Forecasting the evolution of seismicity in southern California:  
Animations built on earthquake stress transfer  

Shinji Toda,1 Ross S. Stein,2 Keith Richards-Dinger,3 and Serkan B. Bozkurt2  

Received 1 September 2004; revised 3 January 2005; accepted 4 February 2005; published 25 May 2005.  

[1] We develop a forecast model to reproduce the distribution of main shocks, aftershocks and surrounding seismicity observed during 1986–2003 in a 300 × 310 km area centered on the 1992 M = 7.3 Landers earthquake. To parse the catalog into frames with equal numbers of aftershocks, we animate seismicity in log time increments that lengthen after each main shock; this reveals aftershock zone migration, expansion, and densification. We implement a rate/state algorithm that incorporates the static stress transferred by each M ≥ 6 shock and then evolves. Coulomb stress changes amplify the background seismicity, so small stress changes produce large changes in seismicity rate in areas of high background seismicity. Similarly, seismicity rate declines in the stress shadows are evident only in areas with previously high seismicity rates. Thus a key constituent of the model is the background seismicity rate, which we smooth from 1981 to 1986 seismicity. The mean correlation coefficient between observed and predicted M ≥ 1.4 shocks (the minimum magnitude of completeness) is 0.52 for 1986–2003 and 0.63 for 1992–2003; a control standard aftershock model yields 0.54 and 0.52 for the same periods. Four M ≥ 6.0 shocks struck during the test period; three are located at sites where the expected seismicity rate falls above the 92 percentile, and one is located above the 75 percentile. The model thus reproduces much, but certainly not all, of the observed spatial and temporal seismicity, from which we infer that the decaying effect of stress transferred by successive main shocks influences seismicity for decades. Finally, we offer a M ≥ 5 earthquake forecast for 2005–2015, assigning probabilities to 324 10 × 10 km cells.  


1. Motivation and Goals  

[2] There is growing evidence that large earthquakes can inhibit or promote failure on nearby faults for decades to centuries, and that the transfer of stress plays a governing role in this interaction [Harris, 1998; Stein, 1999; King and Cocco, 2000; Freed, 2005]. Some progress has also been achieved in explaining the time dependence of observed seismicity by Coulomb stress transfer in an elastic medium coupled with the concepts of Dieterich’s [1994] rate and state friction [Stein et al., 1997; Parsons et al., 1999; Toda and Stein, 2003]. Competing and complementary concepts of dynamic stress triggering [Kilb, 2003; Gomberg et al., 2003], pore fluid diffusion [Peltzer et al., 1996], and viscoelastic stress transfer [Pollitz et al., 2004] have also been explored to explain aftershock and post-seismic geodetic deformation, but here we will restrict ourselves to static stress changes coupled to rate/state friction.  

[3] Here we seek to explore how successive earthquakes redistribute stress, causing large, sudden seismicity rate changes that decay and are subsequently perturbed by other large shocks. To make the process more visible and accessible, we simulate the expected seismicity and compare it with the observations in several animations. To test the ability of the model to capture seismicity observations, we perform statistical analyses on the rate/state predictions and on two control standards. The principal constituents of the model are the background seismicity, the Coulomb stress changes calculated for each large main shock, and the decay constant of aftershocks. With these three elements we attempt to forecast the distribution of seismicity until the next large stress-perturbing main shock strikes. The results are met with some success, and so we suggest that the approach shows promise as a probabilistic forecasting tool.  

[4] At heart, we interpret our results to mean that earthquake stress changes do not simply turn on or off seismicity; rather, the background seismicity rate is enhanced by stress...
increases and suppressed by stress decreases. This, we believe, best explains why seismicity in stress trigger zones is often patchy or discontinuous; why seismicity rate declines in stress shadows are often subtle or absent, and why some aftershock zones expand, migrate or densify.

2. Observations

2.1. Study Site and Test Period

We chose a site with the greatest concentration of large shocks in California during the past 20 years, centered on the 1992 Landers rupture, with boundaries 33.2–36.0°N. latitude, 115.14–118.41°W longitude (Figure 1). Southern California seismicity relocated by Richards-Dinger and Shearer [2000] in this 300 \times 310 km area was updated, providing a catalog from 1 April 1981 through 31 December 2003, within which quarry blasts were systematically removed. To measure seismicity rates we count earthquakes irrespective of magnitude, and so the catalog must exclude earthquakes smaller than the magnitude of completeness \( M_c \) for the period of study. Using ZMAP [Wiemer, 2001], \( M_c = 1.4 \) since 1981 at 95% confidence (Figure 2a).

[6] We first display the evolution of seismicity from 1981 in linear 4-month increments (please see Animation 1). The background seismicity rate appears temporally stable and spatially nonuniform, with concentrations along the San Jacinto and parts of the San Andreas fault, and in the Coso volcanic center north of Ridgecrest. Superimposed on the background seismicity are aftershock sequences that decay rapidly with time. There are also changes in the rate of shocks 25–75 km from the fault ruptures. For example, compare frames 55 and 56 in Animation 1, the intervals before and after the \( M = 7.1 \) 1999 Hector Mine shock; the seismicity rate drops at one site off the Hector Mine rupture and increases at another (Figure 3a), an observation that we seek to explain. However, the majority of aftershocks occur in the first frame after each main shock, so it is difficult to judge how aftershock zones grow, migrate, or change.

2.2. Log Time Seismicity Animation

[7] For an Omori aftershock decay, the number of shocks, \( N \), is given by \( N(t) = \frac{k}{(t + c)^p} \), where \( k \) and \( c \) are constants, and \( p \) is the decay exponent (generally near 1). After several hours following each main shock (once \( t \gg c \)), a similar number of aftershocks would be expected in each log time.

Figure 1. Study area, with active faults and \( M \geq 5.5 \) shocks occurring during 1 April 1981 to 30 December 2003. The eastern California shear zone is wider than portrayed and extends south to Indio.
interval. So in Animation 2, log time increments are used to track aftershock zone migration and expansion. At the time of each main shock, the interval resets to the shortest increment, and subsequently lengthens. We select 3 hours 8 min (3.13 hours, the time between the 1992 M = 7.3 Landers and M = 6.5 Big Bear events) as the first interval after each main shock. Subsequent intervals lengthen in half decibels (6.77, 21.42, 67.7, 214.2, 677 hours, etc.), with the last interval truncated by the succeeding main shock, as shown in Figure 4a, and listed in Table 1. Please see Animation 2.

Among the key phenomena visible in Animation 2 we seek to explain are progressions and migrations of aftershock zones, and seismicity rate changes off the fault ruptures. Key examples include the westward expansion of the North Palm Springs aftershock zone (Figure 5a); and the north, east, and southward expansion of the Joshua Tree aftershock zone (Figure 5b). Within the first 3 hours of the Landers earthquake, off-fault aftershocks occur in isolated sites in the southwest, northwest, and northeast (Figure 5c), where most seismicity will later be concentrated (Figure 5d). So, rather than a migration of aftershocks away from the Landers rupture, the initial distribution becomes more dense. This means that slow processes such as viscoelastic rebound [Freed and Lin, 2002] or pore fluid diffusion [Peltzer et al., 1996] cannot explain the temporal evolution of the Landers aftershocks; viscoelastic rebound might enhance the aftershock rate by the transmission of stress from the lower to upper crust, but stress migration away from the main rupture is too slow to explain the immediate pattern of aftershocks.

Among the most enigmatic features in Animation 2 is the jet of seismicity extending from the Hector Mine shock 140 km to southwest, in frames 32–35 (Figure 6a). The locations of these aftershocks are incompatible with the Coulomb stress change for the Hector Mine earthquake alone, and also with aftershock locations expected because of rupture directivity, for which shocks would extend

Figure 2. Background seismicity (1 April 1981 to 7 May 1986). (a) The frequency-magnitude relation which obeys a power law for M ≥ 1.4 at 95% confidence. This is the minimum magnitude of completeness Mc. (b) Observed background seismicity and smoothed background seismicity used in the rate/state model.
Figure 3. Observed and predicted seismicity and stress in 8-month intervals before and after the 1999 Hector Mine earthquake. (a) Observed seismicity which drops in the rectangle and increases in the other polygons. (b) The Coulomb stress change associated with the (right) Landers and (left) Hector Mine earthquakes (see Figure 7 for source faults). Stress in all polygons is calculated to have risen because of Landers earthquake; the Hector Mine shock drops the stress in the rectangular region but raises it in the others. (c) Rate/state model which incorporates the stress changes of both earthquakes and predicts seismicity that is generally similar to that observed. The predicted number of shocks is also influenced by the background rate of seismicity.
Figure 4. Observed versus predicted seismicity in Animation 3. (a) Time increments for Animation 2 and 3, which lengthen after each main shock. The number of aftershocks is roughly constant with time after a main shock, but the number of background shocks increases. The last increment before each main shock is necessarily truncated. (b) Running linear correlation between the observed and expected seismicity in Animation 3. There are 14,880 regressed points in each frame, 565,440 in all. The control model has the same temporal aftershock decay, but the seismicity diminishes with the inverse squared distance, $d$, from the fault rupture. The rate/state model performs poorly between the Joshua Tree and Landers earthquakes, probably because the $2.5 \times 2.5$ km grid is too coarse to distinguish the Joshua Tree trigger and shadow zones. The background seismicity contributes only modestly to the correlation but is greatest for long time periods. (c) Observed versus predicted number of aftershocks in Animation 3. The first four frames after Landers are dashed because the catalog is incomplete.
beyond the tips of the rupture, particularly to the south [Gomberg et al., 2003; Kilb et al., 2000].

3. Forecast Model

3.1. Earthquake Sources

[10] The static Coulomb stress change \( \Delta \text{CFF} \) caused by a main shock, with simplifying assumptions to account for pore pressure effects [King et al., 1994] is

\[
\Delta \text{CFF} = \Delta \tau + \mu' \Delta \sigma_n
\]  

(1)

where \( \Delta \tau \) is the shear stress change on a given fault plane (positive in the direction of fault slip), \( \Delta \sigma_n \) is the fault-normal stress change (positive when unclamped), and \( \mu' \) is the effective coefficient of friction.

[11] Stress changes associated with the largest earthquakes in the catalog govern the forecast model. Two earthquakes are included that struck before the 1981 start of the catalog because their stress effects are calculated to influence subsequent seismicity for 10–20 years: For the 31 May 1975 M = 4.8 Galway Lake shock [Hill and Beeby, 1977; Lindh et al., 1978] and the 15 March 1979 M = 5.5 Homestead Valley shock, [King et al., 1988; Stein and Lisowski, 1983], we use single rectangular slip patches.

The remaining earthquakes (whose epicenters are shown in Figure 1 and source faults in Figure 7) occurred during the test period: 8 July 1986 M = 6.0 North Palm Springs: 1 patch [Pacheco and Nabelek, 1988]; 23 April 1992 M = 6.2 Joshua Tree: 1 patch [Ammon et al., 1993; Savage et al., 1993]; 28 June 1992 M = 7.3 Landers: 41 patches, simplified from Wald and Heaton [1994], with Iron Ridge and Argos Mountain faults [Hudnut et al., 1993; Peltzer et al., 1998] added; 28 June 1992 M = 6.5 Big Bear: 2 faults with 10 patches [from Jones and Hough, 1995, Figure 12]; and 16 October 1999 M = 7.1 Hector Mine: 28 patches, simplified from Ji et al., 2002a, 2002b. The 1987 M = 5.9 Whittier Narrows and 1991 M = 5.8 Sierra Madre earthquakes are not included because these M < 6 events lie at the western boundary of the test area. The stresses produced by these earthquakes are calculated at 120 \( \times \) 124 elements spaced 2.5 \( \times \) 2.5 km apart. In retrospect, a finer grid would have been preferable, since the rupture lengths of the two M = 6 main shocks are just 8 km long.

3.2. Coulomb Parameters

[12] We use an effective coefficient of friction, \( \mu' = 0.4 \), assumed to be constant for all faults, and a shear modulus of
3.2 × 10^5 bars, in a uniform elastic half-space with Poisson’s ratio of 0.25. The maximum Coulomb stress change over the seismogenic depth of 5–12 km (which we sample at 7 and 11 km) is calculated on the assumption that seismicity will occur at the location and depth where the stress is most increased toward failure. We calculate Coulomb stress changes on optimally oriented vertical strike-slip faults [after King et al., 1994]. Aspects of the optimal-orientation assumption have been validated by near-fault stress rotations observed at Landers [Hauksson, 1994; Wiemer et al., 2002] and by the role of the regional stress in controlling tidal earthquake triggering [Tanaka et al., 2004]. Stress changes are calculated in a regional stress field dominated by a 100-bar horizontal compressional stress oriented N7°E (Figure 7). The orientation is based on the average compression direction in the study area [Hardebeck and Hauksson, 2001; Wiemer et al., 2002; Townend and Zoback, 2004], and its magnitude is chosen to reproduce the Landers near-fault principal stress rotations. Plots of the Coulomb stress change associated with the five main shocks are shown in Figure 7.

[13] The Coulomb stress changes for the Landers and Hector Mine shocks are shown with seismicity before and after the Hector Mine shock in Figure 3b. It appears that

![Figure 5](image-url)
the seismicity rate drops where the Hector Mine shock decreases the Coulomb stress (rectangular box), and the seismicity rate climbs where the Hector Mine shock further increases the Coulomb stress (the two other polygons). Such post-Hector Mine changes are only recognizable where the Landers earthquake had previously increased the Coulomb stress, causing, in our interpretation, a high seismicity rate. Similar seismicity rate drops observations were reported for the 1997 Kagoshima doublet by Toda and Stein [2003] and statistically validated by Woessner et al. [2004]. As in the Landers-Hector Mine case, the Kagoshima seismicity rate drop associated with the second main shock is apparent only where the first main shock increased the seismicity rate.

### 3.3. Rate/State Stress Transfer Model

[14] To consider successive stress changes associated with multiple main shocks, we use the expression for seismicity rate $R$ as a function of the state variable $\gamma$ under a tectonic secular shear stressing rate $\tau_r$ from Dieterich [1994]. Under constant shear stressing rate, the state variable reaches the steady state and is expressed as

$$\gamma_0 = \frac{1}{\tau_r} \quad (2)$$

At steady state, the seismicity rate $R$ is equivalent to the background rate $r$ because $R$ is calculated from

$$R = \frac{r}{\gamma \tau_r} \quad (3)$$

In the absence of a stress perturbation, the seismicity rate is assumed constant. We index the state variable $\gamma$ with time. If an earthquake strikes, it imposes a sudden stress step $\Delta CFF$, and the state variable $\gamma_{n-1}$ changes to a new value $\gamma_n$

$$\gamma_n = \gamma_{n-1} \exp \left( \frac{-\Delta CFF}{A \sigma} \right) \quad (4)$$

---

**Figure 6.** Aftershocks several days after the 1999 Hector Mine earthquake. (a) Within 30 hours after the earthquake, a lineament or jet of aftershocks extends for 140 km southwest of the main shock. (b) Fading effects of the Landers, Big Bear, and North Palm Springs earthquakes, amplified by a Hector Mine stress triggering lobe (seen in Figure 7e), which explains this feature.
where $A\sigma$ is a constitutive parameter times the effective normal stress, set here to be 0.5 bars [Toda and Stein, 2003]. To seek the seismicity at the time of the stress step, we substitute the new state variable in (4). In rate/state friction there is a nonlinear dependence of the time to instability on stress change. A stress increase on a fault causes $\gamma$ to drop, so the fault slips at a higher rate, yielding a higher rate of seismicity. Conversely, a sudden stress drop causes $\gamma$ to

---

**Figure 7.** Stress changes associated with the five M $\geq$ 6 earthquakes in the study area. The maximum Coulomb stress change on optimally oriented vertical strike-slip faults at depths of 7 and 11 km is plotted.
The duration of the transient is inversely proportional to the stressing rate. Given sufficient time (e.g., decades to centuries), the effect of all but the largest stress changes disappears on all but the most slowly stressed faults. A key feature of rate/state stress transfer is that the value of \( \gamma \) before each shock plays a profound role on the role of the stress change on seismicity: the higher the rate of seismicity at the time of a stress increase, the more strongly the seismicity rate will be amplified by the stress change. Further, the effect of each earthquake in a series continues to affect \( \gamma \), and all are used to calculate its current state.

### 3.4. Background Seismicity Rate Matrix

[15] We seek to use the largest number of earthquakes to estimate the background seismicity for the rate/state model, an optimization trading off the minimum magnitude of completeness against time period. Intervals starting before 1981 have a higher \( M_c \) and thus yield a smaller number of total shocks, and so we selected the period, 1 April 1981 to 7 May 1986, ending before the 1986 \( M = 6.0 \) North Palm Springs earthquake (Figure 2b). This leaves an 18-year test interval, 1986–2004. Because such a brief 5-year period is used to define an average long-term rate, we smooth the interval, 1986–2004. Because such a brief 5-year period is used, we adopt \( t_a = 52 \) years (Figure 8a). For the Big Bear earthquake, we isolated a 5.5 × 15 km area centered on the Big Bear aftershocks, and \( t_a = 7 \) years (Figure 8b). For Hector Mine, we used the entire 5 km × 50 km along-fault aftershock zone, with the pre-Landers period, 1981–1982 to calculate the background rate; \( t_a = 66 \) years. We also examined the 20 September 1995 moment magnitude \( M = 5.8 \) Ridgecrest earthquake, which struck in a highly seismic area south of the Coso geothermal field (Figure 1). We calculate \( t_a \) within a 10 × 6 km box (omitting the interval between the 17 August 1995 \( M = 5.5 \) and 20 September 1995 \( M = 5.8 \) Ridgecrest shocks from the background or aftershock period; Figure 1). The decay rate abruptly increases 3 months into sequence, when a \( M = 5.2 \) shock occurs. We find that \( t_a = 17–40 \) years depending on whether or not the first 3 months are included (Table 2). [18] From this exercise we infer that stressing rate is likely related to the proximity to the major transform systems, the San Andreas and eastern California shear zone (Figure 1), with aftershock duration lengthening with distance from the transform, inversely proportional to the presumed fault stressing rate, consistent with equation (6). For comparison, Parsons et al. [2000] found \( t_a \sim 25 \) years for 12 \( M \geq 6.7 \) North Anatolian earthquakes, and \( t_a = 7–11 \) years for 100 \( M \geq 7 \) global events [Parsons, 2002].

### 3.5. Estimated Aftershock Durations

[16] In the rate/state stress transfer model, the fault stressing rate \( \dot{\gamma} \) governs the longevity of the stress effects on seismicity. The stressing rate is related to the aftershock duration \( t_a \) and \( \dot{\gamma} \) [Dieterich, 1994; Dieterich and Kilgore, 1996]:

\[
t_a = \frac{\dot{\gamma}}{\dot{\gamma}}
\]

We thus estimate \( t_a \) for the principal main shocks, permitting equation (5) to be solved for each time frame. We measure the background rate in the area of most abundant aftershocks (within 3 km of the fault rupture), and estimate the time until the aftershock rate will return to this rate [Toda and Stein, 2002]. We find \( t_a \) ranges from 7 to 66 years (Table 2), with durations generally increasing with distance from the San Andreas fault or the eastern California shear zone. Because these estimates are imprecise, we adopt \( t_a = 20 \) years for Animation 3, although slightly better correlations between the observed and predicted seismicity are obtained with a spatially variable \( t_a \).

### 3.6. Predicted Seismicity Animation

[19] The relationship between the observed seismicity, the Coulomb stress, and the expected seismicity density is best understood from Figure 3. The modeled seismicity is strongly influenced by the most recent Coulomb stress changes, amplified by the background seismicity (Figure 3c, left), but the decaying effect of all past \( M \geq 6 \) earthquakes, also influences subsequent seismicity (Figure 3c, right).

[20] The observed and expected seismicity is shown in Animation 3. The expansion of the aftershock zones of the North Palm Springs (frames 2–8) and Joshua Tree (frames 12–17) shocks resembles the rate/state model (Figures 5a and 5b). In both cases, the aftershock zone expands into sites of high background rate (lower right-hand panel in

<table>
<thead>
<tr>
<th>Earthquake</th>
<th>Estimated Aftershock Duration, ( t_a ), years</th>
<th>Distance From San Andreas Fault, km</th>
<th>Distance From Eastern California Shear Zone, km</th>
</tr>
</thead>
<tbody>
<tr>
<td>1992 ( M = 6.5 ) Big Bear</td>
<td>7</td>
<td>10–20</td>
<td>40–50</td>
</tr>
<tr>
<td>1987 ( M = 6.6 ) Superstition Hills</td>
<td>15</td>
<td>20–30</td>
<td>–</td>
</tr>
<tr>
<td>1992 ( M = 7.3 ) Landers</td>
<td>25–50</td>
<td>50–70</td>
<td>0–20</td>
</tr>
<tr>
<td>1999 ( M = 7.1 ) Hector Mine</td>
<td>66</td>
<td>70–80</td>
<td>20–30</td>
</tr>
<tr>
<td>1995 ( M = 5.8 ) Ridgecrest</td>
<td>17–30</td>
<td>140</td>
<td>0–10</td>
</tr>
</tbody>
</table>
Animation 3), where the Coulomb stress changes most strongly increase the modeled seismicity rate. The majority of immediate aftershocks of Landers, both those along the rupture and those more than 25 km from it, lie in areas of expected high seismicity rate (frame 19 and Figure 5c). Two weeks into the Landers-Big Bear aftershock sequence (frame 24 and Figure 5d), clusters of aftershocks appear where Coulomb stress increases coincide with areas of high background seismicity (near Barstow and Hector Mine, and along the northernmost San Jacinto fault and the site of the North Palm Springs earthquake).

Figure 8. Aftershock durations $t_a$ for two main shocks listed in Table 2. Only seismicity within 5 km of the (a) Landers and (b) Big Bear rupture zones is used; the resulting catalog subsets are complete to $M \geq 2.2$ (shocks at $M \geq 2.5$, there are 48 background and $t_a = 34$ years for Landers). (top) Seismicity rate in linear time. The background rate is estimated for the entire 1981–1992 period before each earthquake. (bottom) Seismicity rate fitted with a modified Omori decay. The aftershock duration is estimated from the intersection of the fitted decay curve with background rate. It is evident that Landers aftershocks persist longer than those for Big Bear. The apparent increase in seismicity rate before each main shock in Figure 8 (top) is an artifact of the time bins.

Animation 3), where the Coulomb stress changes most strongly increase the modeled seismicity rate. The majority of immediate aftershocks of Landers, both those along the rupture and those more than 25 km from it, lie in areas of expected high seismicity rate (frame 19 and Figure 5c). Two weeks into the Landers-Big Bear aftershock sequence (frame 24 and Figure 5d), clusters of aftershocks appear where Coulomb stress increases coincide with areas of high background seismicity (near Barstow and Hector Mine, and along the northernmost San Jacinto fault and the site of the North Palm Springs earthquake).

Figure 8. Aftershock durations $t_a$ for two main shocks listed in Table 2. Only seismicity within 5 km of the (a) Landers and (b) Big Bear rupture zones is used; the resulting catalog subsets are complete to $M \geq 2.2$ (shocks at $M \geq 2.5$, there are 48 background and $t_a = 34$ years for Landers). (top) Seismicity rate in linear time. The background rate is estimated for the entire 1981–1992 period before each earthquake. (bottom) Seismicity rate fitted with a modified Omori decay. The aftershock duration is estimated from the intersection of the fitted decay curve with background rate. It is evident that Landers aftershocks persist longer than those for Big Bear. The apparent increase in seismicity rate before each main shock in Figure 8 (top) is an artifact of the time bins.

The previously described jet of shocks extending 140 km southwest of the Hector Mine fault (Figure 6a) corresponds to sites of expected seismicity caused by the interaction of the Hector Mine shock with the decaying effects of the Landers, Big Bear, and North Palm Springs main shocks in regions of high background seismicity (Figure 6b). This pattern would not be predicted by the Hector Mine earthquake stress changes alone, nor would it emerge if the Hector and Landers stress changes were simply added.

The effect of Coulomb stress shadows can be seen by the absence of earthquakes in regions that formerly were seismically active. The predicted seismicity rate at the end of Animation 3 (frame 39) is shown in Figure 9. Comparison of the prediction with the background rate reveals several sites, marked A-D, in which the background rates are high but the modeled rates for frame 39 are low because of the prolonged effect of the stress shadows on seismicity, in accord with the observed seismicity.

It is important to identify the several cases where the model fails to explain observed seismicity rate increases. We
expect a seismicity rate increase near Ridgecrest (Figure 1) of only about 4% (Animation 3, frames 19–29), but a factor of 10 increase is observed. This increase might be explained by dynamic Coulomb stress changes, which are sensitive to the northwestward rupture propagation of the Landers main shock \cite{Gomberg et al., 2003; Kilb et al., 2000}, to local pore fluid effects near the active Coso volcanic and geothermal field \cite{Beeler et al., 2000; Peltzer et al., 1996}, or perhaps to locally anomalous rate/state parameters, such as $A_s$. There is also a prominent lobe of seismicity east of the Hector Mine rupture (frames 34–38, Animation 3) that could only be modeled if there were a gap in the adjacent portion of the Hector Mine fault slip, for which we have no evidence.

4. Validation of the Forecast Model

4.1. Statistical Test for Small Shocks

\cite{24} We seek to assess whether there is a causal relationship between the rate/state stress transfer and the observed seismicity. Here we calculate the spatial correlation coefficient between observed and expected seismicity for all frames of Animation 3. Model successes (where we predict and observe many shocks, or where we predict and observe few), and failures (where we predict many and see none, or predict none and see many) are captured by the regression and reflected in the correlation coefficient. To treat the observations and model similarly, we first smooth the observed seismicity and the calculated Coulomb stress changes with the same 5-km Gaussian filter used on the background seismicity; if we did not do this, the observed seismicity would typically be more confined than the calculated seismicity, such as along a fault rupture. We do not smooth the stress changes for the smaller North Palm Springs and Joshua Tree shocks, because the trigger and shadow zones would be averaged, forfeiting the ability to discriminate between areas of enhanced and reduced seismicity near the main shocks (Figure 7). For the test, we remove the first hour of aftershocks after each main shock except for Big Bear, because the seismic network detection is saturated during this period, causing underreporting of earthquakes. For example, there are no M < 3 shocks in the catalog during the first 3 hours after Landers. (For Joshua Tree, the removal excludes 3 M $\geq$ 4.0 shocks; for Landers, it excludes 3 M $\geq$ 4.0 shocks, 1 M $\geq$ 4.5 shock, 1 M $\geq$ 5.0 shock and 3 M $\geq$ 5.5 shocks; for Hector Mine, 10 M $\geq$ 4.0 shocks, 8 M $\geq$ 4.5 shocks, 1 M $\geq$ 5.0 shock and 1 M $\geq$ 5.5 shock.) However, all recorded shocks, including those in the first hour, are nevertheless shown in the animations. \cite{25} The resulting correlation coefficients are shown in Figure 4b; they average 0.52 for the entire animation, and 0.63 for the period after Landers. Thus the model can explain about one half to two thirds of the observed seismicity distribution. Frame 33, with the highest correlation coefficient, is shown in Figure 6b. The model results can be compared with two control standards (Figure 4b). First, the effect of the background seismicity can be isolated by setting all Coulomb stress changes to zero, yielding an average correlation coefficient of 0.16. The background provides a poor fit immediately after each earthquake, which is dominated by aftershocks, but gradually improves with time, as the aftershock rate fades and the time intervals lengthen.
A more comprehensive control standard is next used in which the number of aftershocks is expected to diminish with the inverse of the squared distance, \( d \), from the fault rupture plane. Because aftershocks are most numerous along the fault rupture, the average correlation coefficient for the \( 1/d^2 \) decay model is equivalent to the rate/state Coulomb model, yielding 0.54 for the entire period. However, unlike the rate/state model, the control does not improve after Landers (0.52). Thus the rate/state model performs better than an empirical aftershock decay model for the three larger earthquakes, but more poorly for North Palm Springs and Joshua Tree. This is presumably because so few of the cells are affected by the Coulomb stress changes for the small shocks, and because the cells smear out Coulomb stress trigger and shadow lobes.

A remarkable feature of the model is evident in Figure 4c, which shows the seismicity rate as a function of time. The peak expected seismicity rates for the M = 6.0 North Palm Springs and M = 7.1 Hector Mine shocks are roughly the same, despite the 45-fold difference in their seismic moments. This is because the background seismicity rate near the San Andreas, where the North Palms Springs shock occurred, is 100 times greater than in the eastern Mojave, where the Hector Mine shock struck (Figure 2b). This expectation is largely borne out by the observed seismicity (Figure 4c). If the background seismicity rate did not influence aftershock production, the two observed decays would look quite different. The model does not, however, capture the longer duration of Hector Mine aftershocks in comparison to North Palms Springs, which we attribute to our use of a uniform \( t_a \) of 20 years, shorter than the \( t_a \) we estimated for Hector Mine (66 years). Since the plot uses log time frames, a typical \( 1/t \) decay appears as a constant slope. In the rate/state model, this slope eventually merges into the background rate of 1.0 at times much larger than \( t_a \). The predicted departure from a \( 1/t \) decay is seen about 8–10 frames after each main shock.

Unlike the correlation coefficients, the fit of the observed to modeled seismicity rates in Figure 4c is better for the two smaller shocks than for the larger shocks. For Landers and Hector Mine, the model predicts too high a rate of aftershocks in the initial 2–4 frames (3 hours to 2 days) after each main shock. This arises because \( M_c \gg 1.4 \) for the first several frames (dashed blue line in Figure 4c). For example, the smallest shock in the first frame after Landers is M = 3.0, resulting in a 1–2 order of magnitude disparity. The modeled rate is also too low for the period of about 1 day to 1 year after Landers and Hector Mine, to which several factors could contribute as follows: (1) because of the short averaging period or a locally higher \( M_c \), our background seismicity rate could be too low in the eastern California shear zone; (2) the 20-year aftershock duration underestimates the longevity of aftershocks, since 25–66 years is more appropriate for Landers and Hector Mine shocks; and (3) we neglect secondary aftershocks (aftershocks of aftershocks) [Ogata, 1988], which would increase the total and prolong the sequence.

4.2. Practical Test for Large Shocks

Irrespective of our statistical test on small shocks, for a forecast method to be societally useful, it must have some predictive power for large earthquakes. Our \( M \geq 6 \) earthquake sample is woefully inadequate for statistical testing, but one can ask instead if the \( M \geq 6 \) main shocks struck at sites where the seismicity rate was expected to be high. In all cases, the epicenter of the future main shock falls where the expected seismicity rate lies above 75% of the cells; for two shocks, the epicenters fall above the 98 percentile; and the average is 93% (Figure 10). One should bear in mind that we are only considering the frames immediately preceding the time when \( M \geq 6 \) shock struck, and we are not claiming that a \( M \geq 6 \) had a high probability of occurrence during any of these frames. Rather, on average, the \( M \geq 6 \) shocks struck in the 7% of the cells with the highest expected seismicity rates at the time of the occurrence.

The Hector Mine epicenter, which lies near the edge of a lobe of high expected seismicity rate, is least consistent with the model. However, the rate/state model fits the epicenter better than by the Landers stress changes alone (compare Figure 7c and Figure 10d). The lobe is a product of the stress change and the background seismicity, and its southern boundary is shifted 4.5 km south of the Coulomb stress lobe reported by Parsons and Dreger [2000] and Harris and Simpson [2002], Felzer et al. [2002], Harris and Simpson [2002], and Kilb [2003] have investigated other possible influences on the location of the Hector Mine shock, including secondary aftershock triggering and dynamic Coulomb stresses.

5. Forecast Model Limitations and Caveats

5.1. Smoothed Background and Aftershock Seismicity

We make an artificial distinction between background seismicity and aftershock seismicity; in fact there should be a continuum between the two, since the background is itself composed of aftershocks and stress effects of past earthquakes. Ideally, this could be overcome by averaging seismicity rates over centuries, or at least a period much longer than the aftershock duration, but this is not possible because all catalogs become more incomplete as one goes back in time, reducing the total number of included shocks.

We are also compelled to smooth the seismicity so that it can be correlated with the stress changes. Small earthquakes are a point process and so the smoothing removes information and introduces artifacts (e.g., smoothing attributes much of the San Andreas fault seismicity to several kilometers of the adjacent crust). The correlation coefficients soon after each main shock are highly dependent on the seismicity closest to the source faults where this smoothing is most influential, and where the Coulomb stress change is least well determined because it depends on the detailed fault slip [Steacv et al., 2004]. One could argue that the observed earthquakes should not be smoothed.

5.2. Spatially Variable Parameters Treated as Uniform

Independent evidence based on observed strain rates suggests that the stressing rate is not uniform across the study area; it is highest on the San Andreas, lower in the eastern California shear zone, and lower still elsewhere in the Mojave block [Savage et al., 2001]. One could create a stressing rate matrix with values proportional to observed strain rates, or alternatively, inversely proportional to after-
Figure 10. Expected distribution of earthquakes in the last frame of Animation 3 before each main shock: (a) 1992 Joshua Tree, (b) 1992 Landers, (c) 1992 Big Bear, and (d) 1999 Hector Mine. The frame number and duration are shown in the bottom left, and the epicenter of the succeeding main shock is shown as an open star. A histogram of the cell frequency as a function of the expected seismicity rate is shown below each panel. The expected seismicity rate at the 2.5 × 2.5 km cell corresponding to the four epicenters is shown as a vertical line in each histogram; the percentage of cells with a seismicity rate lower than the epicenter is indicated. In all cases, the epicenter struck at a location with an expected rate higher than 75% of the cells. In three cases it is >90%, and for two it is >98%. The peak in Figures 10a and 10c is shifted toward higher expected seismicity rates because of the longer frame duration; the percentile range stems from the 2.5 × 2.5 km cell size.
shock duration. Similarly, we treat the friction coefficient as constant, whereas it is probably lowest on the San Andreas [Townend and Zoback, 2004] and highest on immature faults in the Mojave block [Parsons et al., 1999]. We also assume a uniform regional stress tensor, whereas it, too, varies by about 10^6 over the study area [Hardebeck and Hauksson, 2001; Townend and Zoback, 2004]. We also ignore altogether stress changes on reverse faults, some of which are active in the study area [Hauksson, 1994]. Since the optimal orientations are a function of the regional stress, the friction coefficient, and the receiver fault planes, a spatially variable optimal orientation matrix could overcome these limitations.

5.3. Neglect of Inelastic Processes

Although we include the time dependence inherent in rate/state decay, we ignore viscoelastic relaxation [Freed and Lin, 2002; Pollitz et al., 2001], possible postseismic creep [Peltzer et al., 2001], and pore fluid [Cocco and Rice, 2002; Jonsson et al., 2003; Peltzer et al., 1998] or gas diffusion. Viscous relaxation has a timescale on the order of the aftershock duration (decades); fluid diffusion is probably quite brief (months), and creep could take any form. Inclusion of these processes would likely cause the aftershock zones to further diffuse outward and possibly diverge from a 1/t decay rate with time. We have argued, however, that aftershocks during the first 3 hours of Landers strike in almost all localities that later become active (Figures 5c and 5d; or Animation 3, frames 19 and 29), including sites 40–200 km away, which cannot be explained by any of these slow, time-dependent processes. This conclusion is in accord with Jonsson et al. [2003] but differs from that of Miller et al. [2004].

5.4. Neglect of Permanent Probability Changes

Elsewhere, we have attempted to incorporate the permanent effect of the stress changes [Stein et al., 1997; Toda et al., 1998; Toda and Stein, 2002;], in which the interevent time of a large earthquake is changed by the ratio of the stress change to the stressing rate (also termed the clock advance or delay by Gomberg et al. [1998]). However, because the modeled rate/state transient processes dominate over the two-decade span we study, here we seek to avoid additional assumptions about characteristic earthquakes and their the interevent times and variability.

6. Conclusions

6.1. Influence of Background Seismicity on Stress Trigger Zones and Shadows

An important criticism of Coulomb stress modeling has been that aftershocks do not fill in all stress triggering lobes, and that seismicity rate declines are not always apparent in Coulomb stress shadows [Marsan, 2003; Felzer
and Brodsky, 2005). Here we have argued that small stress increases are amplified by high background seismicity rates. This, we believe, explains why several sites, such as Big Bear, Barstow, and Hector Mine, became active immediately after the Landers earthquake despite experiencing only small static stress increases. We also make the corollary argument that large stress increases have little or no effect on seismicity at locations where the background rate is low, such as east of the Hector Mine shock, or in the western Mojave Desert surrounding Edwards (Figure 1 and frame 1 of Animation 3). The ability to observe seismicity rate declines is hampered by low background rates, because the rate must be high enough before the perturbing earthquake to measure a decrease. Thus the seismicity rate decrease in the Coulomb stress shadow of one main shock is only visible where the seismicity rate was first increased by a preceding main shock (Figure 3). Perhaps the most striking example of the influence of the background rate is the similar aftershock productivity for the $M = 6.0$ North Palm Springs and $M = 7.1$ Hector Mine shocks. This observation is independent of the spatial distribution (in other words, of trigger and shadow zones), and in our judgment is best explained by the much higher background rate along the San Andreas fault than in the eastern Mojave desert.

[37] The role of background seismicity, does not, however, explain the dramatic increase of seismicity in the Ridgecrest-Coso area (Figure 1) after the 1992 Landers earthquake, because the increase was much larger than forecast by the rate/state model. Thus other processes that we have not included, such as dynamic Coulomb stresses, shaking or pore fluid effects, also influence seismicity and should be incorporated for a more comprehensive forecast model.

### 6.2. Role of Past Main Shocks on Current Seismicity

[18] The log time animations (Animation 2 and Animation 3) provide a means to see more details of the aftershock triggering process than would otherwise be apparent, including migration and expansion of the aftershock zone in the first few hours to days, and surprising alignments of seismicity (Figures 3a, 5a–5d, and 6a) that in some cases persist for years. We have argued that these features result from the superposition of successive main shock stress changes, the effects of which decay with time. This is perhaps most striking in Figure 5d and Figure 6b. Even though the background seismicity amplifies the stress effects, the seismicity evolves over the 18-year test period, such that the final state of predicted seismicity differs markedly from the pattern due to the background alone. This is most evident by comparing the main and reference boxes in frame 39 of Animation 3. If our approach is valid, then seismicity models that consider only the stress changes of the most recent main shock, or models in which the stress changes of a set of main shocks are simply added, will be inadequate.

### 6.3. $M \geq 5$ Earthquake Forecast for 2005–2015

[19] Finally, we offer a forecast for $M \geq 5$ earthquakes during the next decade, 1 January 2005 to 1 January 2015. We restrict the forecast to the $180 \times 180$ km area in which the earthquake stress changes are greatest, sampling in $10 \times 10$ km cells. The predicted number of $M \geq 1.4$ earthquakes is scaled to $M \geq 5$ events from the observed 1981–2003 frequency-magnitude relation calculated by maximum likelihood within the forecast area by ZMAP [Wiemer, 2001], for which $b = 1.0 \pm 0.02$. The predicted earthquake number, $N$, is related to the probability, $P$, by $P = 1 - \exp(-N)$. The resulting probabilities are shown in Figure 11; the peak probability is 15% and the net probability for the entire area is 92%. (For the period 2005–2010, the peak probability is 9%, and the net probability is 73%.) The 2005–2015 probability in the $25 \times 30$ km area in the southwest corner is 82%. The probability on the Hector Mine rupture has faded, and only where the Hector Mine and Landers triggering zones overlap does the probability rise above 5%. The highest probabilities are found where stresses from three or more main shocks overlap in an area of high background rate. The Poisson formula assumes that the probability in each cell is independent. Strictly speaking, we would need to recalculate the probability after the occurrence of a large shock to reflect the stress transfer, but we will neglect this for the purposes of a blind test.

### 6.4. Forecast Model in Retrospect

[40] How successful is the rate/state model for forecasting seismicity? First we ask the reader to look at the correspondence between observed and forecast seismicity in Animation 3. Next, we show that the average correlation coefficient for the large shocks (for periods after Landers) is encouraging (0.63), and superior to the correlation of a geometric aftershock decay control (0.52). Finally, we find that the four main shocks struck in regions with forecast seismicity rates above, on average, the 90th percentile. With that said, there are periods after Landers when the correlation is only slightly higher than the control, and for the entire period before Landers, the control performs better. While this is a far cry from an earthquake prediction, it is, perhaps, on the road to the more useful and accurate earthquake forecasts that we all seek.

### Acknowledgments

We thank Thomas Dewez for preliminary versions of the animations; Fred Politz, Matthew Gerstenberger, Ruth Harris, John Townend, Debi Kilb, and Andy Freed for thoughtful reviews; Jim Dieterich for guidance; and The Geothermal Program Office of the U.S. Navy for financial support. This is a contribution to the Regional Earthquake Likelihood Models (RELM) program of the Southern California Earthquake Center (contribution 815).

### References


Felzer, K. R., T. W. Becker, R. E. Abercrombie, G. Ekstrom, and J. R. Rice (2002), Triggering of the 1999 Ms, 7.1 Hector Mine earthquake by after-