Self-similar earthquake triggering, Båth's law, and foreshock/aftershock magnitudes: Simulations, theory, and results for southern California

Peter M. Shearer[1]

Received 27 October 2011; revised 3 May 2012; accepted 3 May 2012; published 21 June 2012.

[1] Båth's law, the observation that the largest aftershock is, on average, 1.2 magnitudes smaller than its main shock, independent of main shock size, suggests some degree of self-similarity in earthquake triggering. This behavior can largely be explained with triggering models in which the increased triggering caused by larger magnitude events is exactly compensated for by their decreased numbers, and these models can account for many features of real seismicity catalogs. The Båth's law magnitude difference of 1.2 places a useful constraint on aftershock productivity in these models. A more general test of triggering self-similarity is to plot foreshock and aftershock rates as a function of magnitude m relative to the main shock magnitude, m_{max}, of the largest event in the sequence. Both computer simulations and theory show that these dN/dm curves should be nearly coincident, regardless of main shock magnitude. The aftershock dN/dm curves have the same Gutenberg-Richter b-value as the underlying distribution, but the foreshock dN/dm curves have the same b-value only for foreshock magnitudes less than about $m_{max} - 3$. For larger foreshock values, the dN/dm curve flattens and converges with the aftershock dN/dm curve at $m = m_{max}$. This effect can explain observations of anomalously low b-values in some foreshock sequences and the decrease in apparent aftershock to foreshock ratios for small magnitude main shocks. Observed apparent foreshock and aftershock dN/dm curves for events close in space and time to M 2.5 to 5.5 main shocks in southern California appear roughly self-similar, but differ from triggering simulations is several key respects: (1) the aftershock b-values are significantly lower than that of the complete catalog, (2) the number of aftershocks is too large to be consistent with Båth's law, and (3) the foreshock-to-aftershock ratio is too large to be consistent with Båth's law. These observations indicate for southern California that triggering self-similarity is not obeyed for these small main shocks or that the space/time clustering is not primarily caused by earthquake-to-earthquake triggering.

Citation: Shearer, P. M. (2012), Self-similar earthquake triggering, Båth's law, and foreshock/aftershock magnitudes: Simulations, theory, and results for southern California, *J. Geophys. Res.*, *117*, B06310, doi:10.1029/2011JB008957.

1. Introduction

[2] Earthquakes are observed to cluster strongly in time and space. Aftershocks are one of the most obvious examples of clustering behavior in earthquake catalogs. They occur close to their triggering main shocks and the aftershock rate generally decays with time following the power law relation known as Omori's Law. Large earthquakes generate many more aftershocks than small earthquakes. Aftershock sizes, however, appear to follow the same Gutenberg-Richter (G-R) power law as the rest of the earthquake catalog, i.e., there are many more small aftershocks than large aftershocks. These aspects of aftershock behavior have been incorporated into triggering models, in which the probability of earthquake occurrence at any given time and place is related to the past history of nearby earthquakes. The most well-known of these models is the Epidemic Type Aftershock-Sequences or ETAS model (for reviews, see *Ogata* [1999] and *Helmstetter and Sornette* [2002a]), in which every earthquake, no matter how small, increases the probability of future nearby events. There is no requirement in these models that the aftershocks be smaller than the triggering event. In some cases the main shock (the largest event in the sequence) is triggered by a smaller preceding event, which is then termed a foreshock. In addition, aftershocks can themselves trigger additional aftershocks,

[1]Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California, USA.

Corresponding author: P. M. Shearer, Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093-0225, USA. (pshearer@ucsd.edu)

0148-0227/12/2011JB008957

generating a triggering cascade of activity. However, because aftershock sequences do not last forever or cause a runaway explosion of seismicity, the aftershock productivity of main shocks must be limited. One of the best-known constraints on aftershock productivity is Båth's Law [*Båth*, 1965], which states that the largest aftershock is, on average, 1.2 magnitude units smaller than the main shock. This limits the average total number of aftershocks because a too-large sequence of aftershocks would be expected, by G-R, to generate a larger earthquake than Båth's Law predicts. In addition, the fact that Båth's Law is invariant with respect to main shock size implies a certain degree of self-similarity in the triggering process. In this paper, self-similar triggering is defined to be that which produces constant average *relative* magnitude distributions for foreshocks, main shocks, and aftershocks in triggered sequences, that is, that a M 5 main shock produces the same average number of $3 \geq M < 5$ aftershocks as a M 6 main shock produces $4 \geq M < 6$ aftershocks.

[3] Triggering models can explain many aspects of the clustering seen in real earthquake catalogs and a common approach is to find the best-fitting set of ETAS-model-like parameters to fit a given catalog. However, correlation does not require causality and it is important to also examine if there are aspects of catalog seismicity that are not well explained by triggering models, which could indicate underlying physical processes that are driving earthquake occurrence, rather than seismicity being entirely driven by random background events and earthquake-to-earthquake triggering. Here I focus on the implications of Båth's Law for triggering models and earthquake catalogs, in particular on its constraints on aftershock productivity and on its predictions for the relative numbers of foreshocks and aftershocks of given magnitudes. These results can be tested through comparisons to the over 400,000 earthquakes in the *Lin et al.* [2007] relocated catalog for southern California. Reliable statistics are difficult to obtain for large earthquakes because they occur so infrequently. Therefore, following *Felzer and Brodsky* [2006] and *Shearer and Lin* [2009] I will focus on the much more numerous earthquakes of M 1.5 to 5.5 in the catalog, but will not, however, examine the apparent distance dependence of aftershocks, which has been the source of recent controversy [e.g., *Felzer and Brodsky*, 2006; *Gomberg and Felzer*, 2008; *Richards-Dinger et al.*, 2010]. Rather my focus will be on the implications of Båth's law for aftershock productivity and whether triggering models that satisfy Båth's law produce foreshock and aftershock behavior that agree with southern California seismicity. This is not entirely a new topic and many previous authors have related Båth's Law to triggering models [e.g., *Felzer et al.*, 2002; *Helmstetter and Sornette*, 2003c; *Sornette and Werner*, 2005a; *Saichev and Sornette*, 2005]. Here I will attempt to cite the relevant aspects of this work while making clear the new features of my analysis, which include documentation using differential magnitude versus frequency plots of the expected behavior of foreshock and aftershock magnitudes in synthetic catalogs versus those observed in southern California.

[4] An important conclusion is that while observed foreshock and aftershock magnitudes appear roughly self-similar, they deviate from self-similar triggering models in several key aspects: (1) the 'aftershock' b-values are significantly lower than that of the complete catalog, (2) the number of aftershocks is too large to be consistent with Båth's law, and (3) the foreshock-to-aftershock ratio is too large to be consistent with Båth's law. These results indicate that the space/time clustering associated with M 1.5 to 5.5 earthquakes in southern California is not easily explained with simple earthquake-to-earthquake triggering models and may be more indicative of swarm-like processes.

2. Earthquake Triggering Models

[5] There is an extensive literature on various aspects of earthquake triggering models [e.g., *Ogata*, 1999; *Helmstetter and Sornette*, 2002b; *Felzer et al.*, 2002; *Helmstetter et al.*, 2005]. Here I will review some of the key results required in this paper. A common assumption in these models is that the magnitude of earthquakes, whether occurring as background or triggered events, is a random variable drawn from the Gutenberg-Richter (G-R) distribution. In this case,

$$N(\geq m) = 10^{a-bm} = 10^{a}10^{-bm} \tag{1}$$

where m is the magnitude, a is related to the total number of earthquakes, and b (the b-value) determines the relative number of large quakes compared to small quakes and is generally observed to lie between 0.8 and 1.2. If we define minimum and maximum magnitudes m_1 and m_2 then

$$\begin{aligned} N(m_1, m_2) &= 10^{a}\left(10^{-bm_1} - 10^{-bm_2}\right) \\ &\approx 10^{a-bm_1} \quad \text{for large } m_2 \end{aligned} \tag{2}$$

Note that the number of events per magnitude is given by

$$\frac{dN}{dm} = b\ln(10)10^{a}10^{-bm} \tag{3}$$

In practice, for computer simulations a random event magnitude, m_r, can be computed as

$$m_r = m_1 - \log_{10}x_r \tag{4}$$

where x_r is randomly and uniformly distributed between $10^{m_1-m_2}$ and 1.

[6] A second key aspect of many triggering models is that the average number of direct (first generation) aftershocks, N_{as1} following an event of magnitude m follows a productivity law

$$N_{\mathrm{as1}} = Q10^{\alpha(m-m_1)} \tag{5}$$

where m_1 is the minimum magnitude earthquake that triggers other earthquakes, Q is an aftershock productivity parameter (denoted k by *Sornette and Werner* [2005b]), and α is a parameter that determines the rate of increase in the number of aftershocks observed for larger main shock magnitudes. This type of triggering model assumes that every aftershock has a single parent and is sometime termed the branching model, in which each event independently triggers its own aftershock chain [e.g., *Kagan*, 1991; *Felzer et al.*, 2002]. In contrast, in the ETAS modeling approach of

Ogata [1998] the probability of aftershock triggering is governed by the sum of all previous activity. As discussed in *Sornette and Werner* [2005b], both approaches produce seismicity catalogs with the same statistical properties, although the branching model used here is much faster computationally.

[7] Note that equation (5) provides the average number of direct aftershocks for a given magnitude main shock and typically will have non-integer values. In computer simulations, the actual number of direct aftershocks for a specific event can be computed as a random deviate drawn from a Poisson distribution of mean N_{as1}. The total number of aftershocks generated by a single initiating event will in general be larger than the number of direct aftershocks because each aftershock is capable of spawning its own aftershocks, which in turn can generate additional aftershocks. However, provided the aftershock productivity parameter Q is not set too large for a given m_1 and m_2, this process will eventually converge to a finite total number of aftershocks.

[8] The appropriate value for α to best match real earthquake catalogs has been a source of some discussion [*Helmstetter and Sornette*, 2002b; *Helmstetter*, 2003], but recent work seems to be converging on $\alpha \approx 1$ [*Felzer et al.*, 2004; *Helmstetter et al.*, 2005]. The case of $\alpha = b$ has received special attention because it produces nearly self-similar triggering behavior and yields relatively simple equations to predict the total number of aftershocks that are produced [*Agnew and Jones*, 1991; *Felzer et al.*, 2002, 2004; *Helmstetter et al.*, 2005]. In this case

$$N_{\text{as1}} = Q10^{b(m-m_1)} \tag{6}$$

Now consider a background catalog with a G-R distribution. The average total number of first-generation aftershocks can be expressed as

$$\begin{aligned} N_{\text{tot1}} &= \int_{m_1}^{m_2} Q10^{b(m-m_1)} b\ln(10)10^{a}10^{-bm}dm \\ &= Qb\ln(10)10^{a}10^{-bm_1}\int_{m_1}^{m_2} dm \\ &= Qb\ln(10)(m_2-m_1)10^{a-bm_1} \\ &\approx Qb\ln(10)(m_2-m_1)N_{\text{back}} \quad \text{for large } m_2 \end{aligned} \tag{7}$$

where N_{back} is the number of background events in the catalog. Let us define

$$r = Qb\ln(10)(m_2-m_1) \tag{8}$$

which gives the ratio of the average number of first generation aftershocks to the number of background events (simplified in this case by using the approximation for large m_2). This parameter is denoted n by *Helmstetter and Sornette* [2003b], *Helmstetter et al.* [2003], and *Sornette and Werner* [2005a, 2005b], who call it the branching ratio. Because the aftershocks are also drawn randomly from the same G-R relation as the background catalog, the total average number of 2nd generation aftershocks is simply

$$\begin{aligned} N_{\text{tot2}} &= r\,N_{\text{tot1}} \\ &= r^2\,N_{\text{back}} \end{aligned} \tag{9}$$

For $r < 1$ (the subcritical regime of *Helmstetter and Sornette* [2002b]; the aftershock numbers can go to infinity otherwise), we can sum the geometrical series to obtain

$$N_{\text{tot}} = \frac{1}{1-r}N_{\text{back}} \tag{10}$$

where N_{tot} is the average total number in the complete catalog (background and all aftershocks). The fraction of triggered events in this catalog is simply the parameter r. Similar results for earthquake triggering are found in many previous papers (e.g., *Helmstetter et al.* [2005, equation 19], if one strips out the time dependencies). The result is very simple because setting $\alpha = b$ ensures that every magnitude interval always contributes equally to the aftershock production. That is, the decreased aftershock productivity with smaller magnitudes is exactly compensated for by their increase in numbers.

[9] Equation (10) gives results for an entire catalog with N_{back} background (untriggered) events that follow the G-R distribution. To apply this result to a single main shock, we note that the main shock will generate N_{as1} first-generation aftershocks, which also obey G-R. Thus, if we set the number of background quakes, N_{back} in equation (10) to the number of first-generation aftershocks, N_{as1}, in equation (6), we can obtain the average total number of aftershocks produced by a single main shock of magnitude m:

$$\begin{aligned} N_{\text{as}} &= \frac{1}{1-r}Q10^{b(m-m_1)} \\ &= \frac{r}{1-r}\,\frac{10^{b(m-m_1)}}{b\ln(10)(m_2-m_1)} \end{aligned} \tag{11}$$

where we have substituted for Q using (8).

[10] Note that many computer simulations of triggering assign random times to aftershocks based on Omori's law and sometimes also locations based on observed falloffs in aftershock densities with distance from the main shock. Because my focus here is on the frequency-magnitude distribution of triggered events, details of the time and distance dependence of triggering are largely ignored in both the simulations and the data, but will be the subject of future work.

3. Båth's Law

[11] A widely cited empirical constraint on aftershock behavior is provided by Båth's law [*Båth*, 1965], which states that the largest aftershock is, on average, 1.2 magnitude units smaller than the main shock. The average magnitude of the largest of N events with a G-R distribution is [e.g., *Helmstetter and Sornette*, 2003c]

$$\langle m_{\max}\rangle = m_1 - \int_0^1 \frac{N(1-x)^{N-1}\ln(x)}{b\ln(10)}dx \tag{12}$$

where m_1 is the minimum magnitude. For large N, a good approximation is provided by

$$\begin{aligned} \langle m_{\max}\rangle &= m_1 + \frac{\log_{10}N}{b} + \frac{\gamma}{b\ln 10} \\ &= m_1 + \frac{\log_{10}N}{b} + 0.2507 \end{aligned} \tag{13}$$

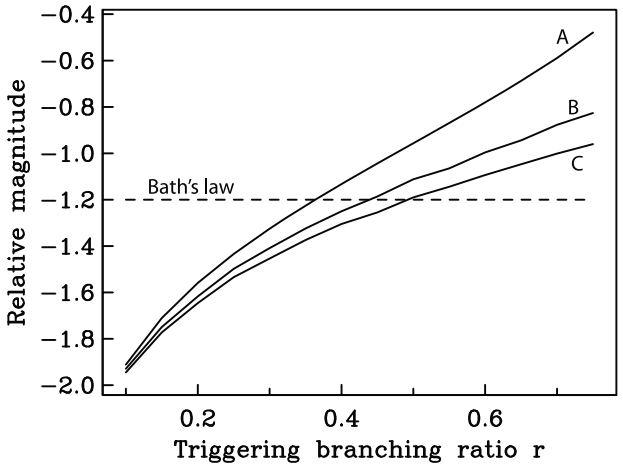

Figure 1. Average differential magnitude of the largest aftershock as a function of the branching ratio r from: (A) equation (14), (B) a numerical simulation of 10,000 M 5 main shocks, and (C) the same numerical simulation, but excluding sequences with aftershocks larger than the starting event. Båth's law, shown by the dashed line, states that the average magnitude of the largest aftershock is 1.2 units smaller than the main shock.

where γ is Euler's constant. Note that the expressions given in *Helmstetter and Sornette* [2003c] and *Sornette and Werner* [2005a] are off by 0.25 magnitude units because they neglect the final term. A derivation of (13) is contained in the Appendix.

[12] Thus the magnitude difference given by Båth's law could (naively) be expressed as

$$\langle \Delta m \rangle = \langle m_{\max} \rangle - m_{\mathrm{MS}} = m_1 + \frac{\log_{10} N_{as}}{b} + 0.2507 - m_{\mathrm{MS}} \quad (14)$$

where m_{MS} is the main shock magnitude and $\langle m_{\max} \rangle$ is the average value of the largest aftershock. Solving (14) for N_{as} and substituting into (11), we can obtain

$$r = \frac{B}{B+1}, \text{where B} = b \ln(10)(m_2 - m_1) 10^{-b(0.2507 - \langle \Delta m \rangle)} \quad (15)$$

For $b = 1$ and $\langle \Delta m \rangle = -1.2$, we have $B = 0.0816(m_2 - m_1)$. For $m_2 - m_1 = 6$, then $r = 0.329$. For $m_2 - m_1 = 8$, this gives $r = 0.395$. This relation (which assumes Båth's law and $\alpha = b$) predicts that the fraction of triggered events among the total events (background and triggered) is less than 50% unless $m_2 - m_1$ exceeds 12.4.

[13] However, for actual simulations and comparisons to earthquake catalogs equations (14) and (15) are inaccurate for several reasons. First, N_{as} in equation (11) is the *average* number of aftershocks produced by a main shock of given magnitude but the actual number of aftershocks for any given earthquake can be quite different. In practice, as recognized by *Helmstetter and Sornette* [2003c], most earthquakes produce fewer aftershocks than N_{as}, but a few generate many more because they happen to trigger one or more unusually large aftershocks. This causes the main shock averaged $\langle m_{\max} \rangle$ to be less than that predicted by equation (14) and the corresponding value of $\langle \Delta m \rangle$ to be less than the predicted value (i.e., larger $|\langle \Delta m \rangle|$ since Δm is defined to be negative). To illustrate this, Figure 1 plots $\langle \Delta m \rangle$ versus r for $m_1 = 0$, $m_2 = 7$, and $m_{\mathrm{MS}} = 5$. Line A shows the prediction of equation (14), compared to the results of a computer simulation of the random aftershock sequences of 10,000 $m_{\mathrm{MS}} = 5$ earthquakes (Line B). The results agree for $r < 0.1$ but the simulated catalog produces smaller $\langle \Delta m \rangle$ than equation (14) for increasing r values, and thus requires larger r values to satisfy Båth's law. In particular, to satisfy $\langle \Delta m \rangle = -1.2$, equation (15) gives $r = 0.36$, while the catalog simulation requires $r = 0.44$.

[14] Second, a fraction of the aftershock sequences will contain events that are larger than the initiating event, as also recognized by *Helmstetter and Sornette* [2003c]. This is expected from the G-R relation and is a common feature of these simulations. However, analyses of real earthquake catalogs often explicitly or implicitly require that the "main shock" be larger than all of the events in the sequence, rather than simply to be the first event. For example, consider a M 3 event that triggers a M 6 event, which then triggers a big aftershock sequence, the largest event of which is M 5. For testing Båth's law, this sequence would probably be counted as having $\Delta m = -1$ rather than $\Delta m = 3$. Requiring that the "main shock" be the largest event in the sequence thus introduces a bias that will decrease $\langle \Delta m \rangle$ by systematically removing positive Δm values. To show this, we repeat the simulation for the 10,000 $m = 5$ initiating events, but in cases where an aftershock is larger than M 5, we consider only the aftershock family of the larger event and compute Δm using the larger event magnitude. As shown in Figure 1, the resulting $\langle \Delta m \rangle$ as a function of r again agrees with the predictions of equation (14) for $r < 0.1$, but

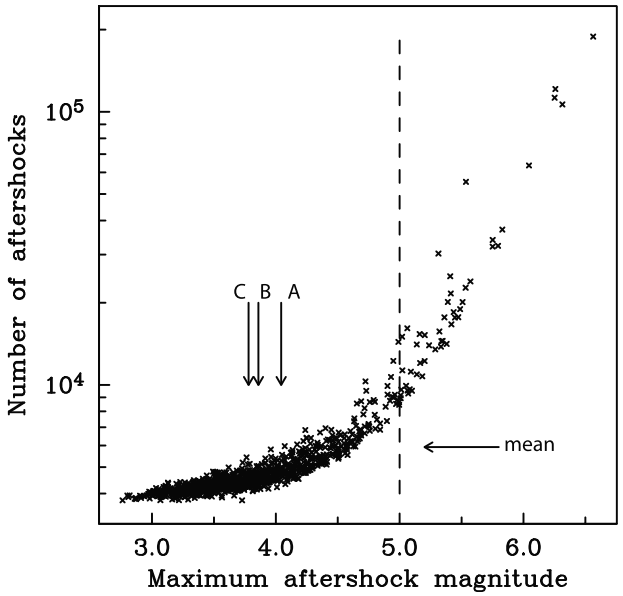

Figure 2. Number of aftershocks versus maximum aftershock magnitude for 1000 random aftershock sequences from a M 5 main shock (assuming $m_1 = 0$, $m_2 = 7$, and $r = 0.5$). The horizontal arrow shows the mean number of aftershocks. The vertical arrows A, B, and C, show different ways to estimate the average magnitude of the largest aftershock (see text).

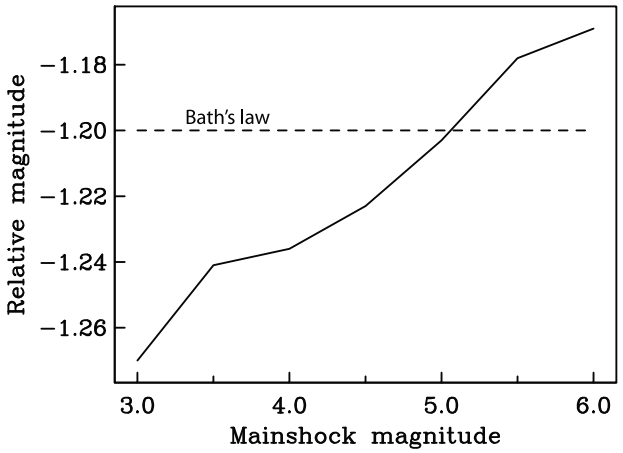

Figure 3. Average differential magnitude of the largest aftershock as a function of the main shock magnitude for random aftershock sequences, generated using $m_1 = 0$, $m_2 = 7$, and $r = 0.494$.

produces larger values for larger r values. In particular, the simulation requires $r = 0.49$ to satisfy $\langle \Delta m \rangle = -1.2$.

[15] To illustrate how the differences plotted in Figure 1 come about, Figure 2 plots the number of aftershocks, N, as a function of the magnitude of the largest aftershock, m_{max}, for 1000 random trials with $r = 0.5$, $m_1 = 0$, $m_2 = 7$, and $m_{MS} = 5$. The vertical arrows, labeled A, B, and C, correspond to the curves in Figure 1, and show (A) the predicted m_{max} from equation (14), (B) the average m_{max} over all the main shocks, and (C) the average m_{max} excluding the events larger than the main shock. The distribution of the results for the different main shocks is highly skewed in both N and m_{max}. A small fraction of runs have extremely large values of N, which occurs when they happen to trigger an unusually large aftershock. The average value of N is 5918 (approximated by equation (11) and shown by the horizontal arrow in Figure 2), which reflects the influence of these very large values of N and overestimates N for the vast majority of the runs. Thus, estimates of average m_{max} based on the average value of N are biased toward higher magnitudes and the true $\langle m_{max} \rangle$ is less than that predicted by equation (14). A further reduction in $\langle m_{max} \rangle$ is seen if sequences with an aftershock of larger magnitude than the main shock are excluded, i.e., those to the right of the vertical dashed line in Figure 2.

[16] These effects introduce a small amount of non-self-similarity to the predicted triggering results, even for $\alpha = b = 1$. This is illustrated in Figure 3, which plots $\langle \Delta m \rangle$ versus main shock magnitude for $m_{MS} = 3$ to 6, assuming r is fixed at 0.494, and performing 10,000 trials for each m_{MS} value (m_1 and m_2 remaining at 0 and 7). The simulation predicts that $\langle \Delta m \rangle$ increases with main shock magnitude, from about –1.27 at $m_{MS} = 3$ to –1.17 at $m_{MS} = 6$. To maintain a constant $\langle \Delta m \rangle$ in the simulated catalog would require slightly more productivity for the smaller events, i.e., a value of a slightly less than one. It is doubtful, however, that the small change in $\langle \Delta m \rangle$ apparent in Figure 2 could be seen in real earthquake catalogs.

[17] A possible source of bias in comparing simulations to real earthquake catalogs is the uncertainty associated with magnitude measurements. There will always be some amount of random measurement error in assigning magnitudes. Any random scatter in assigned magnitudes will tend to bias $\langle \Delta m \rangle$ estimates to larger values. This is illustrated in Figure 4, which plots $\langle \Delta m \rangle$ for normally distributed magnitude errors with standard deviation, σ, ranging from 0 to 0.5 magnitude units ($r = 0.494$, $m = 5$, $m_1 = 0$, $m_2 = 7$, 10,000 trials for each data point). The bias can be as much as 0.2 magnitude units for $\sigma = 0.5$. The appropriate value to use for σ will vary from catalog to catalog. For the Southern California Seismic Network (SCSN) magnitudes analyzed later in this paper, σ is estimated at 0.1 (E. Hauksson, personal communication, 2010), in which case the predicted bias is very small, and is neglected for the remainder of this paper.

[18] Throughout this section, I have used $m_1 = 0$ and $m_2 = 7$, for which $r = 0.5$ provides a reasonable fit to Båth's law. The corresponding value of aftershock productivity, Q, is 0.030. However, different values for m_1 and m_2 require different r values to provide the same fit. In particular, larger values of the minimum magnitude m_1 require lower values of r and increased values of Q to satisfy Båth's law. Here I use $m_2 = 7$ because this is close to the size of the largest events in the modern southern California earthquake catalog (the last M 8 event was in 1857). The appropriate value for the minimum magnitude m_1 has been the subject of some discussion [e.g., *Felzer et al.*, 2002; *Sornette and Werner*, 2005a, 2005b]. The exact choice of m_1 is not critical for this paper, and $m_1 = 0$ is chosen to be well below most catalog thresholds, while not excessively increasing the computation cost of the computer simulations. My intention is not to completely explore the model parameter space, but to find a reasonable set of parameters satisfying Båth's law that can be used to explore predicted aftershock and foreshock behavior in more detail.

4. Differential Magnitude Plots

[19] If we define the main shock as the largest event in a triggered sequence (rather than the initiating event), then

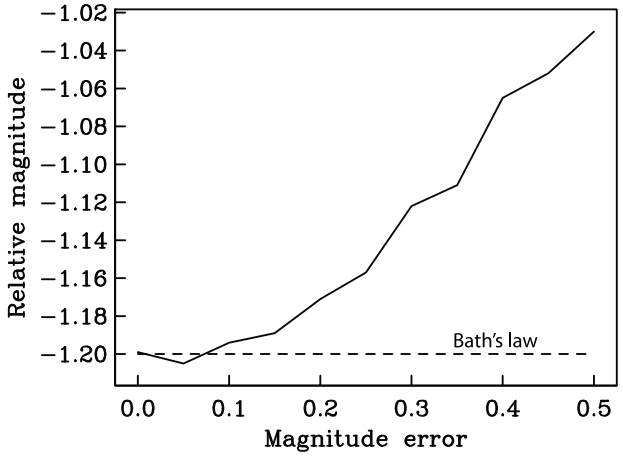

Figure 4. Average differential magnitude of the largest aftershock as a function of catalog magnitude error (standard deviation) for random aftershock sequences, generated using $m_1 = 0$, $m_2 = 7$, and $r = 0.492$.

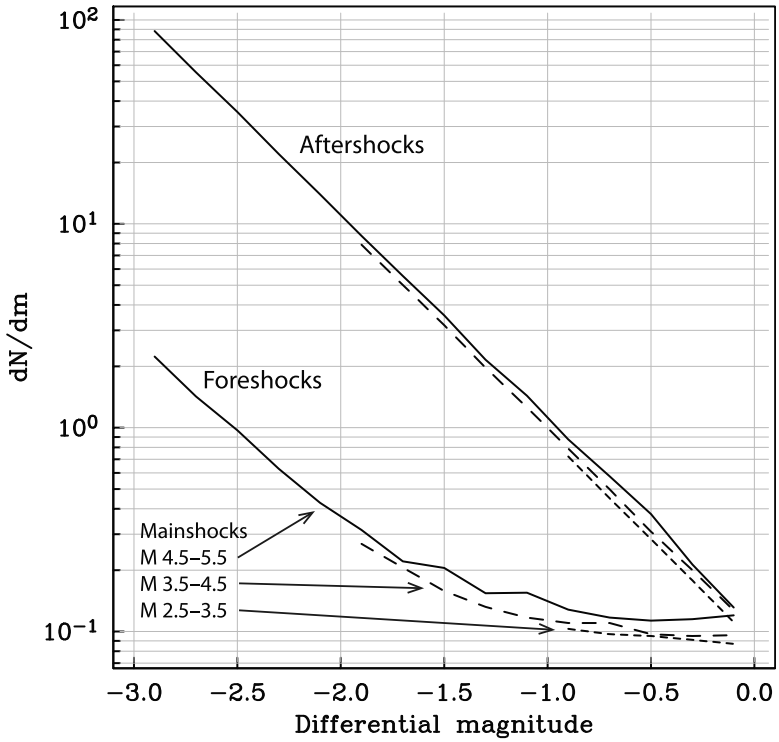

Figure 5. Average numbers of foreshocks and aftershocks as a function of differential magnitude with respect to the main shock (defined as the largest event in the sequence) as computed from a computer simulation of 10^8 randomly triggered sequences. The solid lines show results for M 4.5–5.5 main shocks, the long-dashed lines show results for M 3.5–4.5 main shocks, and the short-dashed lines show results for M 2.5–3.5 main shocks. The minimum catalog magnitude is 1.5, which limits the lower differential magnitude of some of the lines. Notice that the aftershock distribution is linear on a log scale, consistent with the model b-value of one, but that the foreshock distribution bends to join the aftershock distribution at zero differential magnitude. This explains why foreshock sequences are often observed to have anomalously low apparent b-values and why foreshock-to-aftershock ratios increase as the differential magnitude approaches zero.

both foreshocks and aftershocks are smaller than the main shock. Båth's Law is consistent with triggering self-similarity, as defined above, because the average magnitude difference of 1.2 between the main shock and its largest aftershock does not depend upon main shock magnitude. But self-similar triggering imposes a more general constraint on triggering behavior than fixing the average size of the largest aftershock. If aftershock generation is indeed self-similar with respect to magnitude, as discussed earlier, then the average number of both foreshocks and aftershocks within a given differential magnitude range should be the same regardless of the main shock magnitude. For example, although a M 5 main shock will generate many more aftershocks than a M 4 main shock above a given minimum aftershock magnitude, under self-similarity on average the M 5 main shock will generate the same number of aftershocks with magnitudes above 5–Δm as the M 4 generates with magnitudes above 4–Δm, where Δm is a fixed magnitude difference. A M 4 main shock should have the same average number of M 2–4 aftershocks as a M 5 main shock has of M 3–5 aftershocks.

[20] This provides a way to test for self-similarity by plotting the average number of foreshocks and aftershocks as a function of the magnitude difference $\Delta m = m - m_{MS}$, where m_{MS} is the main shock magnitude and m is the foreshock or aftershock magnitude. Figure 5 shows the results for a computer-generated set of triggered sequences, generated using $r = 0.5$, $m_1 = 0$, $m_2 = 7$, and 10^8 starting 'background' events (randomly distributed between M 0 and 7 assuming a b-value of one). In each sequence, the largest event is defined as the main shock, prior events are defined as foreshocks and later events as aftershocks. Aftershocks are assigned times based on the probabilities implied by Omori's law, thus some first generation aftershocks can occur after second or higher generation aftershocks. We count the number, N, of events as a function of Δm, using a magnitude binning interval of 0.2. We assume a minimum catalog magnitude of 1.5 (the triggering calculations are done using a minimum magnitude of 0.0, but we only output events of $m \geq 1.5$). We then compute average dN/dm curves for main shock magnitudes of 2.5–3.5, 3.5–4.5, and 4.5–5.5. As expected for triggering self-similarity, both the foreshock and aftershock dN/dm curves are nearly coincident as a function of main shock magnitude.

[21] The constant slopes of the aftershock dN/dm functions are consistent with the model b-value of one. However, the foreshock dN/dm curves are concave upward. Only for $\Delta M < -3$ do they have a slope of minus one. At larger

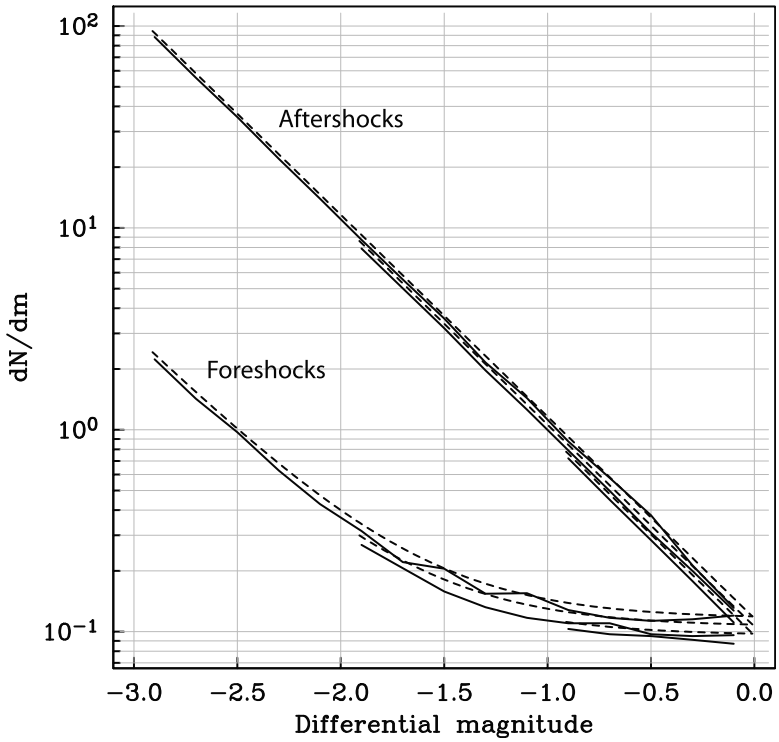

Figure 6. A comparison between the aftershock and foreshock *dN*/*dm* curves measured from a computer simulation (solid lines) with those predicted by analytical approximations (dashed lines).

relative magnitudes, they have shallower slopes and appear to intersect the aftershock curves at $\Delta M = 0$. This may seem surprising, but in fact it is a consequence of the main shock selection criteria, which defines the main shock as the largest event in the sequence, rather than the first event [*Knopoff et al.*, 1982; *Helmstetter and Sornette*, 2003a]. Appendix B explains this behavior in terms of the triggering theory presented earlier.

[22] The main difference in the slope of the foreshock and aftershock *dN*/*dm* curves is due to the fact that the initiating event is included in the foreshock distribution but is not included in the aftershock distribution. As several authors have noted [e.g., *Vere-Jones*, 1969; *Console et al.*, 2003], for Båth's law to hold, the main shock cannot be drawn from the same G-R distribution as the aftershocks. In particular, its average value is larger than expected purely from the largest event in a random G-R sequence. Thus, including the initiating event lowers the apparent *b*-value for foreshock sequences. As an aside, note that these relationships are easier to understand when presented in terms of *dN*/*dm* rather than *N*(*m*). A complication arises in *N*(*m*) curves when they approach a limiting upper magnitude. The upper magnitude limit causes them to curve downward in log plots and not plot as straight lines, even when their *b*-value is constant. In contrast, *dN*/*dm* curves with constant *b*-value remain straight on log plots even when they approach a fixed upper magnitude limit.

[23] Figure 6 plots the theoretical approximations given in Appendix B versus the computer-simulated results for 10^8 random sequences with $r = 0.5$, $m_1 = 0$, $m_2 = 7$, and target event magnitudes of 2.82, 3.82, and 4.82 (the average magnitudes in the 2.5–3.5, 3.5–4.5, and 4.5–5.5 bins, which of course have more events closer to their lower limits). Notice the generally good agreement between the approximations and the random realizations, except that the theory very slightly overpredicts *dN*/*dm.* Numerical simulations show that the approximations work best for small values of *r* and increasingly overpredict the actual numbers of aftershocks and foreshocks as *r* approaches one. This is probably because the theory does not fully account for the correlation between the number of events in the sequences and the largest aftershock, as discussed earlier and illustrated in Figure 2. Note that in general the behavior of the aftershock *dN*/*dm* curve for $m - m_{\max} > -1$ and the foreshock *dN*/*dm* curve for $m - m_{\max} > -2$ will be hard to resolve for individual main shocks because the expected total number of events is typically less than one. It is only by averaging the results over many main shocks that one can resolve these details.

[24] A number of studies have claimed to observe anomalous *b*-values in foreshock sequences [e.g., *Suyehiro*, 1966; *Papazachos*, 1975; *Smith*, 1981] with most studies finding lower *b*-values for foreshocks compared to aftershocks. However, *Knopoff et al.* [1982] proposed that a lower foreshock *b*-value is an expected artifact of the data selection procedure (i.e., assigning the largest event in the sequence as the main shock means the preceding events are no longer statistically independent of the main shock), an argument further developed by *Helmstetter and Sornette* [2003a]. My results support this interpretation and document the expected

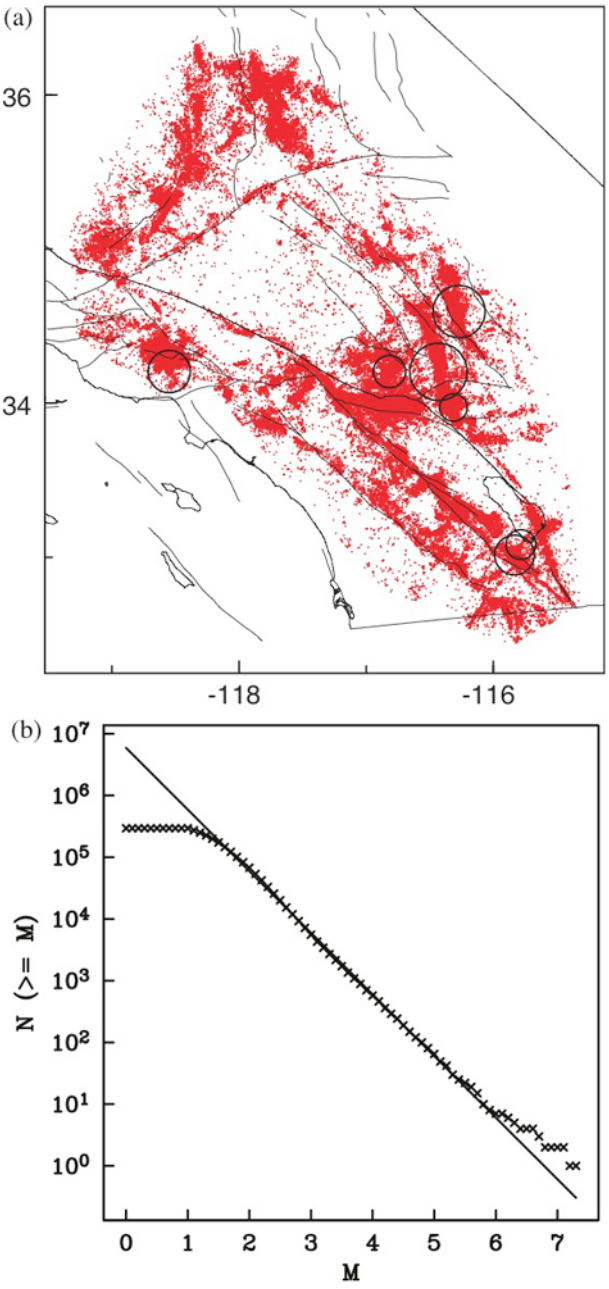

Figure 7. (a) Event locations from the LSH catalog of *Lin et al.* [2007], windowed to only include events with good station coverage. Events of $M > 6$ are shown as black circles. (b) The number of events as a function of minimum magnitude, plotted at 0.1 magnitude intervals. The best-fitting line for $2 \le M \le 5$ is plotted and has a slope of -1.00.

form of the magnitude distribution of foreshocks for the specific case of $\alpha = b$ triggering models.

5. Comparison to Southern California Seismicity

[25] A recent catalog [*Lin et al.*, 2007] for southern California computed using waveform cross correlation provides relative location accuracy of 100 m or less among nearby events. This catalog (here termed LSH) spans 1981 to 2005 and includes 433,166 events over a magnitude range from less than 1 to over 7. My event section procedure is the same as that used in *Shearer and Lin* [2009]. To obtain a more uniform data set, the catalog is windowed to include only events of $M \ge 1.5$ that are located inside the network and identified as local earthquakes by the network operators (i.e., excluding quarry blasts), reducing the catalog to 173,058 events (see Figure 7a). A plot of the number of events versus magnitude for this catalog (see Figure 7b) can be fit reasonably well with a constant b-value ($b = 1.00$) down to M 1.5, suggesting that the catalog is nearly complete down to that magnitude, although there are certainly at least some missing events near the lower limit (more about this later). A b-value of 1.0 agrees with a recent analysis of southern California seismicity by *Hutton et al.* [2010]. Because catalog completeness often suffers following major earthquakes owing to the high seismicity rate [e.g., *Kagan*, 2004], I exclude target events during certain specified time periods. Specifically, I exclude target earthquakes for 1 month following the 1987 M 6.2/6.6 Elmore Ranch/Superstition Hills and 1992 M 6.1 Joshua Tree earthquakes, 2 months following the 1994 M 6.7 Northridge earthquake, and 3 months following the 1992 M 7.3 Landers and 1999 M 7.1 Hector Mine earthquakes. This process should remove at least some of the coincidentally contemporaneous event pairs within major aftershock sequences, cited by *Richards-Dinger et al.* [2010] in their commentary on *Felzer and Brodsky* [2006].

[26] Next, I search through the catalog for target events in bins of M 2.5–3.5, 3.5–4.5, and 4.5–5.5, requiring that the target events have no larger event occurring within 20 km and 3 days before and 1/2 day after, resulting in 6185, 540, and 61 events in the three magnitude bins respectively. For each target event, I then count 'foreshocks' occurring within 5 km and 12 hours before the target event, and 'aftershocks' occurring within 5 km and 12 hours after the target event. This process does not, of course, capture all of the foreshocks and aftershocks, but rather is intended to obtain a sample of their relative numbers and magnitude dependence. By selecting relatively short distance and time cutoffs, the goal is to maximize the ratio of triggered earthquake activity to 'background' activity. Next, I bin the foreshocks and aftershocks in relative magnitude (Δm) intervals of 0.2 and compute the average value of dN/dm for each bin (this includes normalizing by the number of target events). The results are plotted in Figure 8.

[27] The aftershock dN/dm curves for the three different target event magnitudes are nearly coincident, consistent with self-similarity. However, the aftershock curves have a b-value of about 0.8 to 0.85, significantly less than that of the complete catalog (which is well fit with $b = 1.00$). I will discuss the implications of this difference later. The foreshock curves are more irregular, reflecting a relatively small number of events (especially for the M 3.5–4.5 and M 4.5–5.5 bins at small $|\Delta m|$ values. The M 3.5–4.5 and M 4.5–5.5 foreshock curves are roughly coincident, consistent with self-similarity in the foreshock behavior, but the M 2.5–3.5 curve is offset to significantly higher amplitudes. The M 2.5–3.5 results are defined by a large number of events, so this appears to be a robust departure from self-similar behavior. There are significantly more M 2–3 'foreshocks' to M 3 events than there are M 3–4 'foreshocks' to M 4 events. Note that none of the curves appear to flatten significantly

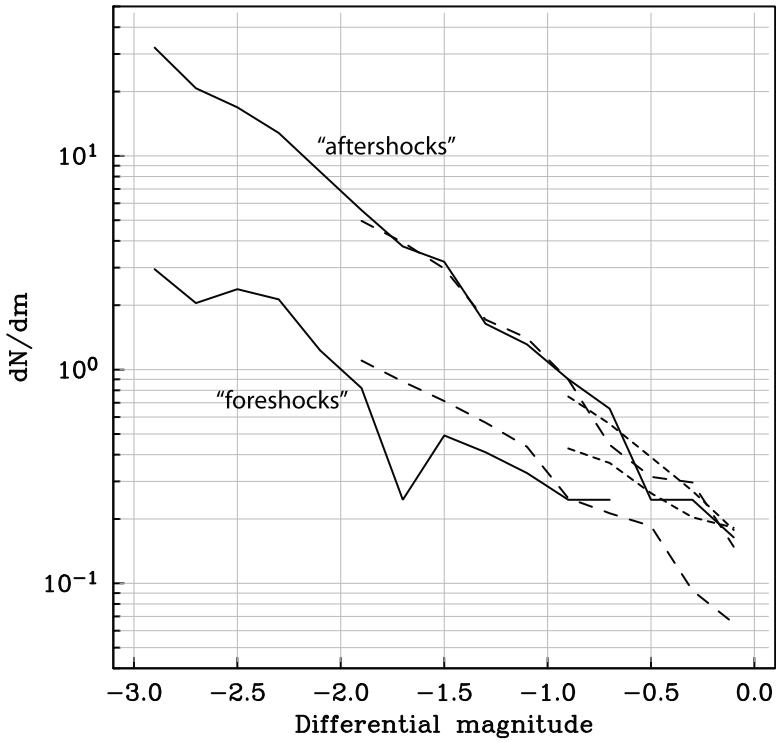

Figure 8. Average numbers of 'foreshocks' and 'aftershocks' as a function of differential magnitude relative to the 'main shock,' as measured from 1981–2005 seismicity in southern California. Events are counted that occur within 5 km and 12 hours of the assumed main shocks. The solid lines show results for M 4.5–5.5 main shocks, the long-dashed lines show results for M 3.5–4.5 main shocks, and the short-dashed lines show results for M 2.5–3.5 main shocks. The left edge of the foreshock lines is determined by the minimum catalog magnitude of 1.5. The M 4.5–5.5 foreshock line is incomplete near zero differential magnitude owing to a lack of events.

near their left edges, which might have been expected if catalog incompleteness was an important factor near the lower magnitude limit (M 1.5) of the catalog.

[28] To test the sensitivity of the results plotted in Figure 8 to the event selection criteria, I repeated the analyses using windows that are twice as strict. In this case, the target events are required to have no larger event occurring within 40 km and 6 days before and 1 day after, resulting in 4558, 509, and 60 events in the three magnitude bins respectively. For each target event, 'foreshocks' are counted occurring within 2.5 km and 6 hours before the target event, and 'aftershocks' occurring within 2.5 km and 6 hours after the target event. As one might expect, the tighter foreshock/aftershock window yields a proportionally smaller average number of foreshocks and aftershocks, and somewhat noisier foreshock curves when the total number of foreshocks is reduced to one or two within a differential magnitude bin. But the main features, including the aftershock b-value and the foreshock-to-aftershock ratio are similar to those seen in Figure 8 (see Figure S1 in the auxiliary material).[1]

[29] Figure 9a compares the synthetic results plotted in Figure 5 with the real southern California catalog results in Figure 8, and reveals some important differences. First, as mentioned above, the b-value for the data 'aftershocks' is about 0.8 to 0.85, significantly less than the value of 1.0 for the full southern California catalog. In contrast, the synthetic catalog has an aftershock b-value of 1.0, as is expected because all magnitudes were randomly assigned using b = 1.0. Second, the number of data aftershocks is comparable to the number of synthetic aftershocks for differential magnitudes between −1.5 and 0. This is surprising, given that the data are selected from a fairly tight space/time window around the main shock (5 km and 12 hours), thus counting what is presumably only a fraction of the total number of aftershocks, whereas the synthetics are computed from compete triggered sequences, i.e., *all* aftershocks are counted. Thus the number of 'aftershocks' (with magnitudes within 1.5 units of the main shock) in the data is much larger than might be expected from the triggering model. Third, the curvature of the foreshock curves seen in Figure 5 is only weakly suggested in Figure 9a. The reason for this is not clear, but the lack of many foreshock events at differential magnitudes between −1 and 0 limits the reliability of curvature estimates. Finally, the ratio of apparent foreshock-to-aftershock activity is higher for the real seismicity than for the synthetic catalog, which was produced with triggering parameters (r = 0.5, m_1 = 0, m_2 = 7) which reproduce Båth's law. The validity of Båth's law was recently verified for $M \geq 5.5$ earthquakes in California by *Shcherbakov and Turcotte* [2004] who obtained an average Δm value of −1.16.

[1] Auxiliary materials are available in the HTML. doi:10.1029/2011JB008957.

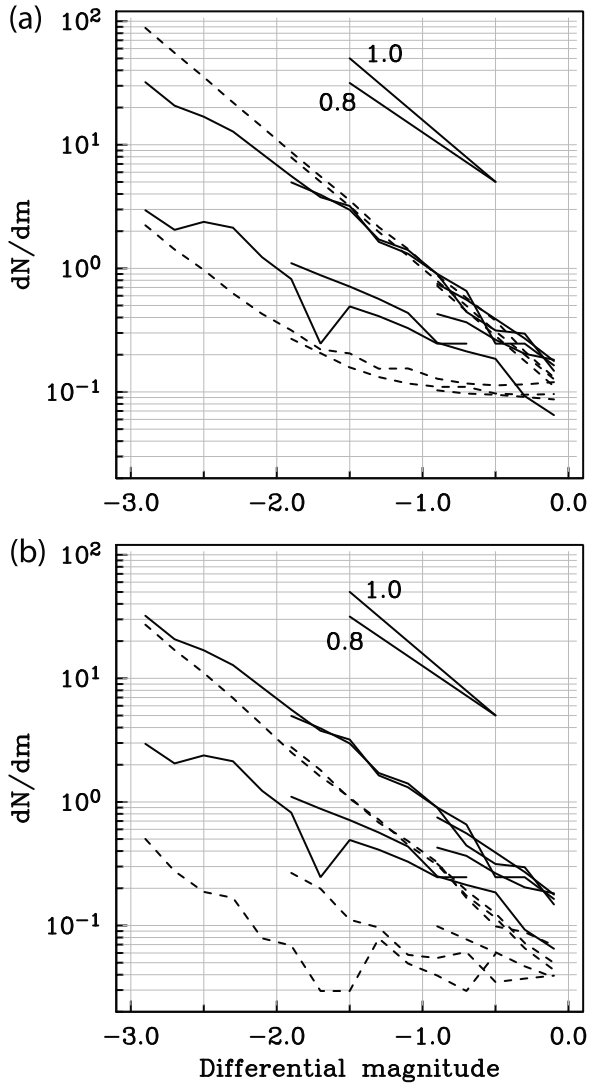

Figure 9. A comparison between the observed windowed numbers of foreshocks and aftershocks in southern California seismicity (solid lines) to model predictions (dashed lines) as computed using: (a) complete foreshock/aftershock counts from a triggering model that obeys Båth's law, and (b) windowed foreshock/aftershock counts (using the same space/time window as applied to the data), derived from a triggering model similar to that in Figure 9a, but which simulates the complete spatial and temporal distribution of events in southern California. For comparison, *b*-value slopes of 0.8 and 1.0 are also plotted.

[30] Could these differences be caused somehow by the fact that a space/time windowing scheme was applied to the data but not to the synthetics? To test this possibility, I generated a synthetic catalog with actual event times and locations as follows: 10 million background events of M 0 to 7 were assigned uniformly distributed random times over a 100 year period. Locations were randomly chosen from the actual data catalog, but random location scatter of 2 km was added to each event. Aftershocks were computed using the same model parameters as before ($r = 0.5$, $m_1 = 0$, $m_2 = 7$) but delay times were randomly assigned assuming Omori's Law with the rate given by $1/(t + c)$, where $c = 86$ s and with minimum and maximum delay times of 9 s and 1000 days. The aftershock locations were randomly assigned with respect to the main shock location using a linear decay in aftershock density with distance of $r^{-1.37}$, as suggested by the results of *Felzer and Brodsky* [2006] and *Brodsky* [2011], and minimum and maximum distances of 30 m and 100 km, respectively. This calculation is not intended to fully explore the range of possible models (tests of the range dependence of seismicity clustering will be the target of future work), but to perform a first-order test of the effects of windowing the data on our results.

[31] This catalog was then processed using exactly the same target-event selection criteria and foreshock/aftershock windowing method as was applied to the real data. The resulting catalog has about 680,000 events of $M \geq 1.5$ in 100 years, or about the same number of events per year as the data catalog (173,058 in 25 years). Figure 9b compares the resulting differential foreshock and aftershock curves to the data. As expected, the synthetic aftershock curves are shifted downward (by about a factor of three), consistent with the fact that the space/time windows are not capturing all the aftershocks. The curves are substantially below the corresponding data aftershock curves for most of the plotted differential magnitude range. The windowed synthetic aftershock curves have the same *b*-value (1.0) as the unwindowed synthetics. Thus, the change in *b*-value seen in the data 'aftershocks' compared to the overall catalog is unlikely to be caused by the windowing.

[32] The windowed synthetic foreshock curves are noisier than the aftershock curves, reflecting the smaller numbers of events. However, like the aftershock curves, they are shifted downward compared to the unwindowed synthetics, and the foreshock-to-aftershock ratio remains substantially lower than that observed for the data. Interestingly, the foreshock curves appear to split apart to some extent, with more foreshocks observed for the smaller target events. This suggests that the foreshock counts may be contaminated by unrelated 'background' seismicity that happens to fall in the space/time window around the main shock, but is not part of the triggered event sequence containing the main shock. This provides a possible explanation for the analogous splitting of the foreshock curves seen for the real catalog.

[33] The fact that the synthetic differential dN/dm curves plotted in Figure 9b are below the data curves suggests that a better fit might be obtained for a model with greater aftershock productivity (i.e., a higher branching ratio than 0.5). Indeed this is the case, as illustrated in Figure 10, which plots results for a branching ratio of $r = 0.8$ (still with $m_1 = 0$, $m_2 = 7$). The overall dN/dm levels and the ratio of foreshock-to-aftershock activity agree much better with the data. However, the increased model aftershock activity translates to a predicted largest aftershock about 0.9 units smaller than the main shock, rather than the generally accepted Båth's Law value of 1.2. This r value of 0.8 also predicts that over 25% of earthquakes are followed by an aftershock of larger magnitude, much more than is typically observed [e.g., *Jones*, 1985]. In addition, the synthetic aftershocks continue to have a *b*-value of one and thus cannot fit the $b \sim 0.8$ slope of the data dN/dm curves. Finally, note that the apparent foreshock-to-aftershock ratio for the M 2.5–3.5 main shocks is simply too large to explain with any likely amount of

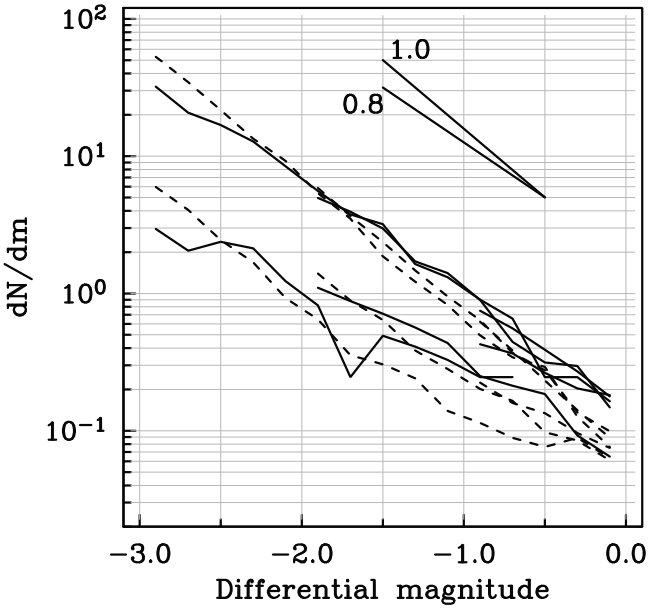

Figure 10. Similar to the data versus synthetics comparison of Figure 9b, except computed for a simulated catalog with much greater aftershock productivity than is consistent with Båth's law. Solid lines show observed windowed numbers of foreshocks and aftershocks in southern California seismicity; dashed lines show model results windowed the same as the data. For comparison, *b*-value slopes of 0.8 and 1.0 are also plotted.

increased triggering. Taken together, these results suggest that it is unlikely that the apparent foreshocks and aftershocks obtained from tight space/time windows around *M* 2.5 to 5.5 main shocks in southern California can be fully explained with self-similar triggering models of the type considered here.

6. Discussion

[34] Earthquake triggering models can explain many aspects of real earthquake catalogs and exploring the complexities of model behavior as a function of their parameterizations has become a field of study in itself. However, in seismology it is important to recognize that these models are idealizations, which do not always capture all the details in real earthquake catalogs. For example, aftershock productivity typically varies among different earthquakes, with some events generating much larger sequences than other events of the same magnitude, even without triggering an anomalously large aftershock. Swarms, in which activity starts, continues, and then stops, without an obvious initiating main shock are also difficult to explain with triggering models tuned to fit average aftershock behavior, such as those presented here. Allowing triggering model parameters to vary with space and time can fit some of this behavior, at the cost of increasing the model complexity and making the results specific to particular regions or times and less universally applicable. It is important to examine possible limitations to the triggering models because this can point to aspects of earthquake occurrence that reflect physical differences and possible driving mechanisms rather than seismicity being a completely random process.

[35] A fundamental distinguishing attribute of aftershock sequences is their time asymmetry—many more events occur following a large earthquake than occur as foreshocks before the earthquake. Because aftershock productivity decreases with main shock magnitude, aftershock sequences of single events are only obvious for large earthquakes. For example, a typical *M* 6 earthquake might generate 1000 aftershocks above magnitude 1.5, while a *M* 3 earthquake only generates one, and only about 10% of *M* 2 earthquakes generate even a single aftershock (above *M* 1.5). To obtain resolvable results for small events, it is thus necessary to average results from hundreds to thousands of small 'main shocks.' In addition, often these smaller events occur in areas of ongoing seismicity, motivating the application of narrow time windows to distinguish likely triggered events from background events. In southern California, this was the approach used by *Felzer and Brodsky* [2006] to identify likely aftershocks, and the approach used to resolve apparent foreshocks and aftershocks in Figure 8. However, for smaller main shocks the foreshock-to-aftershock ratio increases and the results become more time symmetric. For example, *Richards-Dinger et al.* [2010] note in their commentary on *Felzer and Brodsky* [2006] that 7134 southern California *M* 2–3 'main shocks' are preceded by 319 $M \geq 2$ 'foreshocks' and followed by 364 'aftershocks' within ± 5 minute windows.

[36] We have shown here that increased foreshock/aftershock time symmetry is expected from standard triggering models as the main shock magnitude approaches the minimum catalog magnitude. This is shown by the convergence of the foreshock and aftershock *dN*/*dm* curves in Figure 5. However, these same triggering models predict foreshock-to-aftershock ratios that are much less than that observed in southern California data for *M* 2.5–5.5 main shocks. This difference is most likely a result of the presence of a small amount of 'background seismicity' that does not participate in triggering related to the targeted 'main shock.' This seismicity could be independently triggered late aftershocks from a large event (e.g., from the Landers or Hector Mine sequences), or swarm-related activity. Distinguishing among these possibilities will be a goal of future work, which will consider the time and distance dependence of the foreshock and aftershock activity.

[37] The lowered *b*-value of the apparent 'aftershocks' of *M* 2.5 to 5.5 target events in southern California compared to the complete catalog is a robust result. A relatively low *b*-value for aftershocks was also observed by *Reasenberg and Jones* [1989] for 62 California aftershock sequences of $M \geq 5$, who obtained a mean *b*-value of 0.90 ± 0.02. This *b*-value difference between aftershocks and 'background' events does not appear in synthetic triggering calculations, even when space/time windowing is applied to simulate the processing applied to the real data. Taken at face value, this suggests that these events are not drawn from the same distribution of magnitudes as the general event population. This contradicts a common assumption in triggering models and suggests two possibilities: (1) The triggered events obey different G-R statistics than the overall catalog, or (2) many of the apparent foreshocks and aftershocks are not related to the 'main shock' by earthquake-to-earthquake

triggering, but are clustered due to an underlying physical process that generates events with a lower *b*-value than the complete catalog. As suggested by *Richards-Dinger et al.* [2010] it is likely that some of the data event pairs considered here (and in *Felzer and Brodsky* [2006]) are in swarms, which are often hypothesized to result from an underlying physical driving mechanism, such as slow slip or fluid flow. Swarm *b*-values have been observed both higher [e.g., *Sykes*, 1970] and lower [e.g., *Hainzl and Fischer*, 2002] than 1.0. However, an analysis of the 6950 events contained in the 71 southern California swarms identified by *Vidale and Shearer* [2006] yields a *b*-value indistinguishable from the complete catalog, and thus inclusion of swarm events in the data considered here does not explain the low *b*-values observed. This issue warrants further study.

7. Conclusions

[38] My emphasis here has been on the magnitude dependence of foreshocks and aftershocks in southern California compared to that predicted by a specific class of triggering models that approximate self-similar triggering (i.e., having $\alpha = b$), without fully considering the larger issues of their Omori's law time dependence or their distribution with distance. However, even within this relatively narrow focus some results are clear:

[39] 1. As previous authors have shown, earthquake triggering models, in which the increased triggering caused by larger magnitude events is exactly compensated for by their decreased numbers (i.e., in which $\alpha = b$), lead to relatively simple analytical expressions for average aftershock abundances. Model parameters can be adjusted so as to reproduce Båth's law, that is the average magnitude of the largest aftershock is 1.2 units smaller than the main shock. However, computer simulations of individual triggered sequences exhibit large variations in the numbers of aftershocks.

[40] 2. A good test of triggering self-similarity is to plot foreshock and aftershock rates as a function of differential magnitude relative to the magnitude of the largest event in the sequence. These *dN*/*dm* curves are nearly coincident for $\alpha = b$ models, regardless of main shock magnitude.

[41] 3. These simulated aftershock *dN*/*dm* curves have the same Gutenberg-Richter *b*-value as the underlying distribution, but the foreshock *dN*/*dm* curves have the same *b*-value only for foreshock magnitudes less than about $m_{\max} - 3$. For larger foreshock magnitudes, the *dN*/*dm* curve flattens and converges with the aftershock *dN*/*dm* curve at $m = m_{\max}$. This effect can explain observations of anomalously low *b*-values in some foreshock sequences and the decrease in apparent aftershock-to-foreshock ratios as the foreshock and aftershock magnitudes approach the target event magnitude.

[42] 4. Observed apparent foreshock and aftershock *dN*/*dm* curves for events close in space and time to *M* 2.5 to 5.5 main shocks in southern California appear roughly self-similar, but differ from triggering simulations is several key respects: (1) the aftershock *b*-values are significantly lower than that of the complete catalog, (2) the number of aftershocks is too large to be consistent with Båth's law, and (3) the foreshock-to-aftershock ratio is too large to be consistent with Båth's law. These observations indicate for southern California that either triggering self-similarity is not obeyed for these small main shocks or that a significant fraction of the space/time clustering is not primarily caused by earthquake-to-earthquake triggering.

Appendix A: Mathematical Details

[43] Here is a derivation of equation (12) provided by G. Backus (personal communication, 2010). We start with the integral

$$S = \int_0^1 dx \ln x(1-x)^{N-1} \tag{A1}$$

Using $\int_0^1 dx \ln x = -1$, we may write

$$\begin{aligned} S+1 &= \int_0^1 dx \ln x\left[(1-X)^{N-1} - 1\right] \\ &= \int_0^1 \ln x\, d\left[-\frac{1}{N}(1-x)^N - x + \frac{1}{N}\right] \\ &= \left[\ln x\left(-\frac{1}{N}(1-x)^N - x + \frac{1}{N}\right)\right]_0^1 \\ &\quad + \int_0^1 \frac{dx}{x}\left[\frac{(1-x)^N}{N} + x - \frac{1}{N}\right] \\ &= \int_0^1 dx \frac{(1-x)^N - 1}{Nx} + \int_0^1 dx \end{aligned} \tag{A2}$$

and thus

$$S = \int_0^1 dx \frac{(1-x)^N - 1}{Nx} \tag{A3}$$

Letting $x = 1 - y$, we have

$$\begin{aligned} S &= -\int_1^0 dy \frac{y^N - 1}{N(1-y)} = \frac{1}{N}\int_0^1 dy \frac{y^N - 1}{1-y} \\ &= -\frac{1}{N}\int_0^1 dy\left(1 + y + \cdots + y^{N-1}\right) \\ &= -\frac{1}{N}\left[1 + \frac{1}{2} + \cdots + \frac{1}{N}\right] \\ &= -\frac{1}{N}\left[\ln N + \gamma + O\left(\frac{1}{N}\right)\right] \end{aligned} \tag{A4}$$

where $\gamma = 0.5772157$ is Euler's constant, the limiting difference between the natural logarithm and the harmonic series. Substituting into equation (11) provides the large N approximation for the average maximum magnitude

$$\langle m_{\max} \rangle = m_1 + \frac{\log_{10} N}{b} + \frac{\gamma}{b \ln 10}. \tag{A5}$$

Appendix B: Theory for Aftershock and Foreshock *dN*/*dm* Curves

[44] Consider equation (11) for the average total number of aftershocks produced by a starting (first) event of magnitude m_F in the case of triggering with $\alpha = b$:

$$N_{\mathrm{AS}} = \frac{r}{1-r} \frac{10^{b(m_F - m_1)}}{b \ln(10)(m_2 - m_1)} \tag{B1}$$

where r is the branching ratio and m_1 and m_2 are the lower and upper magnitude limits, respectively. For such a sequence, the G-R distribution can be expressed as

$$N(\geq m) = \frac{r\left[10^{-b(m-m_F)} - 10^{-b(m_2-m_F)}\right]}{(1-r)\,b\ln(10)(m_2-m_1)} \tag{B2}$$

and its derivative as

$$\frac{dN}{dm} = \frac{r}{(1-r)(m_2-m_1)}\ 10^{-b(m-m_F)} \tag{B3}$$

Defining $\dot{N}_0$ as dN/dm at $m = m_F$, we can express this as

$$\frac{dN}{dm} = \dot{N}_0 10^{-b(m-m_F)} \tag{B4}$$

where

$$\dot{N}_0 = \frac{r}{(1-r)(m_2-m_1)} \tag{B5}$$

Notice that the expressions for dN/dm only depend upon magnitude relative to the first event (i.e., $m - m_F$).

[45] The average number of triggered events larger than the first event is given by

$$N_0 \equiv N(m \geq m_F) = \frac{r}{1-r}\ \frac{1-10^{-b(m_2-m_F)}}{b\ln(10)(m_2-m_1)} \tag{B6}$$

For the example that we considered earlier ($r = 0.5$, $m_1 = 0$, $m_2 = 7$), this gives values of $N_0 = 0.062$, 0.062, 0.061, and 0.056, for $m_F = 3$, 4, 5, and 6, respectively. These are the average number of aftershocks expected that are larger than the starting event. Note that, consistent with self-similarity, there is little dependence on magnitude except when the target event magnitude approaches the upper magnitude limit m_2. These values can be compared with the often-cited result of *Jones* [1985], that there is a 6% probability that a M 3–5 earthquake in southern California will be followed by a larger event within 5 days and 10 km, and the more recent study of *Christophersen and Smith* [2008], who found that the probability of a M 2–4.5 event in southern California triggering a larger event within one day is about 4%.

[46] Things become considerably more complicated if we compute magnitudes relative to the largest event in the sequence, $m_{\max}$, rather than the starting event. This is because some of the events described by equation (B3) with $m < m_F$ are nonetheless preceded or followed by events that are larger than m_F. In addition, equation (B3) only provides the average aftershock distribution, while as we saw earlier, the number of events in individual sequences is random and varies widely. In particular, a small number of main shocks that happen to generate large aftershocks contain many more events than typical aftershock sequences and contribute disproportionately to the averages in equation (B3). Computer simulations have the advantage of fully accounting for all these complexities, but they are computationally intensive and provide limited intuition regarding foreshock and aftershock behavior.

[47] To estimate the aftershock dN/dm curve with respect to the largest event in a triggered sequence, we need to apply corrections to equation (B3) that account for: (i) those events that were triggered by a previous event with $m > m_F$, and (ii) those events that are followed by an event of $m > m_F$. For case (i), recall that for $\alpha = b$, all magnitudes within a G-R distribution contribute equally to aftershock production. Thus some of the $m < m_F$ aftershocks included in equation (B3) were triggered not by the initiating event of magnitude m_F, but by an aftershock of $m > m_F$. To correct equation (B3) for this effect, we need to distinguish between the first generation aftershocks and the later aftershocks. It is only the later aftershocks that are possibly triggered by a larger event. From equations (6) and (11), the ratio of later aftershocks to total aftershocks is simply r. The later aftershocks are triggered by previous aftershocks that obey a G-R distribution between m_1 and m_2. Because $\alpha = b$, the fraction of later aftershocks (a fraction r of the total) that are preceded by an event larger than m_F is $(m_2 - m_F)/(m_2 - m_1)$. Thus, the aftershock dN/dm distribution for events not preceded by a triggered event larger than the initiating event is reduced by a correction factor of $1 - r(m_2 - m_F)/(m_2 - m_1)$.

[48] For case (ii), note that on average there are N_0 events larger than the first event. Thus, for small N_0 we can approximate the reduction in dN/dm using the factor $1 - N_0$. Taking both effects into account we have the corrected aftershock dN/dm function

$$\frac{dN'_{\rm AS}}{dm} \approx \frac{r[1-N_0]10^{-b(m-m_F)}}{(1-r)(m_2-m_1)}\left[1 - r\frac{m_2-m_F}{m_2-m_1}\right] \tag{B7}$$

or

$$\frac{dN'_{\rm AS}}{dm} \approx \dot{N}'_0\ 10^{-b(m-m_F)} \tag{B8}$$

where $\dot{N}'_0$ is $dN'_{\rm AS}/dm$ at $m = m_F$, i.e.,

$$\dot{N}'_0 \equiv \dot{N}_0\left[1 - r\frac{m_2-m_F}{m_2-m_1}\right][1-N_0] \tag{B9}$$

The dN/dm function defined by equation (B7) includes some events with $\Delta m > 0$ (as the equation only approximately corrects for the occurrence of these larger aftershocks). The average total number of these events is given by

$$N'_0 = N_0\left[1 - r\frac{m_2-m_F}{m_2-m_1}\right][1-N_0] \tag{B10}$$

Equation (B7) for $m \leq m_F$ gives the expected aftershock distribution following the largest event in the sequence, even if this event is larger than the initiating event. This is because these larger events will themselves generate aftershock sequences that follow equation (B3). The aftershock $\frac{dN}{dm}$ curve is a simple power law decay with slope $-b$ on a $\log_{10}$ plot.

[49] The predicted foreshock $\frac{dN}{dm}$ curve is more complicated because the initiating event (of magnitude m_F) is counted in the sequence. An approximation is provided by

$$\frac{dN'_{\rm FS}}{dm} \approx \dot{N}'_0 + \frac{1}{2}N'_0\dot{N}'_0 10^{-b(m-m_{\max})},\ m < m_{\max} \tag{B11}$$

Here the first term comes from the starting event and the second term comes from the triggered events that occur before the maximum magnitude event. Although this

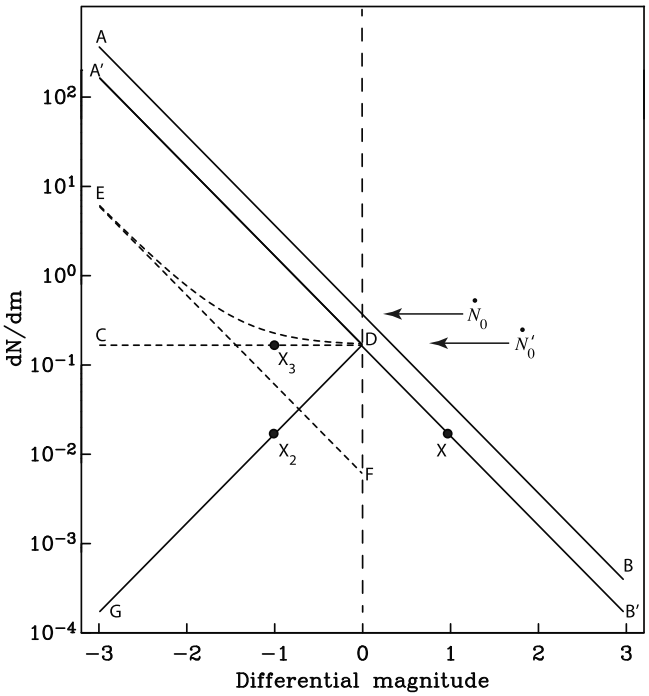

Figure B1. A diagram to explain the relationships among various *dN*/*dm* curves (see text).

approximation makes intuitive sense (see below) and is reasonably accurate in explaining the results of numerical triggering simulations (especially at small r values), I have not yet succeeded in deriving it analytically.

[50] Figure B1 illustrates where this equation comes from and many of the other relationships in this section. Line AB describes the average aftershock distribution, dN/dm, from an initiating event, plotted on a log scale as a function of magnitude relative to the starting event, $m - m_F$ (equation (B3)). The value of dN/dm at zero differential magnitude is $\dot{N}_0$. However, some of these events were triggered by an earlier aftershock that was larger than the starting event or are followed by an aftershock larger than the starting event. Thus, if we exclude these events, this line is reduced by a scaling factor to the line $A'B'$ (see equation (B7)). For this line, the value of dN/dm at zero differential magnitude is $\dot{N}_0'$. The left-hand portion of this line, the segment $A'D$, is the aftershock distribution for the largest event in the sequence. The right-hand portion of this line (DB') cannot be used directly, because these events are larger than the starting event. These larger events have associated aftershock and foreshock sequences. Their aftershock sequences have identical average properties as those from initiating events of the same magnitude and thus are described by $A'D$. Their foreshock sequences involve smaller preceding events. The first of these events is the initiating earthquake itself, which is not included on the line $A'B'$. Considering now events relative to the maximum magnitude event, then the first event contributes to the foreshock dN/dm curve. The line segment DB' maps to the equivalent segment DG and a point X with positive differential magnitude +1 becomes the point X_2 with negative differential magnitude −1. However, the larger magnitude events represented by X occur more rarely than the starting events in the sequence. Thus, to determine the average number of foreshocks per target event, we divide by a smaller number of target events. Correcting for this effect moves the contribution from X_2 up to the point X_3. A similar argument applies for all the other points along DB'. The decrease along DG is exactly compensated for by the proportional decrease in the number of target events. Thus the contribution of the starting events in the sequence to the foreshock dN/dm curve is the horizontal line CD, the first term in equation (B11).

[51] Next consider the segment $A'X$, which represents events smaller than the new reference magnitude $m_{\max}$, defined by the position of X. There are a total of N'_0 of events larger than the initiating event. If a larger event does occur, on average it occurs halfway through the sequence of smaller events defined by $A'D$. Because here we are considering only foreshock behavior, the fact that this larger event may generate a lengthy aftershock sequence does not matter. These factors contribute to the second term in equation (B11) and define the line EF, which is parallel to $A'D$, but reduced by the factor $\frac{1}{2}N'_0\dot{N}'_0$. The sum of these two terms gives the curve ED, which defines the foreshock dN/dm curve. Notice that for magnitudes very much smaller than the reference magnitude $m_{\max}$, it has the same b-value as the aftershock sequence. However, its slope is reduced as m approaches $m_{\max}$ and it intersects the aftershock dN/dm curve $A'D$ at zero differential magnitude. Notice that the expected ratio of the number of aftershocks to the number of foreshocks is magnitude dependent and approaches $\frac{1}{2}N'_0 = r/(1-r)(m_2 - m_1)$ when $m - m_{\max} < -3$, but approaches one as $m - m_{\max}$ approaches zero. Some intuitive insight as to why this occurs is given by the fact that, in the limit as Δm approaches zero, magnitude $m_{\max} - \Delta m$ events must on average trigger the same number of magnitude $m_{\max}$ events as magnitude $m_{\max}$ events trigger magnitude $m_{\max} - \Delta m$ events.

[52] **Acknowledgments.** This research was supported by the USGS/NEHRP program and by the Southern California Earthquake Center. Andrew Michael and an anonymous reviewer provided helpful comments. Karen Felzer provided a copy of her AftSimulator program, which was very helpful in understanding how computer simulations of triggering work, although I did not use AftSimulator itself for the results presented here.

References

Agnew, D. C., and L. M. Jones (1991), Prediction probabilities from foreshocks, *J. Geophys. Res.*, *96*, 11,959–11,971.

Båth, M. (1965), Lateral inhomogeneities of the upper mantle, *Tectonophysics*, *2*, 483–514.

Brodsky, E. E. (2011), The spatial density of foreshocks, *Geophys. Res. Lett.*, *38*, L10305, doi:10.1029/2011GL047253.

Christophersen, A., and E. G. C. Smith (2008), Foreshock rates from aftershock abundance, *Bull. Seismol. Soc. Am.*, *98*, 2133–2148, doi:10.1785/0120060143.

Console, R., A. M. Lombardi, M. Murru, and D. Rhoades (2003), Båth's law and the self-similarity of earthquakes, *J. Geophys. Res.*, *108*(B2), 2128, doi:10.1029/2001JB001651.

Felzer, K. R., and E. E. Brodsky (2006), Decay of aftershock density with distance indicates triggering by dynamic stress, *Nature*, *441*, 735–738.

Felzer, K. R., T. W. Becker, R. E. Abercrombie, G. Ekstrom, and J. R. Rice (2002), Triggering of the 1999 Mw 7.1 Hector Mine earthquake by aftershocks of the 1992 Mw 7.3 Landers earthquake, *J. Geophys. Res.*, *107*(B9), 2190, doi:10.1029/2001JB000911.

Felzer, K. R., R. E. Abercrombie, and G. Ekstrom (2004), A common origin for aftershocks, foreshocks, and multiplets, *Bull. Seismol. Soc. Am.*, *94*, 88–98, doi:10.1785/0120030069.

Gomberg, J., and K. Felzer (2008), A model of earthquake triggering probabilities and application to dynamic deformations constrained by ground motion observations, *J. Geophys. Res.*, *113*, B10317, doi:10.1029/2007JB005184.

Hainzl, S., and T. Fischer (2002), Indications for a successively triggered rupture growth underlying the 2000 earthquake swarm in Vogtland/NW Bohemia, *J. Geophys. Res.*, *107*(B12), 2338, doi:10.1029/2002JB001865.

Helmstetter, A. (2003), Is earthquake triggering driven by small earthquakes?, *Phys. Rev. Lett.*, *91*, 058501.

Helmstetter, A., and D. Sornette (2002a), Diffusion of epicenters of earthquake aftershocks, Omori's law, and generalized continuous-time random walk models, *Phys. Rev. E.*, *66*, 061104.

Helmstetter, A., and D. Sornette (2002b), Subcritical and supercritical regimes in epidemic models of earthquake aftershocks, *J. Geophys. Res.*, *107*(B10), 2237, doi:10.1029/2001JB001580.

Helmstetter, A., and D. Sornette (2003a), Foreshocks explained by cascades of triggered seismicity, *J. Geophys. Res.*, *108*(B10), 2457, doi:10.1029/2003JB002409.

Helmstetter, A., and D. Sornette (2003b), Predictability in the Epidemic-Type Aftershock Sequence model of interacting triggered seismicity, *J. Geophys. Res.*, *108*(B10), 2482, doi:10.1029/2003JB002485.

Helmstetter, A., and D. Sornette (2003c), Båth's law derived from the Gutenberg-Richter law and from aftershock properties, *Geophys. Res. Lett.*, *30*(20), 2069, doi:10.1029/2003GL018186.

Helmstetter, A., D. Sornette, and J.-R. Grasso (2003), Mainshocks are aftershocks of conditional foreshocks: How do foreshock statistical properties emerge from aftershock laws, *J. Geophys. Res.*, *108*(B1), 2046, doi:10.1029/2002JB001991.

Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2005), Importance of small earthquakes for stress transfers and earthquake triggering, *J. Geophys. Res.*, *110*, B05S08, doi:10.1029/2004JB003286.

Hutton, K., J. Woessner, and E. Hauksson (2010), Earthquake monitoring in southern California for seventy-seven sears (1932–2008), *Bull. Seismol. Soc. Am.*, *100*, 423–446, doi:10.1785/0120090130.

Jones, L. M. (1985), Foreshocks and time-dependent earthquake hazard assessment in Southern California, *Bull. Seismol. Soc. Am.*, *75*, 1669–1679.

Kagan, Y. Y. (1991), Likelihood analysis of earthquake catalogs, *Geophys. J. Int.*, *106*, 135–148.

Kagan, Y. Y. (2004), Short-term properties of earthquake catalogs and models of earthquake source, *Bull. Seismol. Soc. Am.*, *94*, 1207–1228, doi:10.1785/012003098.

Knopoff, L., Y. Y. Kagan, and R. Knopoff (1982), b-values for foreshocks and aftershocks in real and simulated earthquake sequences, *Bull. Seismol. Soc. Am.*, *72*, 1663–1676.

Lin, G., P. M. Shearer, and E. Hauksson (2007), Applying a three-dimensional velocity model, waveform cross correlation, and cluster analysis to locate southern California seismicity from 1981 to 2005, *J. Geophys. Res.*, *112*, B12309, doi:10.1029/2007JB004986.

Ogata, Y. (1998), Space-time point-process models for earthquake occurrence, *Ann. Inst. Stat. Math.*, *50*, 379–402.

Ogata, Y. (1999), Seismicity analysis through point-process modeling: A review, *Pure Appl. Geophys.*, *155*, 471–507.

Papazachos, B. C. (1975), Foreshocks and earthquake prediction, *Tectonophysics*, *28*, 213–226.

Reasenberg, P. A., and L. M. Jones (1989), Earthquake hazard after a mainshock in California, *Science*, *243*, 1173–1176, doi:10.1126/science.243.4895.1173.

Richards-Dinger, K., R. S. Stein, and S. Toda (2010), Decay of aftershock density with distance does not indicate triggering by dynamics stress, *Nature*, *467*, 583–586.

Saichev, A., and D. Sornette (2005), Distribution of the largest aftershocks in branching models of triggered seismicity: theory of the universal Båth's law, *Phys. Rev. E*, *71*, 056127.

Shcherbakov, R., and D. L. Turcotte (2004), A modified form of Båth's law, *Bull. Seismol. Soc. Am.*, *94*, 1968–1975.

Shearer, P. M., and G. Lin (2009), Evidence for Mogi doughnut behavior in seismicity preceding small earthquakes in southern California, *J. Geophys. Res.*, *114*, B01318, doi:10.1029/2008JB005982.

Smith, W. D. (1981), The *b*-value as an earthquake precursor, *Nature*, *289*, 136–139.

Sornette, D., and M. J. Werner (2005a), Constraints on the size of the smallest triggering earthquake from the epidemic-type aftershock sequence model, Båth's law, and observed aftershock sequences, *J. Geophys. Res.*, *110*, B08304, doi:10.1029/2004JB003535.

Sornette, D., and M. J. Werner (2005b), Apparent clustering and apparent background earthquakes biased by undetected seismicity, *J. Geophys. Res.*, *110*, B09303, doi:10.1029/2005JB003621.

Suyehiro, S. (1966), Difference between aftershocks and foreshocks in the relationship of magnitude to frequency of occurrence for the great Chilean earthquake of 1960, *Bull. Seismol. Soc. Am.*, *56*, 185–200.

Sykes, L. R. (1970), Earthquake swarms and sea-floor spreading, *J. Geophys. Res.*, *75*, 6598–6611.

Vere-Jones, D. (1969), A note on the statistical interpretation of Båth's law, *Bull. Seismol. Soc. Am.*, *59*, 1535–1541.

Vidale, J. E., and P. M. Shearer (2006), A survey of 71 earthquake bursts across southern California: Exploring the role of pore fluid pressure fluctuations and aseismic slip as drivers, *J. Geophys. Res.*, *111*, B05312, doi:10.1029/2005JB004034.