Decay of aftershock density with distance indicates triggering by dynamic stress

K. R. Felzer
U. S. Geological Survey
525 S. Wilson, Pasadena, CA 91106

E. E. Brodsky
Dept. of Earth Science, UC Santa Cruz
1156 High St., Santa Cruz, CA 95060

February 10, 2006
The majority of earthquakes are aftershocks, yet aftershock physics is not well understood. Most studies suggest that static stress changes trigger aftershocks, but recent work suggests that shaking (dynamic stresses) may play a role. We measure the decay of aftershocks as a function of distance from M 2–6 mainshocks in order to clarify the aftershock triggering process. We find that for short times after the mainshock, when low background seismicity rates allow for good aftershock visibility, the decay is well fit by a single inverse power law over distances of 0.2 km – 50 km. The consistency of the trend indicates the same triggering mechanism is working over the entire range. Since static stress changes at the more distant aftershocks are negligible (< 10 Pa) this suggests that dynamic stresses are triggering all of the aftershocks. If the linear density of potential hypocenters is constant in space, the observed aftershock density is consistent with the probability of triggering aftershocks being nearly proportional to seismic wave amplitude. The data is not well-fit by models that combine static stress change with the evolution of frictionally locked faults.

Previous studies of how aftershock density decays with distance from the mainshock have found a range of functions, including power laws and combinations of power laws with constants and exponentials. The ability to study the decay with improved clarity has recently been provided by the publication of large catalogs with precise earthquake locations. We use the Shearer et al. relocated 1984–2002 Southern California catalog. Standard location error is on the order of tens of meters, similar to that obtained in more localized studies using similar relocation techniques. For most of the analysis we use M 2–4 mainshocks and M ≥ 2 aftershocks. We prefer small mainshocks because their large number allows for good statistical averaging and for the use of a small difference between mainshock and aftershock magnitude, which improves catalog completeness. For near field measurements, where larger mainshocks are necessary for appropriate range and precision, we use M 5–6 mainshocks and M ≥ 3 aftershocks. Earthquakes are used as mainshocks if they are sufficiently isolated from larger earthquakes in time and space (see Methods). To improve statistics we combine the aftershocks of each unit magnitude range of mainshocks (Figure 1).
The spatial density of a point process can be measured in any dimension. For instance, a density could be the number of points per length, per area, or per volume. We choose to measure linear density, i.e., the number of aftershocks per length. This is done by collapsing all of the aftershocks onto a single line with their position on the line equal to their distance from the mainshock. The number of aftershocks per unit length can then be measured at different positions using standard statistical tools (See Methods).

We first study earthquakes occurring within 5 min and 50 km of M 2–4 mainshocks. The short time window minimizes the amount of background seismicity\textsuperscript{13}, i.e., earthquakes not aftershocks of the designated mainshocks. We approximate the mainshocks as point sources and measure the distance, $r$, between mainshock and aftershock hypocenters. From 0.2 to 50 km the data is well fit by

$$
\rho(r) = cr^{-n},
$$

where $n = 1.37 \pm 0.1$ for the $3 \leq M < 4$ mainshocks and $1.35 \pm 0.12$ for the $2 \leq M < 3$ mainshocks, and $c$ is a constant that varies with the number of aftershocks (Figure 2). The error is the 98% confidence interval based on 5000 bootstraps. For $r < 0.2$ km the point source approximation is no longer accurate (Supplementary Figures 1 and 2).

We also check the applicability of this function to longer times. For 30 minutes of post-mainshock data the time window is still short, and an inverse power law, with $n = 1.36 \pm 0.07$, fits the $3 \leq M < 4$ mainshocks. For $2 \leq M < 3$ mainshocks background seismicity begins to level the decay around 16 km, but a clear inverse power law with $n = 1.37 \pm 0.15$ is evident at shorter distances (Supplementary Figure 3). Background interference is worse for smaller mainshocks because of the lower aftershock productivity per mainshock. For 30 minutes–25 days of post-mainshock earthquakes there is universal deviation from a pure power law; however we find that a combined aftershock/background function fits the data well, at 95% and 65% confidence (Supplementary Figure 4). This suggests that the rate of aftershock decay with distance does not change with time.
To study the densities of aftershocks within one fault length of the mainshock we use M 5–6 mainshocks that have a Harvard CMT focal mechanism solution and aftershock distributions that clearly delineate the preferred fault plane. We estimate the length and width of the faults from empirical relationships\textsuperscript{14}, and center the fault planes at the median aftershock location. We measure aftershock density from 0.2 km to 12 km from the fault plane, (\(~0.05\) fault lengths of an M 5 earthquake to 1 fault length of an M 6 earthquake). As before, we recover an inverse power law, with \(n = 1.34 \pm 0.25\) (Figure 3). The decay levels out at \(r < 0.2\) km due to error in fault plane location and catalog incompleteness in the very near field.

To verify that the decay we observe is due to aftershock physics we repeat the analysis for the M 2–4 mainshocks with a time-randomized catalog. This produces a large scatter of points (Figure 4), indicating that the pattern in Figure 2 is aftershock related. To verify that a pure inverse power law is a good functional fit we use the Kolmogorov-Smirnoff test (see Methods). We also test the fit of a composite power law of the form \(\rho(r) = a_1 r^{-n_1} + a_2 r^{-n_2}\). The Bayesian Information Critierion\textsuperscript{15,16} prefers the single power law fit. To check if our results are catalog dependent we verify that aftershocks in unrelocated Japanese and Northern California catalogs also follow an inverse power law decay (Supplementary Figure 5).

The consistent aftershock decay relationship observed from distances of 0.2 km to 50 km — from within 0.05 fault lengths of M 5 mainshocks to over 100 fault length of M 2–3 mainshocks — implies that static stress change is not triggering the aftershocks. Static stresses decay rapidly. At 3 km from a M 4 earthquake the static stress change is at most 4 kPa, comparable to tidal stresses that have not been found to trigger earthquakes\textsuperscript{17}; at 10 km from a M 3 the static stress change is 3 orders of magnitude lower. Triggering by static stress in the near field and dynamic stress in the far field would require a discontinuity in the aftershock decay. Only uniform triggering by dynamic stress matches the observation of a single, consistent decay that traverses a wide range of distances.

The hypothesis of aftershock triggering by static stress change has received strong support in part because there is a physical model, rate and state friction,\textsuperscript{3} which explains how static stress
changes could result in the power law distribution of aftershock times\textsuperscript{18}. Our observations indicate, however, that this model does not fit the distribution of aftershocks in space. At very short times, using a point source approximation for the static stress in a whole space, the model predicts that the density of aftershocks from an earthquake of moment $M_0$ is

$$\rho(r) = B(r)e^{c_2r^{-3}}$$

(2)

where $B(r)$ is the background seismicity per kilometer per unit time as a function of distance from the mainshock and $c_2 = M_0/4\pi A\sigma$ where $\sigma$ is normal stress and $A$ is a frictional constant. At a given distance from a set of mainshocks, the observed density in a combined dataset is the sum of the individual ones.

The functional form of $B(r)$ can be estimated from distance measurements between random earthquakes (Figure 4). The scatter is large, but the data can be fit with the same functional form used by others\textsuperscript{19, 20},

$$B(r) = c_1 r^{(D-1)}.$$  

(3)

This function describes points randomly scattered on a structure with effective dimension $D$. We substitute Equation 3 into Equation 2 and use a gridsearch to find the best fit to the 30 min/16 km aftershocks of M 3-4 mainshocks over a wide parameter range (see Methods). For the best fit case, with $D = 0.1$, the summed squared error is 1.4 times worse than for the best inverse power law fit, and the correlation of the data residuals is high ($r = 0.45$; 601 data points) (Supplementary Figure 6). This correlation has less than 0.01% chance of occurring for a good functional fit. For the residuals of the inverse power law $r = 0.0079$. For more realistic values of $D=1–2$, the fit is much worse (Supplementary Figure 6). The point source approximation may produce inaccurate representation of the static stress change within one fault length of the source (e.g. within 1 km for M 4 mainshocks and 0.1 km for M 2 mainshocks), but the functional shape and associated misfit is problematic at further distances as well.

The poor fit of the data to Equation 2 indicates that the aftershocks are not triggered by static
stress change coupled with rate and state friction, at least at distances beyond one fault length. We also find more model-dependent evidence that the number of aftershocks triggered varies linearly with dynamic stress change amplitude. Linear aftershock density, $\rho(r)$, can be separated into geometric and physical terms,

$$
\rho(r) = \frac{N_{\text{aft}}}{\Delta r} = \frac{N_{\text{hyp}}}{\Delta r} \times \frac{N_{\text{aft}}}{N_{\text{hyp}}} = cr^{-1.4}
$$

(4)

where $N_{\text{aft}}$ is the number of aftershocks in a shell of width $\Delta r$ centered at distance $r$ from the mainshock and $N_{\text{hyp}}$ is the number of potential hypocenters, or places where aftershocks could be triggered, in the same shell. If the background seismicity is roughly evenly distributed over the active fault population, the functional form of the background earthquake distribution (Equation 3) can be substituted for the geometric term,

$$
\rho(r) = c_1 r^{D-1} c_3 r^m = cr^{-1.4}
$$

(5)

The trade off between $D$ and $m$ prevents a direct solution for $D$. We instead estimate $D$ by fitting the data in Figure 4. We find a better fit with $D = 1$ than with $D = 2$ or 3; i.e. the linear density is independent of distance. A $D$ value close to 1 is also suggested by geometric considerations (see Methods).

If $D \approx 1$, the fraction of potential hypocenters triggered decays as $r^{-1.4}$, i.e. at a rate slightly stronger than $1/r$. The maximum amplitude of seismic waves also decay somewhat faster than $1/r$. The standard Richter relationship for maximum short period seismic wave amplitude is well-fit by a combination of $r^{-1.2}$ decay and anelastic attenuation with $Q=300$. Anelasticity is less important at the depth of the aftershocks so the peak amplitude of shaking at depth falls off as $r^{-1.2}$ (Figure 2). The power law decay of local seismic amplitude is primarily due to the geometrical spreading of the S-waves, wave focussing, and surface wave amplitude decay. Given the uncertainty in $D$, the data is consistent with the probability of triggering an aftershock being
proportional to the amplitude of the seismic wave. Scaling the probability of triggering with the amplitude of the waves is also consistent with the number of aftershocks triggered increasing by a factor of 10 with each mainshock magnitude unit as has been observed\textsuperscript{24, 25, 20}.

In summary, the decay of aftershock linear density with distance from M 2–6 mainshocks is well fit by an inverse power law. The trend exists at least from 0.2 to 50 km for the first 5 minutes of the aftershock sequence. At longer times background seismicity makes distant aftershocks more and more difficult to detect, but we can trace the decay trend for at least 16 km for the first 30 minutes, and to at least 10 km for the first 2 days. If the linear density of faults is independent of distance ($D \approx 1$) then the data indicates that the probability of triggering an aftershock is directly proportional to the amplitude of seismic shaking. Whether or not $D = 1$, dynamic triggering is preferred by the data. The similarity of aftershock decay from distances of 0.05 to over 100 fault lengths implies a single physical triggering mechanism, and dynamic stress change is the only plausible agent over most of this range.

**Methods**

**Mainshock and aftershock selection.** Earthquakes are used as mainshocks if they are separated from larger earthquakes by at least $L$ km or by $t_1$ days if the larger earthquake comes first, and $t_2$ days if it comes after. This separation minimizes contamination from aftershocks of larger earthquakes. $L$ is set at 100 km; larger values (at least up to 500 km) produce the same results. The results are insensitive to the values of $t_1$ and $t_2$ as long as $t_1 << t < t_2$ where $t$ is the time after the mainshock for which we use aftershock data. For $t = 5$ min and $t = 30$ min we obtain the same aftershock decay for values of $t_1$ between 3 and 100 days. We use $t_1 = 3$ days, which, within this range, maximizes the number of qualified mainshocks, and $t_2 = 0.5$ days. For Supplementary Figure 4, where $t$ can be as large as 25 days, we set $t_1 = 100$ days and $t_2 = 26$ days. For Figure 3, where $t = 2$ days, we use $t_1 = 30$ days and $t_2 = 2$ days.

We use a uniform distance cutoff for measuring aftershocks of all mainshocks, even though the mainshock magnitudes vary. This is justified by our observation in agreement with previous authors\textsuperscript{26, 8} that the distribution of aftershock distances is independent of mainshock magnitude (Supplementary Figure 7).

**Linear density measurement.** To measure linear density we use the nearest neighbor method\textsuperscript{27}, in
which densities are estimated by taking the inverse of the width of the box required to contain $k$ neighboring points. The edges of sequential boxes meet between data points. Smoothing is controlled by $k$. We find that our results are constant for $k = 1–20$, although the fitting error increases with $k$ (Supplementary Figure 8). We use $k = 1$, which produces the smallest error and minimum bias. The advantages of the nearest neighbor method are that the number and location of measurements are determined by the location of the data points, smoothing is uniform, and there are no empty bins.

The definition of linear density averages over all azimuths and is therefore insensitive to radiation pattern.

**Catalog completeness.** We check catalog completeness by fitting the Gutenberg-Richter magnitude frequency relationship, with $b = 1$. The aftershocks of M 2–3 mainshocks are complete to M 2. For M 3–4 mainshocks 10%– 15% of the smallest aftershocks (M 2.0–2.2) measured over the first 30 minutes may be missing, but there is no systematic loss with distance and thus no expected effect. Over the first 5 minutes closer to 30% to 40% of the smallest aftershocks of the M 3–4 mainshocks may be missing, but again the loss is not significantly systematic with distance. The $M \geq 3$ aftershocks used for the M 5–6 mainshocks are complete.

**Goodness of fit.** If the proposed single inverse power law fit is adequate the residuals should have a distribution similar to that of data drawn from a pure power law distribution via Monte Carlo simulation. We test for similarity between the residuals of simulated data and the largest robust data set in this study (Southern California 30 min/16 km aftershocks of M3-4 mainshocks) with the Kolmogorov-Smirnoff test. At 95% and 65% confidence the test indicates that the null hypothesis that the observed and simulated residuals come from the same distribution cannot be rejected.

We test the composite power law $\rho(r) = a_1 r^{-n_1} + a_2 r^{-n_2}$ by fitting $a_1$ and $a_2$ with a nonlinear least square algorithm for every value of $n_1$ and $n_2$ between 0.5 and 3. We find that the single power law is preferred (Bayesian Information Criterion).

**Testing rate and state friction.** To fit the rate and state friction equation to the aftershock density data we try the parameters: $c_1$: $10^{-5.72} - 10^{-1.172}$, stress drop: 0.1–10 MPa, normal stress 10–1000 MPa, $A$: 0.005–0.012, $D$: 0.1–2. The resulting range of $c_2$ is 0.008–200 $\times (M_0/4\pi)$. We achieve a minimum least squared error at $D = 0.1$, $c_1 = 0.0034$, and $c_2 = 0.33(M_0/4\pi)$. 

8
Potential hypocenter distribution effective dimension $D$. Because of the limited seismogenic depth of Southern California, at length scales over $\sim 10$ km the system would effectively be 2D if potential hypocenters were randomly scattered throughout the crust. In reality earthquakes concentrate on planar faults, whose width is also limited by the seismogenic depth. At distances longer than $\sim 10$ km–20 km, effective $D$ for earthquakes randomly scattered on a fault tends towards 1. Multiple faults increase $D$, but the concentration of earthquakes in streaks and clusters on faults decreases it\textsuperscript{11}. Thus we expect $D \approx 1$ at large distances, and the lack of any break in the slope of the aftershock density decay, from 200 m to 100 km, for a wide range of mainshock magnitudes, suggests that $D \approx 1$ throughout.

References


**Supplementary Information** is linked to the online version of the paper at www.nature.com.nature

**Acknowledgements** We thank R. Abercrombie, A. Felzer, N. Field, M. Gerstenberger, J. Gomberg, S. Gross, J. Hardebeck, M. Harrington, A. Helmstetter, S. Hough, L. Jones, Y. Kagan, H. Kanamori, I. Main, S. Prejean, P. Shearer, R. Stein, J. Vidale, and A. Yong for comments. This work was supported in part by the National Science Foundation.

**Author Information** The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.R.F. (kfelzer@gps.caltech.edu).
Figure 1: Combined aftershocks of M 3-4 mainshocks. To create a composite aftershock data set we move all of the mainshocks to the origin in space and time and move their aftershocks with them. Data here is for the first 30 min of aftershocks of M 3-4 mainshocks. The gray star gives the locations of the mainshocks, at the origin.

Figure 2: Distance from the mainshock hypocenter vs. aftershock linear density. Aftershocks are $M > 2$ and occur in the first 5 minutes. (a) $2 \leq M < 3$ mainshocks. The plot uses 7396 mainshocks and 104 aftershocks. (b) $3 \leq M < 4$ mainshocks. The plot uses 2355 mainshocks and 199 aftershocks. The data is fit with an inverse power law with an exponent of $-1.35 \pm 0.12$ for the $2 \leq M < 3$ data and $-1.37 \pm 0.1$ for the $3 \leq M < 4$ data (black lines). The fit is made from 0.2 to 50 km for both plots. For comparison, dashed gray lines give the decay of maximum seismic wave amplitude, a proxy for dynamic stress, as derived from the standard Richter relationship \(^2\).

Figure 3: Aftershock density vs. distance from the closest point on the fault planes of M 5–6 mainshocks. The plot uses 9 mainshocks and 104 $M > 3$ aftershocks that occurred within 2 days of the mainshock. A fit to the data is made from 0.2 to 12 km, or from 0.05 fault lengths of an M 5 to 1 fault length of an M 6 earthquake. The decay rate shows good agreement with the far field decay found for M 2-4 mainshocks.

Figure 4: Distance vs. earthquake linear density for a time randomized catalog. Distances are measured between 10 random earthquakes from the mainshock data set and random catalog earthquakes, producing 346 earthquake pairs. Unlike the aftershock data, there is no systematic decay of density with distance. The sum of squared residuals is lowest when the data is fit with the relationship that linear earthquake density is constant (solid line), as opposed to linear density equals $r$ (dashed line), which would correspond to hypocenters being located randomly in 2D, or $r^2$ (dashed-dotted line), which would correspond to random locations in a volume. (See Eq. 3.)
\( \rho(r) = c r^{-1.35 \pm 0.1} \)

\( \rho(r) = c r^{-1.37 \pm 0.1} \)
$\rho(r) = cr^{-1.34 \pm 0.25}$
Distance from mainshocks (km)

Linear density (earthquakes/km)

Felzer_Brodsky Figure 4
Complete data for M 3-4 Southern California mainshocks. In Figure 2 in the main paper we show the 5 minute aftershocks of M 2—4 mainshocks over the range 0.2 km – 50 km. This is the range that most accurately represents how aftershock density decays with distance. In order to provide the full breadth of the most important data set in this paper, we plot the densities of the five minute M>2 aftershocks of 3≤M<4 mainshocks from 0 to 200 km. Black dots are data as analyzed in Fig. 2; Grey triangles are the same data with an alternative mainshock exclusion criterion of L=500 km (see Methods).

At distances less than a fault length, the observed density is constant rather than following the power law discussed in this paper. The extreme nearfield behavior is a result of measuring aftershock distances from the hypocenter rather than the fault plane (See Supp. Fig. 2 and Fig. 4 in the main text).

At far distances there are more background earthquakes than aftershocks, and thus the aftershock decay can no longer be observed. Since the mainshock selection criteria only eliminated larger mainshocks within 100 km (see Methods), there is an artificial increase in apparent background seismicity at distances beyond 100 km. Performing the same analysis with a 500 km exclusion zone results in nearly constant background seismicity in the 100-300 km range and thus in nearly level observed farfield earthquake density (grey triangles).
Supplementary Figure 2

Demonstration via simulation that the decrease in the decay rate at short distances is due to hypocentral versus fault plane distance measurements. For the Monte Carlo simulation data (gray triangles) aftershocks are assigned a power law distribution of distances from points on planar vertical faults. The same number of mainshocks and same distribution of mainshock magnitudes is used as in the real data, and the strike of each mainshock is allowed to vary randomly. A normally distributed location error with a standard deviation of 20 m is added to each aftershock location. Distances are measured from the aftershock hypocenter to the mainshock hypocenter. We do not simulate variations in mainshock dip.

Catalog incompleteness is modeled to mimic observations of the five minute aftershock catalog data where completeness is determined by comparison with the Gutenberg-Richter relationship. In the simulation 75% of the aftershocks of M 3—3.5 mainshocks closer than one fault length are eliminated and 93% of aftershocks of M 3.5—4 mainshocks closer than one fault length are eliminated. As stated in the Methods section, in the distance range used to fit Equation 1 (0.2-50 km), fewer aftershocks are missing and the percentage of missing earthquakes is not strongly distance dependent. Simulation trials in which no earthquakes are eliminated also show the key features of this figure, i.e., a corner and decrease in the decay rate at one fault length due to the breakdown of the point source approximation.
Supplementary Figure 3

Distance from the mainshock vs. linear aftershock density for 30 min of aftershocks after each mainshock. All aftershocks are M ≥ 2. The data is fit with an inverse power law (solid gray lines) over 0.2-16 km. Background interference is small for these datasets in this distance range (see below). (A) Results for 2 ≤ M < 3 mainshocks. This plot uses 7396 mainshocks and 219 aftershocks. (B) Results for 3 ≤ M < 4 mainshocks. This plot uses 2355 mainshocks and 610 aftershocks.

There are many more mainshocks than observable aftershocks. This is because the minimum aftershock magnitude is kept very close to the mainshock magnitude to ensure catalog completeness at distances used for the fit, so the large number of aftershocks that are much smaller than their mainshock are not included. The observed number of aftershocks is consistent with standard productivity relations for the small mainshock magnitudes.

Background seismicity becomes apparent for aftershocks of the 2 ≤ M < 3 mainshocks at a closer distance (16 km) than it does for aftershocks of the the 3 ≤ M < 4 mainshocks (> 50 km) because many more 2 ≤ M < 3 mainshocks are used in the analysis, thus the total time represented in the plot (30 minutes per mainshock x 7396 mainshocks = 154 days) is longer than the total time used for the 3 ≤ M < 4 plot (49 days). This longer time allows more background earthquakes to accumulate. The larger quantity of background earthquakes means that the aftershock signal is obscured at shorter distances in the far field where aftershock numbers are small. The larger background is not a significant factor in the near field, where the aftershock to background earthquake ratio is high.
Aftershocks over longer time windows. Data is from the Shearer et al. relocated Southern California catalog; mainshocks are M 3-4, aftershocks are M>2. We study distances from 0.2 – 16 km, following Figure 2 in the main text. Because this data covers a longer time period stricter rules are used to find mainshocks that are isolated from larger earthquakes, and so fewer mainshocks can be used (see Methods). There are 191 mainshocks in this data set.

The longer time spans include a larger number of background earthquakes. We fit the data with \( p(r) = cr^{-1.4} + b \), where the fit parameters \( c \) and \( b \) are controlled by the number of aftershocks and background earthquakes, respectively. We minimize the least-square error to fit \( b \) and \( c \). Because aftershock sequences decay rapidly we expect the percentage of aftershocks to drop with time.

(A) Data from 30 minutes to 1 day after the mainshock is best fit with 48% aftershocks, given by the gray line (\( c = 13.0, b = 3.5 \)). A very small correlation coefficient of the data residuals (\( r = -0.004 \)) and a Kolmogorov-Smirnoff test (at 95% and 65% confidence) confirms a good fit. There are 107 aftershocks in this plot. (B) Data from 1-3 days is best fit with 34% aftershocks, \( c = 9.9, b = 4.7; r = -0.008 \). 114 aftershocks used. (C) From 3-25 days, the best fit is with 4% aftershocks, \( c = 4, b = 25; r = 0.007 \). 452 aftershocks are used. (D) Combined 30 minutes – 25 days, \( c = 24, b = 33 \). 630 aftershocks are in this plot. If our previous solutions are correct, then 17% of this total should be aftershocks; the best fit gives 15%. As an additional check of consistency we do a weighted average of the daily background rate from panels (A) through (C), assigning weights by the number of days given in each panel. The result is an average
background rate of 1.3 earthquakes/day/km for the combined data set; divided by the number of mainshocks this gives a background rate of 0.0068 earthquakes/day/km (M≥2) in real time. This predicts that the composite figure in panel (D) should be best fit with b = 1.3 x 25 = 32.5, which agrees well with the b=33 solved for in the inversion. The goodness of fit is again verified by a low correlation coefficient for the data residuals (r = 0.03) and a Kolmogorov-Smirnoff test, both of which indicate that the null hypothesis that the underlying aftershock decay rate remains constant with time cannot be disproved at 95% and 65% confidence. The Kolmogorov-Smirnoff test is done by comparing the data to Monte Carlo simulated data drawn from the combination of a pure inverse law and a constant background distribution.

Most importantly, our data shows no evidence that the near field aftershocks transition to a steeper decay rate with time, as might be expected if there were a temporal transition from dynamic to static stress change triggering.
Supplementary Figure 5

Distance from the mainshock vs. aftershock linear density for un relocated catalog data from Japan (JMA catalog, 1997-2003) excluding earthquakes deeper than 30 km (A) and Northern California (B), excluding Long Valley caldera (ANSS catalog, 1984-2004). Distances are measured between hypocenters. Aftershocks used in Northern California are M>2, in Japan M>2.5. Aftershocks occur within the first 30 minutes of their respective mainshocks. (We use the first 30 minutes rather than the first 5 minutes to obtain a reasonable amount of data for these sequences). M 3-4 mainshocks are used. Like the 30 minute aftershock data in Figure 3 in the main text, we can fit the densities with inverse power laws (gray lines). Based on a sample of reported catalog errors, the average 3D location error (98% confidence) is about 2.5 km for Northern California and 2.2 km for Japan. Thus the shortest distance fit is 2 km for both data sets. The fits extend to 50 km. The Japan data set uses 7305 mainshocks and 366 aftershocks; the Northern California data set 3014 mainshocks and 369 aftershocks.

While the slope of the Japanese data is consistent with the slope we find in Southern California, the decay in Northern California is slightly steeper. Fitting the trend of background seismicity in Northern California using the same technique as in Figure 4 (main text) we find that the fault dimension, $D$, is about 0.3 smaller than in Southern California, indicating less complex fault structure. Assuming the same aftershock triggering mechanism in both locations, Equation 5 predicts that the exponent of the aftershock density decay in Northern California should be -1.7, which is within our measurement error bars.
Predictions for variation of aftershock density with distance by the static stress change plus rate and state friction aftershock triggering model with various parameters (grey lines), plotted against the 30 minute aftershocks of M 3—4 aftershocks from Supplementary Figure 3 (black circles). All of the rate and state curves are associated with summed least squared errors which are higher than the error for a simple power law fit and the data residuals for all of the curves have high correlation coefficients, indicating poor fit. Density is aggregated for the entire dataset so the 30 minute background seismicity is equivalent to the number of mainshocks used (2355) multiplied by the 30 minute window, i.e., 49 days.

(A) Best fitting rate and state curve when a wide range of parameters are tried (see Methods). Best fit parameters are $D = 0.1$, background rate ($c_1$) = 0.0039 earthquakes/km/30 minutes or 0.19 $M \geq 2$ earthquakes/km/day, and $c_2/M_0^{4\pi}$, or $A\sigma$, = 0.33. The negative slope at long distances is due to the very low value of $D$. (Any value of $D$ other than 1 will produce a trend in linear density in the farfield.) For a good fit to 601 data points at 98% confidence, the correlation coefficient of the data residuals, $r$, should be less than 0.1. For the fit in (A), $r = 0.45$. (For a pure inverse power law fit $r = 0.0079$). Since $D=0.1$ is physically extreme and has not been observed, we also look for best fit solutions with set values of $D$. (B) The best fit with $D = 1$. Background rate = 1.6 earthquakes day/km, $A\sigma = 0.83$. $r = 0.71$. (C) The best fit with $D = 2$. Background rate = 3.2 earthquakes/day/km and $A\sigma = 1.7$. $r = 0.59$. (D) The best fit with $D = 1$ and the background rate constrained to the estimated real earthquake rate at 16 km distance of 0.03 earthquakes/day/km. The best fit solution gives $A\sigma = 1.9$, $r = 0.12$, and a summed least squared error that is 23 times higher than the pure inverse power law fit.
The distance from the mainshock hypocenter vs. aftershock linear density for 200 aftershocks of M 2-3 mainshocks (black circles) and 200 aftershocks of M 3-4 mainshocks (gray triangles). Aftershocks are M > 2 aftershocks from the first 30 minutes of each aftershock sequence. When the same total number of aftershocks is used from each data set, the data points lie on top of each other. This indicates that the distribution of aftershock distances is independent of mainshock magnitude. Aftershock sequences of individual small mainshocks may appear compact simply because they have few aftershocks and the density decreases with distance, but there is in fact no correlation between the distance at which an aftershock occurs and the magnitude of its mainshock (correlation coefficient=-0.017, 290 data points between 0.2 and 16 km).
Demonstration that our results are not highly sensitive to our choices of density measurement method and smoothing parameter. We use the first 30 minutes of data from M 3-4 mainshocks, over the distance range of 0.2 - 16 km. The densities in (A) and (B) are measured with the nearest neighbor method, with k = 10 (A) and k = 20 (B), meaning that the densities are estimated by taking the inverses of the lengths of the boxes needed to contain groups of 10 and 20 neighboring earthquakes, respectively. Densities in (C) are measured by counting the number of earthquakes in bins of equal length on a logarithmic scale (e.g. from making a histogram on a logarithmic scale). The slopes in these three plots are the same as in figures in the main text, where we measured density using the nearest neighbor method with k = 1. An inverse power law remains an excellent fit to the data, and the best fit power law exponents are consistent with the value of $n$ found in the main text. The primary difference is that the data is more strongly smoothed. This allows for better visual presentation, but preserves less information, leading to a somewhat less well constrained fit for the inverse power law exponent.