durations of aftershock sequences (Fig. 2C) were computed by (i) calculating, for all the mainshocks, the delays after which the last direct and last indirect aftershocks occurred, and (ii) averaging these delays conditioned on the magnitude of the mainshock. The duration of direct aftershock sequence was largely independent of the mainshock magnitude and was generally short (on the order of 10 to 15 days for $m \geq 3$ aftershocks). The dressed sequence, however, lasted longer for larger mainshocks, following a $10^{0.35m}$ increase. This implies that short-lasting triggering mechanisms, acting at the time scale of a few days, could be the key process, along with the cascading effect, in controlling earthquake dynamics.

The slow expansion of aftershock zones has been reported in previous studies (16, 17). We measured the mean distance $r$ between mainshock and aftershock with time $\delta t$ separating the two earthquakes. This distance was constant for bare aftershocks, whereas for dressed aftershocks it slowly grew as $r \sim \delta t^{0.19}$ (Fig. 2E). We reached similar conclusions when considering the distance normalized by the bare influence length $L$ of the mainshock (Fig. 2F). This shows that cascading triggering drives the expansion of aftershock zones: The spatial pattern of direct triggering was almost constant with time, ruling out triggering by fluid movements or viscoelasticity at the time scales examined here.

The number of earthquakes directly triggered by all the earthquakes of a given magnitude slowly decreased with this magnitude, demonstrating the importance of small shocks in controlling the regional seismicity. The collective production scales as $10^{(a-b)m}$; here, $a = 0.6$ for the bare aftershocks (Fig. 2B) and $b = 1.05$ (Fig. 3B).

The background spatial function $g(r)$ decayed as $r^{-0.43}$, which is equivalent to saying that the background earthquake epicenters are fractally distributed with dimension $D = 1.57$. This is confirmed by the correlation integral of the background earthquakes (Fig. 3A), counting the number of pairs of earthquakes with distance less than a given value, each pair $(i, j)$ being weighted by $w_{0, i} \times w_{0, j}$. The similar $b$ values of the Gutenberg-Richter laws (18) for the total and the background earthquakes (Fig. 3B) suggest that the dynamic rupture extent is not a priori controlled by the triggering mechanism (either previous earthquakes or aseismic processes such as tectonic loading) at work.

We obtained a background rate of 0.17 $m \geq 3$ earthquakes per day in southern California over the years 1984 to 2002, which corresponds to 19.5% of the total rate of $m \geq 3$ earthquakes. The remaining ~80% can therefore be considered as resulting from stress transfer and fault interaction processes, causing the seismicity to be heavily clustered in time. Such a proportion at the magnitude cutoff $m_c = 3$ is coherent with the estimates in the range 18% to 24% we obtained using other nonparametric methods (19). The values computed using the present method, however, depend on $m_c$: Cutting at larger magnitudes increases the relative proportion of background earthquakes to 32% at $m_c = 4$ and 68% at $m_c = 5$. A larger cutoff causes the removal of small triggering earthquakes; earthquakes that were triggered by small shocks are then more likely to be seen as background earthquakes (20). Inversely, decreasing $m_c$ would yield smaller percentages of background events. The 19.5% proportion at $m_c = 3$ is therefore an overestimation of the actual background contribution.

Declustering of earthquake catalogs aims at removing the aftershock clusters, keeping only statistically independent mainshocks. The declustering algorithms by Gardner and Knopoff (4) and Reasenberg (5) are the most classical methods. We ran these two methods along with ours (Fig. 4). The present method is better at identifying an underlying Poisson process, and it efficiently removes the aftershock clusters following large mainshocks such as the 1992 Landers and 1999 Hector Mine earthquakes. Moreover, it does not rely on any parameterization.

Cascading of aftershock triggering is an essential component of seismicity. It has a scale-invariant structure, making earthquake declustering an ill-defined problem: The statistical dependence between earthquakes increases when decreasing the value of $m_c$, so that the remaining set of declustered mainshocks heavily depends on $m_c$. Because of

this cascading, the aftershock sequence initiated by a mainshock is substantially extended, mostly in time. Conversely, what appears at first as an aftershock cluster related to a well-identified mainshock is in fact mostly caused not by the mainshock itself, but rather by intermediate aftershocks. When decreasing $m_c$, the direct triggering effect due to large mainshocks could potentially be even further reduced relative to direct triggering by small shocks. This is particularly critical for understanding the physical mechanisms that cause earthquake triggering: The testing and validation of models first require the correct relation of the aftershocks to their trigger, rather than to an older, generally bigger, ancestor in the triggering chain.

**References and Notes**

1. K. R. Felzer, T. W. Becker, R. E. Abercrombie, G. Ekström, J. R. Rice, *J. Geophys. Res.* **107**, 2190 (2002).
2. D. Marsan, *Geophys. J. Int.* **163**, 141 (2005).
3. A. Helmstetter, Y. Y. Kagan, D. D. Jackson, *J. Geophys. Res.* **110**, B05S08 (2005).
4. J. K. Gardner, L. Knopoff, *Bull. Seismol. Soc. Am.* **64**, 1363 (1974).
5. P. Reasenberg, *J. Geophys. Res.* **90**, 5479 (1985).
6. S. D. Davis, C. Frohlich, *Geophys. J. Int.* **104**, 289 (1991).
7. G. M. Molchan, O. E. Dmitrieva, *Geophys. J. Int.* **109**, 501 (1992).
8. J. Zhuang, Y. Ogata, D. Vere-Jones, *J. Am. Stat. Assoc.* **97**, 369 (2002).
9. See supporting material on *Science* Online.
10. P. Shearer, E. Hauksson, G. Lin, *Bull. Seismol. Soc. Am.* **95**, 904 (2005).
11. G. Ouillon, D. Sornette, *J. Geophys. Res.* **110**, B04306 (2005).
12. B. E. Shaw, *Geophys. Res. Lett.* **20**, 907 (1993).
13. A. Helmstetter, *Phys. Rev. Lett.* **91**, 058501 (2003).
14. K. R. Felzer, R. E. Abercrombie, G. Ekström, *Bull. Seismol. Soc. Am.* **94**, 88 (2004).
15. C. H. Scholz, *The Mechanics of Earthquakes and Faulting* (Cambridge Univ. Press, Cambridge, 1991).
16. F. Tajima, H. Kanamori, *Phys. Earth Planet Inter.* **40**, 77 (1985).
17. A. Helmstetter, G. Ouillon, D. Sornette, *J. Geophys. Res.* **108**, 2483 (2003).
18. B. Gutenberg, C. F. Richter, *Bull. Seismol. Soc. Am.* **142**, 185 (1944).
19. S. Hainzl, F. Scherbaum, C. Beauval, *Bull. Seismol. Soc. Am.* **96**, 313 (2006).
20. D. Sornette, M. J. Werner, *J. Geophys. Res.* **110**, B09303 (2005).
21. We thank G. Daniel, A. Helmstetter, G. Ouillon, and D. Sornette for constructive discussions. Supported by European Community project VOLUME (FP6-2004-Global-3) (O.L.).

**Supporting Online Material:**

**« Extending earthquake' reach through cascading » by D. Marsan and O. Lengliné.**

We here test the method, using synthetic catalogs (section 1). We modify the method as described in the manuscript to account for the non-uniform distribution of background earthquakes over space (section 2). We run an analysis of the California earthquake dataset using 3D distances between the aftershock hypocenters and the mainshock rupture plane instead of the 2D epicentral distances as in the manuscript (section 3). Finally, we discuss the linearity hypothesis on which the method is based (section 4).

**1 – A test of the method:**

We demonstrate the ability of the method to decipher the multiple connections between earthquakes, by analyzing a synthetic catalogue using the Epidemic-Type Aftershock Sequence (ETAS) model (*S1, S2*). This model postulates that every earthquake triggers its own, local, aftershock sequence, therefore resulting in a cascade of triggering. Earthquakes are generated with a time and space-dependent rate

density $\lambda(\underline{x},t)=\lambda_0+\sum_{t_i<t} K\,e^{\alpha m_i}(t-t_i+c)^{-p}f(r,m_i)$ computed by summing the influence (kernels) of all past earthquakes of indices i. The rate density $\lambda(\underline{x},t)$ integrated over a given space-time volume

$V=(\underline{X},T)$ yields the mean number $\Lambda=\int_X d\underline{x}\int_T dt\,\lambda(\underline{x},t)$ of earthquakes occurring in $V$, so that the actual number is the realization of a Poisson law with mean $\Lambda$. Parameter $\lambda_0$ is the rate density of background earthquakes, i.e., earthquakes that occur randomly in space and time and that are

not triggered by previous earthquakes. We denote by $r=\left|\underline{x}-\underline{x_i}\right|$ the distance between earthquakes i

and j, and by $m_i$ and $t_i$ the magnitude and occurrence time of earthquake i. Parameters $K$ and

$\alpha$ constrain the direct aftershock productivity of each mainshock. The rate of directly triggered

aftershocks follows Omori-Utsu's law $(t+c)^{-p}$ with exponent $p$ and time cut-off $c$. The spatial

density is $f(r,m)=\dfrac{1}{2\pi r L \ln(1+\dfrac{1}{L})\times(1+\dfrac{r}{L})}$ for $0\leq r\leq 1$, and $f(r>1,m)=0$, so that

$\int_0^1 dr\, 2\pi r\, f(r,m)=1$, with $L\sim 10^{0.5m}$ the rupture length at magnitude $m$. For this test, we

used the parameters $\lambda_0=0.25$ earthquakes per unit time and unit surface, $\alpha=2$, $c=10^{-2}$,

$p=1.2$, $K=0.0094$, giving a branching ratio (mean number of direct aftershock triggered by an

earthquake) of 0.9. The magnitudes are distributed according to a Gutenberg-Richter law (*18*), hence a

probability density $f_m(m)=b\ln(10)\times 10^{-bm}$, with b-value equal to 1, and minimum magnitude

$m=0$. A total of $N=7010$ earthquakes were thus generated, over a duration $T=10^3$ (in

arbitrary units; Fig. S1). The largest earthquake has magnitude $m=4.61$, and a rupture length set to

$L=0.1$ (in arbitrary units). The earthquakes are constrained to occur on a $S=2\times 2$ surface, with

periodic boundaries, i.e., a torus.

We apply the algorithm to this simulated catalogue. We further simplify the method by decoupling the

spatial and the temporal dependence of the kernel: $\lambda_i(\underline{x},t)=\lambda_s(\left|\underline{x}-\underline{x_i}\right|,m_i)\times\lambda_t(t-t_i,m_i)$ . This

allows for a more robust estimation of the weights, by significantly reducing the number of kernel

values to determine. Since a direct dependence on magnitude can be ambiguously carried by either the

spatial density $\lambda_s$ (number of triggered earthquakes per unit area) or the temporal rate $\lambda_t$ (number

of triggered earthquakes per unit time), we constrain the spatial density to be normalized:

$$\int_0^\infty\int_0^\infty dx\,dy\,\lambda_s(x,y,m)=\int_0^\infty dr\,2\pi r\,\lambda_s(r,m)=1$$ . The initial guess for the rates / densities is done by

democratically setting equal weights to all preceding earthquakes (including the background term),

with no condition on their magnitudes: $w_{0,j}=w_{i,j}=\dfrac{1}{j},\forall i<j$ , and then starting the first iteration of

the algorithm directly at step 2 (i.e., updating of the rates). We postulate that convergence is reached

when the logarithm of the densities $\lambda_s(\left|\Delta\underline{x}\right|,m)$ or rates $\lambda_t(\Delta t,m)$ do not change by more than

1% for any given separation $\left|\Delta\underline{x}\right|$ or time $\Delta t$ , between two consecutive iterations. This criterion is

found to be well adapted: using more constraining criteria did not lead to visually different rate

estimates.


The solution obtained is in close agreement with the theoretical rates (Fig. S2). The method is therefore

well able to separate the individual triggering of each mainshock in a complex time series which

involves multiple, cascading, triggering. Again, this estimation was performed with no a priori

constraints on the shape of the rates / densities. Discretization in space, time and magnitude for

computing the kernels is the only arbitrary input.

A realisation of the background seismicity, i.e., the earthquakes that initiate clusters of aftershocks, is

computed by using the background probabilities $w_{0,i}$ . Such a realisation is compared to the total set

of earthquakes (Fig. S1). The background seismicity is clearly devoid of aftershock clusters as

expected, and occurs at a nearly constant rate density of 0.242, which is close to the true value

$\lambda_0 = 0.25$ . Monte-Carlo simulations were performed to check whether the estimated $\lambda_0$ is always

correctly determined. Forty runs of the ETAS model (keeping the same parameters) were analysed,

yielding an estimate of 0.248 +/- 0.010 for the background rate density. No significant correlation (at

the 90% confidence level) is observed between the estimate of $\lambda_0$ and the total number of

earthquakes produced in each simulation (that ranged from 1747 to 9539).

The convergence is equivalent to minimizing a cost function $J$ , classically defined as the opposite of

the logarithm of a likelihood function:

$$J = \lambda_0 \times S \times T + \sum_{i=1}^{N} \int_{0}^{T-t_i} dt \; \lambda_t(t, m_i) - \sum_{i=1}^{N} \log\left(\lambda_0 + \sum_{j=1}^{i-1} \lambda_t(t_i - t_j, m_j) \times \lambda_s(r_{ij}, m_j)\right)$$ , where $T$ is the total

duration of the earthquake catalog (Fig. S3). Changing the spatial discretization does not significantly

affect the obtained temporal rates: even in the case of a purely temporal method, i.e., not using any

information on distances between earthquakes, the final rates are close to the theoretical rates (Fig. S4).

Exploiting the spatial information mostly speeds up the convergence, as can be seen by comparing

Figure S4 to Figure S3.

Finally, the solution does not depend on the initial choice for the kernel. In particular, the same final

solution is obtained starting from the democratic kernel as defined above, or starting from the

discretized theoretical kernel (Figs. S3 and S4). The latter is not the best solution: it has a cost function

value $J$ of the order of +100 compared to the best solution found during the convergence of the

algorithm.

**2 – Non-uniform spatial distribution of background earthquakes:**

We modify the method described in the main text to account for the fact that background earthquakes

do not densely cover the region, although we still consider that they occur at a stationary rate (i.e.,

forming a Poisson process). The background term $\lambda_0(\underline{x}) = \lambda_{0,t} \times \lambda_{0,s}(\underline{x})$ is then the product of the

constant rate $\lambda_{0,t}$ with the local density $\lambda_{0,s}(\underline{x})$ conditioned on the position $\underline{x}$ relative to the

background earthquakes: $\lambda_{0,s}(\underline{x}) = \sum_{i=1}^{N} g(|\underline{x_i} - \underline{x}|) w_{0,i}$, with the normalisation $\int_0^\infty dr\, 2\pi r\, g(r) = 1$.

The goal is then to estimate $g(r)$. This is done by modifying the two steps of the algorithm: (step 1)

attribute the background weights $w_{0,j} = \alpha_j \lambda_0(\underline{x_j})$, and (step 2) compute $\lambda_{0,t} = \frac{1}{T} \sum_{i=1}^{N} w_{0,i}$ and

$$g(r) = \frac{\sum_{i,j \in A} w_{0,i} w_{0,j}}{S(r, \delta r) \sum_{i,j} w_{0,i} w_{0,j}}$$ , where $A$ is the set of earthquakes pairs such that

$\left| \underline{x_i} - \underline{x_j} \right| = r \pm \delta r$ . Additionally, a correction for changes in completness magnitude in the first few

hours after large mainshocks ( $m \geq 6$ ) is introduced: each aftershock counts for $10^{b(m_c(t)-3)}$ events,

where the b-value of the Gutenberg-Richter law (*16*) is 1.05, and $m_c(t) \geq 3$ is the completness

magnitude at the time $t$ of the aftershock.

## 3 – Analysis of California earthquakes using distances to faults:

We here perform the same analysis of California earthquakes as in the manuscript, but, instead of using

inter-epicentral distances, we here rather use distances between the causative fault and the target

hypocenter. This allows to account for the spatial anisotropy of the triggering.

Computing the distances:

The distance $r_{i,j}$ 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, see Table

S1 for references. All source models are provided by Martin Mai's finite source rupture model database

([www.seismo.ethz.ch/srcmod](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 all the other ( $m<6$ ) earthquakes, the fault plane is computed as followed: we first

selected earthquakes occurring within half a day following the considered mainshock and with

epicentral distances less than twice the length $d=10^{0.5\times(m-4)}$ in km. Note that this scaling is

consistent with the scaling found with the method when using inter-epicentral distances, see main

manuscript. The fault plane is then defined as the 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 hypocenter of the mainshock. The size of the (square) fault plane is

defined by its half-length $d$ as defined above.


Results of the analysis:

We use the same correction coefficients as with the analysis using inter-epicentral distances. Figures

S5 and S6 are the equivalent, for the new analysis, of Figures 1 and 2 of the main manuscript. In

comparison to the previous results, we note the following changes and similarities:


- the bare rates are larger for the $m \geq 6$ mainshocks (Fig. S5 a). The bare productivity of those

  earthquakes is significantly increased, yielding a scaling $\chi \sim 10^{0.86 m}$ (Fig. S6 b), hence a larger $a$

  parameter (increasing from 0.60 to 0.86). The dressed rates are very similar to the previous ones, at

  least for $t>0.1$ days; for shorter time intervals, they are increased for $m \geq 6$ mainshocks and

decreased for $m<6$. The dressed productivity scales as $\chi \sim 10^{0.73\,m}$, which is close to the

$a=0.66$ found previously. This does not change our earlier conclusion that small mainshocks

collectively trigger more aftershocks than their larger counterparts, since the exponent $a$ is still

significantly less than the b-value of the Gutenberg-Richter law for southern California ( $b=1.05$

, Fig. 3b).

• The mean aftershock sequence duration is magnitude dependent for both the bare and the dressed

kernels, but the latter is 6 to 10 times longer than the former (Fig. S6 c).

• The background rate of $m \geq 3$ earthquakes is lowered down to 0.11 day$^{-1}$, giving a 13%

probability of being a background earthquake at this magnitude cut-off.

• The spatial densities cannot be directly compared between the two analyses, as one is a 2D while

the other is a 3D kernel. The bare densities (at least at short distances $r<10$ km) and the dressed

densities decay with distance as $\lambda_s(x,y,z,m) \sim r^{-3}$, at all magnitudes (cf. dashed lines in Figure

S5 d). The median distance from the rupture plane to the directly triggered aftershocks is very

small, ranging between 1.0 km and 3.4 km, with no clear dependence on magnitude, implying that

most direct aftershocks occur very close to the rupture plane. The mean triggering distances are

increased by an average factor of 7 when considering the full aftershock cascade rather than only

the directly triggered aftershocks (Fig. S6 d). As with the previous analysis using epicentral

distances, aftershock diffusion away from the mainshock is only observed for the dressed kernel, although the process is significantly sub-diffusive, i.e., the mean distance grows as $r \sim t^{0.20}$ (Fig. S6 e).

- Interestingly, all significant differences between the two analyses affect the bare kernel. The dressed kernel characteristics are nearly identical: almost the same exponents are obtained for the duration scaling, for the aftershock zone expansion, and for the productivity. Since most changes to the bare kernel are found for large mainshocks while small mainshocks keep the same bare characteristics, this clearly confirms that the dressed aftershock sequences are dominated by small shocks.

In conclusion, considering the distance to the causative fault rather than the simpler epicentral distance affects the results of the analysis, in particular by giving more weight to the large mainshocks. This does not however modify the main conclusions of the previous analysis, i.e., (1) small earthquakes collectively trigger more direct and indirect aftershocks than large earthquakes; (2) the cascade of triggering significantly extend aftershock sequences (here by a factor of about 7 to 10) both in time and space, and causes a slow diffusion of the aftershock zone; (3) the background contribution is small, of the order of 13% to 19% at $m_c = 3$, and is expected to decay to even lower values at smaller magnitude cut-off.

**4 – Linearity hypothesis:**

The main assumption of this method is the linearity of the triggering process: the collective triggering of a set of triggers is simply the sum of their individual triggering, which stays the same whatever the history of the process. This assumption goes against the rate-and-state friction model (*S8*), which is non-linear. However, accounting for non-linearity would require an underlying model, hence making the method model-dependent. There is unfortunately no simple way to evaluate the degree of non-linearity of the triggering process directly from the data. The scaling of the productivity parameter $\chi$ however gives some hints: in case of linear triggering, this scaling (parameter *a*) must be the same for bare and for dressed aftershocks. We find that $a = 0.60 \pm 0.07$ (bare) and $a = 0.66 \pm 0.04$ (dressed) when looking at epicentral distances (see main manuscript), and $a = 0.86 \pm 0.09$ (bare) and $a = 0.73 \pm 0.04$ (dressed) when using distances to the fault (see above). We therefore cannot reject the hypothesis of linearity based on the value of this parameter.
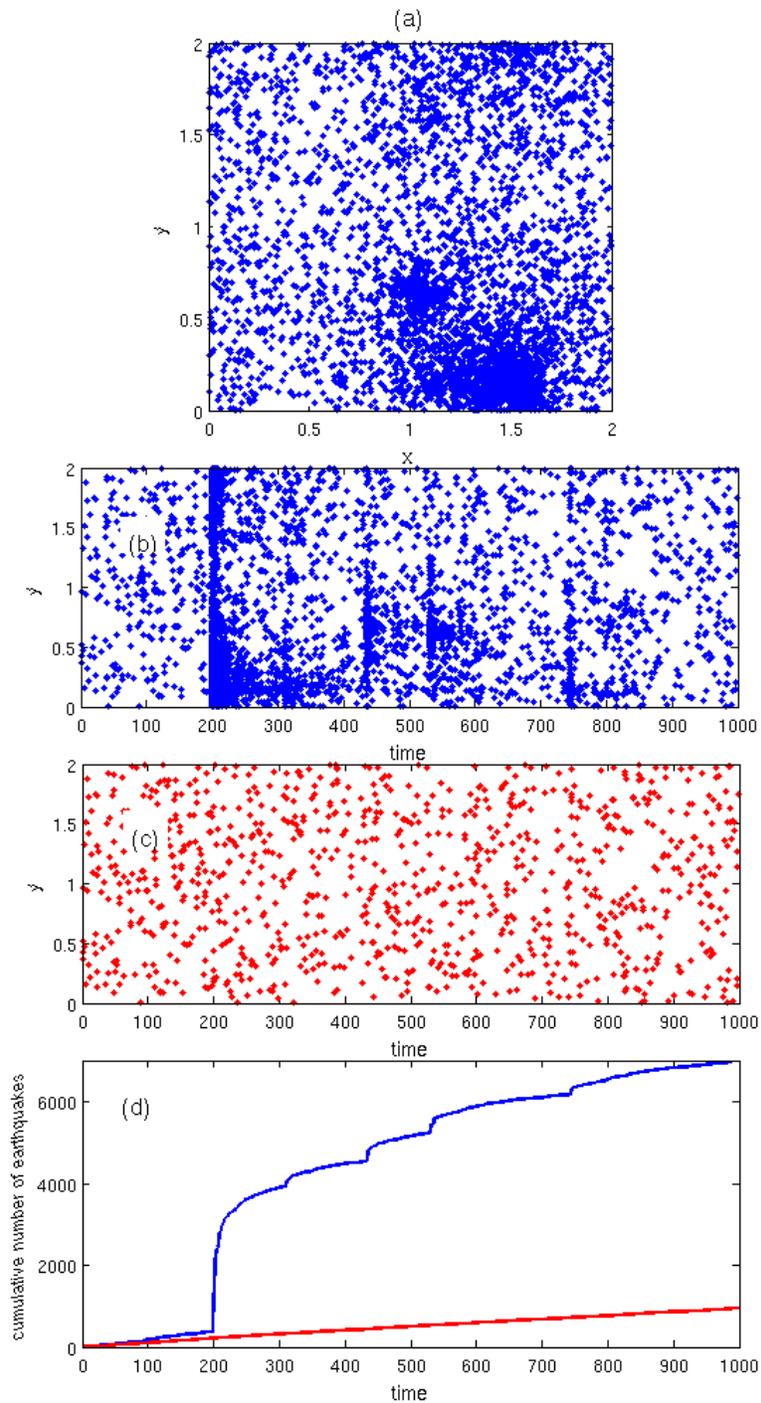
Figure S1: **synthetic earthquake catalogue.** (Blue) simulated catalogue using an ETAS model, compared to the catalogue declustered using the present method (in red). (a) Location of the epicentres; (b) and (c) space-time plots (along the y-direction); (d) cumulative time series. Aftershocks are efficiently removed by the declustering. The remaining background seismicity occurs at an estimated rate of 0.968 events / unit time, very close to the true 1 event / unit time imposed in the model.
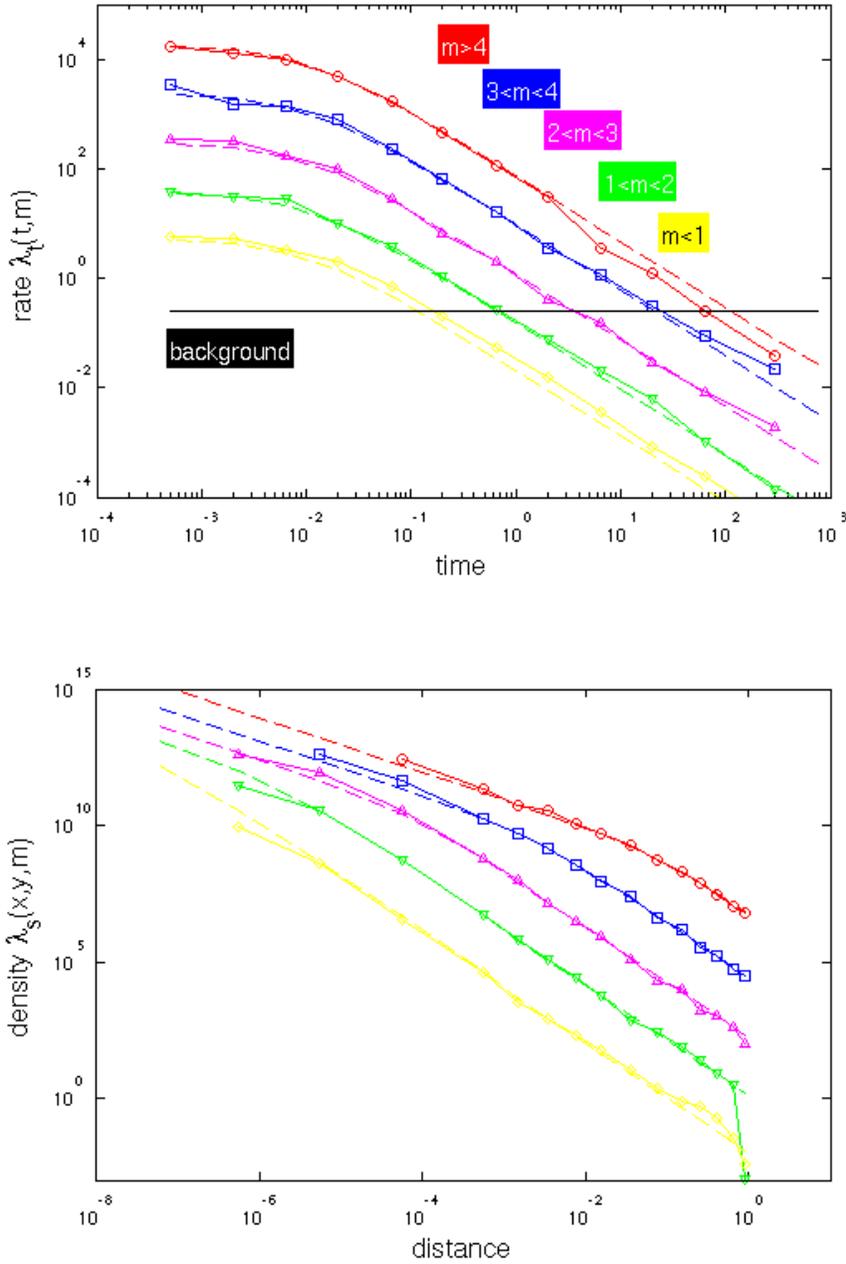
Figure S2: **estimated rates and densities for the synthetics.** Temporal rates $\lambda_t(t,m)$ (top graph, in earthquakes / unit time) and spatial densities $\lambda_s(x,y,m)$ (bottom graph, in earthquakes / unit area) estimated using the present method, compared to the theoretical $\lambda_t(t,m)=K\,e^{\alpha m}(t+c)^{-p}$ and $\lambda_s(x,y,m)=f(r,m)$ (dashed lines), for the synthetic catalogue of Figure S1. The spatial densities are shifted for clarity.
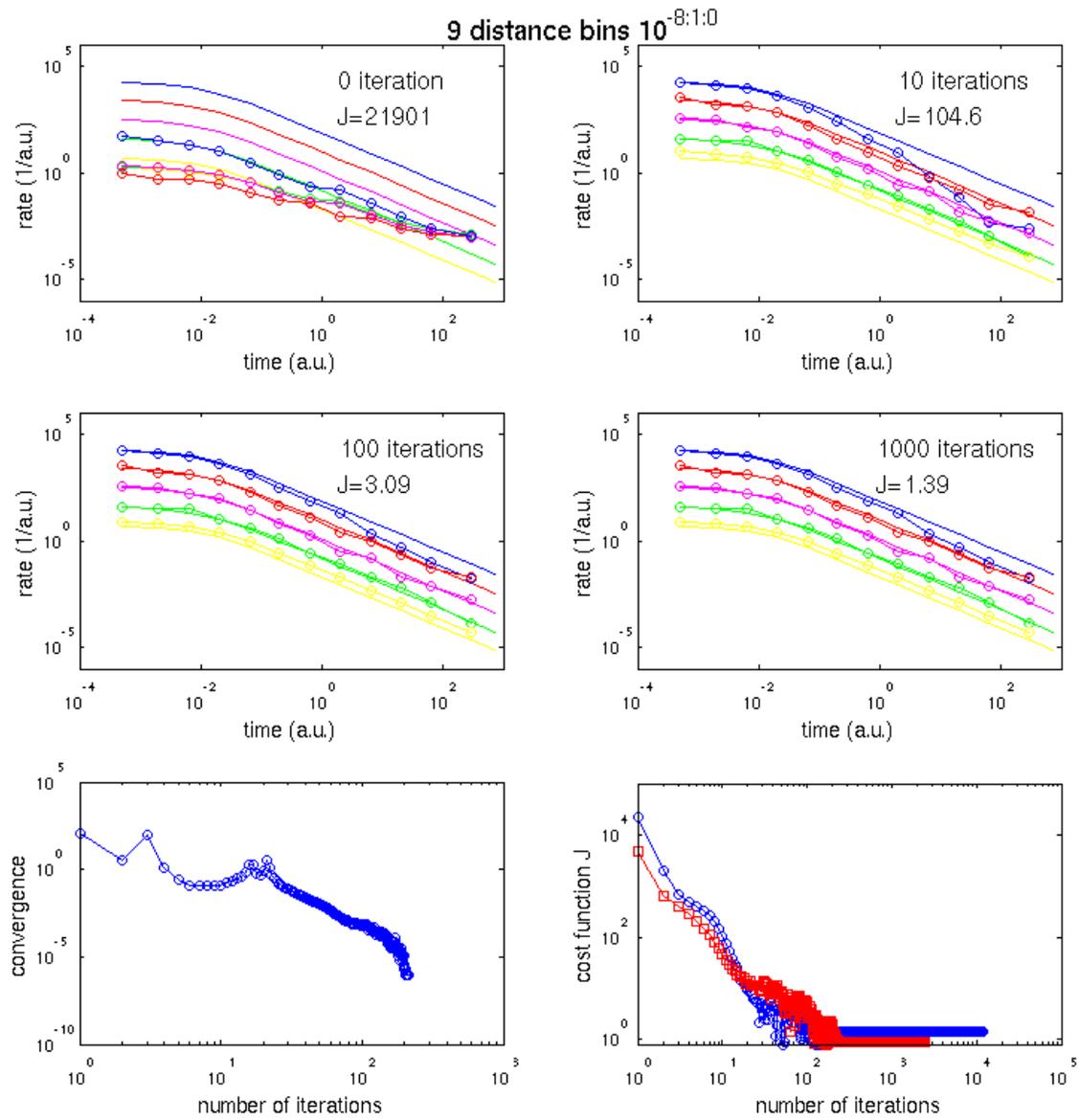
Figure S3: **snapshots of the temporal rates during convergence, for a space-time kernel discretized using 9 distance intervals.** Top four graphs: rate $\lambda_t(t,m)$ at starting point, and after 10, 100 and 1000 iterations, compared to the theoretical rates. Bottom left graph: convergence criterion vs. number of iterations. Bottom right graph: cost function $J$ vs. number of iterations, for two different initial conditions (democratic kernel in blue, theoretical kernel in red). The cost $J$ is shifted so that $J=0$ corresponds to the minimum value found during the convergence (the same for the two starting points).
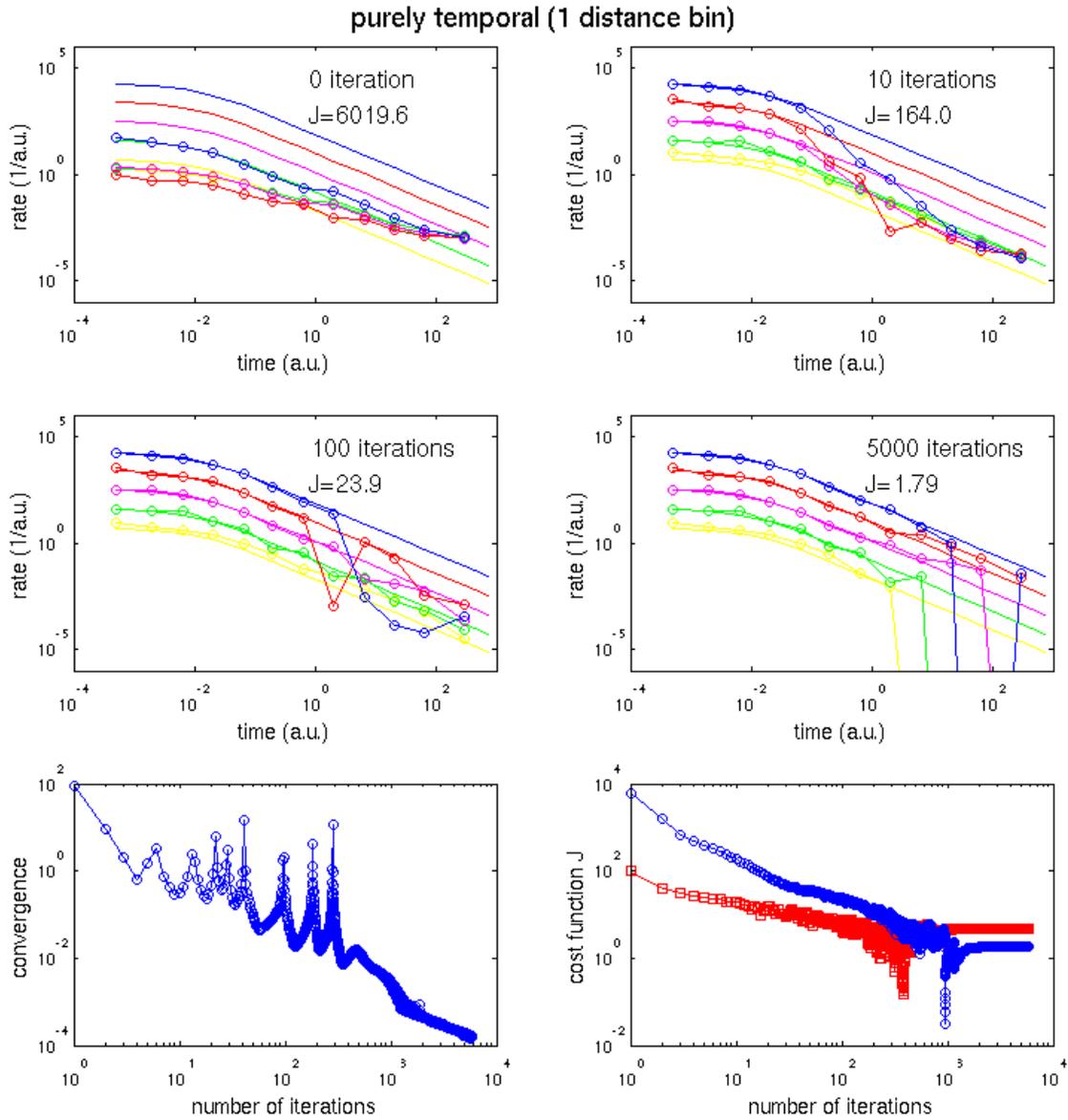
Figure S4: **snapshots of the temporal rates during convergence, for a purely temporal kernel.**
Same as Figure S3, but for a kernel not using any information on distances between earthquakes. The snapshots are taken at starting point, and after 10, 100 and 5000 iterations.
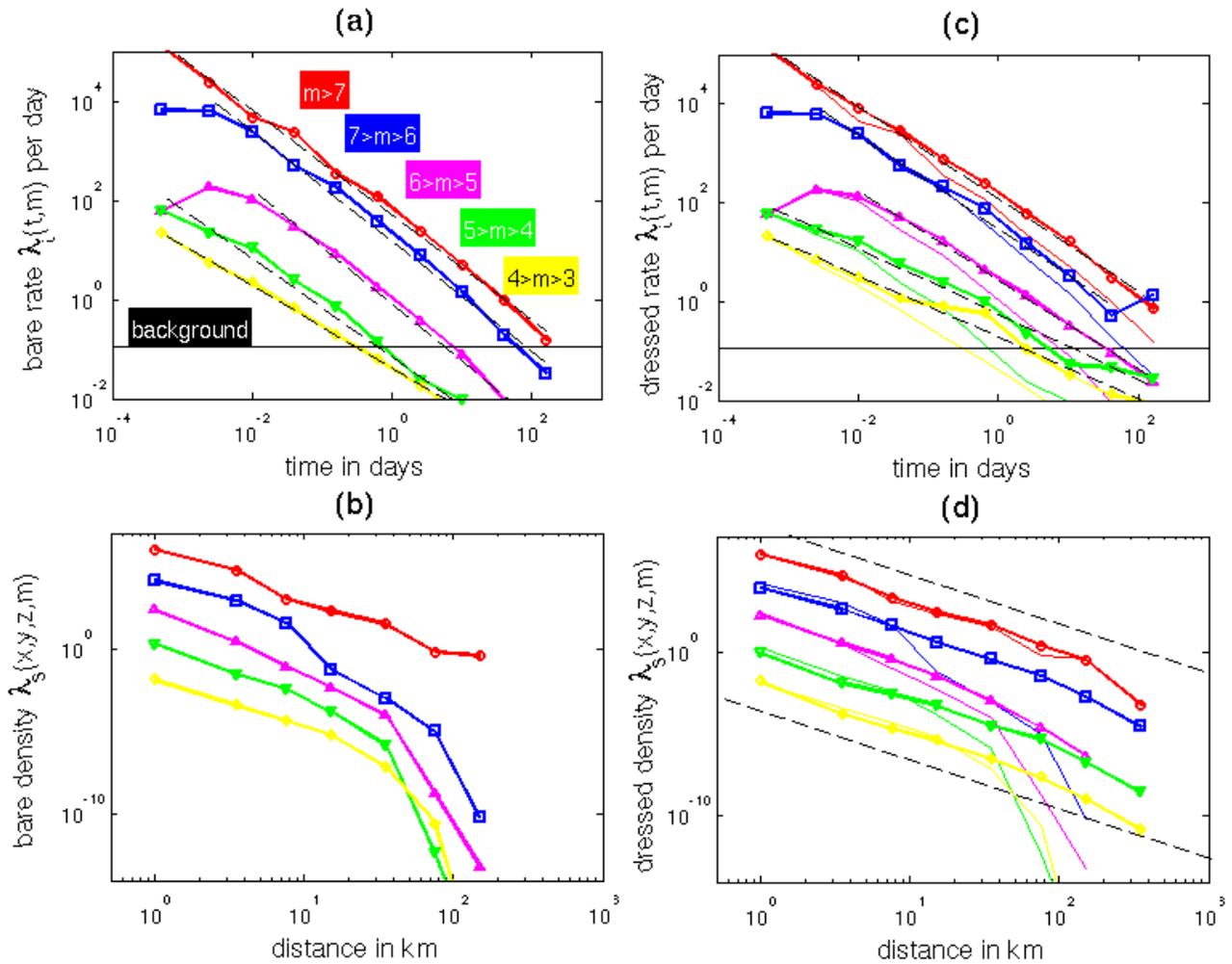
Figure S5: **rates and densities for California earthquakes using distances to faults.** As in Figure 1, for distance between earthquakes computed as the shortest distance from the causative fault to the target hypocenter (rather than the epicentral distance). The dashed lines in graph (d) indicate a $1/r^3$ scaling.
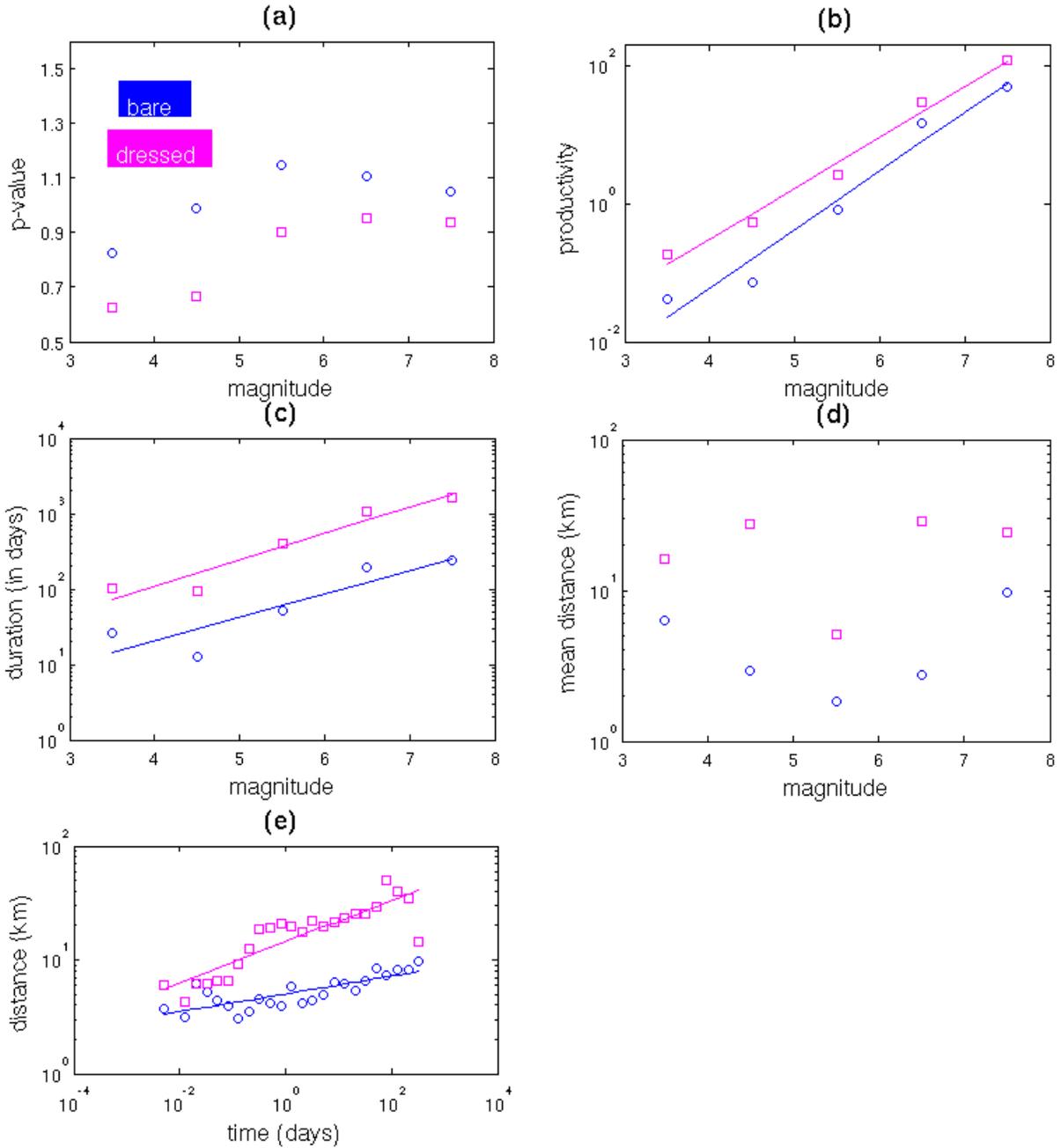
Figure S6: **aftershock sequence characteristics for California earthquakes usin distances to faults.** As in Figure 2, for distance between earthquakes computed as the shortest distance from the causative fault to the target hypocenter (rather than the epicentral distance). The fits are: (b) productivity $\chi \sim 10^{0.86\,m}$ (bare rate) and $\chi \sim 10^{0.73\,m}$ (dressed rate), (c) mean duration $t \sim 10^{0.31\,m}$ (bare) and $t \sim 10^{0.35\,m}$ (dressed), (e) mean mainshock – aftershock distance $r \sim t^{0.07}$ (bare) and $r \sim t^{0.20}$ (dressed).

| Date | Location | Magnitude | Reference |
|:---:|:---:|:---:|:---:|
| November 23, 1987 | Elmore Ranch | 6.2 | S6 |
| November 24, 1987 | Superstition Hill | 6.6 | S6 |
| April 22, 1992 | Joshua Tree | 6.1 | S3 |
| June 28, 1992 | Landers | 7.3 | S7 |
| June 28, 1992 | Big Bear | 6.3 | S4 |
| January 17, 1994 | Northridge | 6.7 | S7 |
| October 16, 1999 | Hector Mine | 7.1 | S5 |
| October 16, 1999 | Hector Mine aftershock | 6.7 | S5 |

Table S1: references to rupture models for all m ≥ 6 earthquakes.

**References:**

S1. Y. Y. Kagan, L. Knopoff, *J. Geophys. Res.* **86**, 2853 (1981).

S2. Y. Ogata, *J. Am. Stat. Assoc.* **83**, 9 (1988).

S3. S. E. Hough, D. S. Dreger, *Bull. Seismol. Soc. Am.* **85**, 1576 (1995).

S4. L. E. Jones, S. E. Hough, *Bull. Seismol. Soc. Am.* **85**, 668 (1995).

S5. A. Kaverina, D. Dreger, E. Price, *Bull. Seismol. Soc. Am.* **92**, 1266 (2002).

S6. S. Larsen, R. Reilinger, H. Neugebauer, W. Strange, *J. Geophys. Res.* **97**, 4885 (1992).

S7. Y. Zeng, J. Anderson, Evaluation of numerical procedures for simulating near-fault long-period ground motions using Zeng method, *Report 2000/01 to the PEER Utilities Program, available at* *http://peer.berkeley.edu* (2000).

S8. J. Dieterich, *J. Geophys. Res.* **99**, 2601 (1994).