AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Solid Earth

RESEARCH ARTICLE

10.1002/2016JB013807

Key Points:
- The number of triggered aftershocks follows a lognormal distribution
- Its coefficient of variation is about 1.0 to 1.1 for southern California earthquakes
- A positive dependence of this variability with stress drop variability is expected

Correspondence to:
D. Marsan,
david.marsan@univ-savoie.fr

Citation:
Marsan, D., and A. Helmstetter (2017), How variable is the number of triggered aftershocks?, *J. Geophys. Res. Solid Earth*, *122*, 5544–5560, doi:10.1002/2016JB013807.

Received 7 DEC 2016
Accepted 26 JUN 2017
Accepted article online 29 JUN 2017
Published online 31 JUL 2017

How variable is the number of triggered aftershocks?

D. Marsan[1] and **A. Helmstetter**[2]

[1]Université Grenoble Alpes, Université Savoie Mont Blanc, CNRS, IRD, IFSTTAR, ISTerreU, Chambéry, France, [2]Université Grenoble Alpes, Université Savoie Mont Blanc, CNRS, IRD, IFSTTAR, ISTerre, Grenoble, France

Abstract Aftershock activity depends at first order on the main shock magnitude but also shows important fluctuations between shocks of equal magnitude. We here investigate these fluctuations, by quantifying them and by relating them to the main shock stress drop and other variables, for southern California earthquakes. A method is proposed in order to only count directly triggered aftershocks, rather than secondary aftershocks (i.e., triggered by previous aftershocks), and to only quantify fluctuations going beyond the natural Poisson variability. Testing of the method subjected to various model errors allows to quantify its robustness. It is found that these fluctuations follow a distribution that is well fitted by a lognormal distribution, with a coefficient of variation of about 1.0 to 1.1. A simple model is proposed to relate this observed dependence to main shock stress drop variability.

1. Introduction

The number of aftershocks $N(m)$ in an aftershock sequence is known to vary with the magnitude of the main shock m as $N(m) = K \exp(\alpha m)$. Many studies have investigated this productivity law, by stacking over large numbers of main shocks of equal magnitudes, in order to recover an ensemble average relationship. Parameter α is a key parameter in aftershock sequence modeling, and more generally in modeling seismicity dynamics. Its estimated value strongly depends on the type of method used, with direct counting within magnitude-dependent, fixed-sized space-time windows following main shocks typically leading to high values of $\alpha \simeq 2$ to 2.3 [*Felzer et al*., 2004; *Helmstetter et al*., 2005], while maximum likelihood methods generally find lower α values [e.g., *Hainzl and Marsan*, 2008, and references therein]. The two approaches provide different perspectives on the productivity law. If one assumes that any earthquake of magnitude m can directly trigger on average $N(m)$ aftershocks, then these aftershocks also trigger their own aftershocks and so on, so that one must distinguish between the number of directly triggered aftershocks (i.e., our $N(m)$) and the total number of aftershocks (i.e., direct and indirect). Maximum likelihood methods estimate numbers of direct aftershocks, while space-time window stacking counts total numbers. In theory, both methods should provide the same α exponent, if the productivity law $N(m)$ of direct triggering remains stationary. However, while based on more objective processing than window methods, maximum likelihood methods are more sensitive to model errors, in particular regarding to the use of point-like or finite rupture sources [*Hainzl et al*., 2008], to magnitude uncertainties [*Werner and Sornette*, 2008], to the exclusion of potential triggering earthquakes outside the studied period and area [*Wang et al*., 2010], and to possible occurrences of transient increases in background earthquake rate [*Hainzl et al*., 2013].

While the ensemble averaged productivity law $N(m) = K \exp(\alpha m)$ has thus been studied in some details, the variability of the productivity prefactor K from one main shock to the other has attracted much less attention. However, main shocks of equal magnitudes can trigger variable number of aftershocks, this variability going beyond the simple Poisson fluctuations around the mean number $N(m)$ [cf. e.g., *Marsan et al*., 2014, Figure 3]. Much of this variability owes to the occurrence (or absence) of large aftershocks, which in turn trigger more aftershocks, strengthening the natural variability of the process. It is therefore of particular importance, when investigating the variability around the mean productivity law, to properly account for this cascade of triggering, hence to evaluate which aftershocks are effectively directly triggered by the considered main shock.

Several dependences of aftershock productivity have been demonstrated or suggested in past studies. (i) Since aftershock occurrence is probably conditioned to afterslip [e.g., *Perfettini and Avouac*, 2007], a dependence on postseismic slip rate is expected [*Lange et al*., 2014]. (ii) Heat flow has been found by *Yang and Ben-Zion* [2009] to inhibit aftershock productivity, by favoring viscous instead of brittle deformation. For the five areas in southern California studied by these authors, a factor of about 10 is observed between the most and the least productive sequences, which is about twice the related ratio in heat flow (cf. their

Figure 6b), implying that this dependence could generate significant local variability of the aftershock production. This is particularly well evidenced in the case of oceanic transform faults, which are known to host relatively weak aftershock sequences [*Boettcher and Jordan*, 2004; *McGuire et al*., 2005]. (iii) Dependence of K on faulting style is described by *Tahir and Grasso* [2014, 2015]; thrust earthquakes are shown to trigger more aftershocks, which could be related to the larger increase of Coulomb stress change in this context [*Lin and Stein*, 2004]. (iv) Supershear ruptures have been shown to be depleted of on-fault aftershocks [*Bouchon and Karabulut*, 2008], the shock wave generated during the earthquake instead triggering off-fault aftershocks. It can be expected that this specific pattern has a direct influence on the number of triggered aftershocks. Inversely, slow ruptures have been observed to trigger fewer earthquakes [*Pollitz and Johnston*, 2006] than fast ruptures of equivalent size. It is however well known that such slow ruptures can also trigger abundant seismicity, as is for example the case of the off-Boso swarms [e.g., *Hirose et al*., 2012], the depth at which slow slip occurs likely exerting a major control on aftershock triggering [*Delahaye et al*., 2009]. (v) Finally, there exists a positive correlation between the numbers of foreshocks and of aftershocks, possibly related to preslip continuing in the postseismic phase [*Marsan et al*., 2014]. It can also be expected that the variability in the number of aftershocks is partly linked to other characteristics of the main shock, in particular its stress drop, and the spatial heterogeneity of the slip distribution.

Here we aim at estimating the variability of the productivity factor K. We will thus investigate direct aftershocks only, in an attempt to grasp the dependence of $N(m)$ on main shock characteristics. This analysis therefore needs to account for (1) the Poisson, natural variability of $N(m)$: the latter is the mean number, while the observed number of aftershocks is a random realization of a Poisson distribution with this mean $N(m)$; and (2) the triggering of indirect aftershocks, i.e., so to only count the earthquakes in the aftershock sequence that are effectively triggered by the main shock, not by previous aftershocks. This then allows to evaluate the residual variability beyond these two phenomena. We analyze this variability for earthquakes that occurred in California and study more particularly how it relates to stress drop variability.

2. Method

We here describe how we estimate the variability of the productivity prefactors K_i from one earthquake i to another. As a working hypothesis, we assume that the statistics of K do not depend on the main shock magnitude, so that the term $e^{\alpha m}$ in the productivity law $N(m)=Ke^{\alpha m}$ encompasses all the dependence on m. To quantify the variability of K around its mean $\overline{K}$, we seek to estimate the coefficient of variation $C = \frac{\sigma_K}{\overline{K}}$ where σ_K is the standard deviation.

The first step consists in estimating the numbers N_i of (directly) triggered aftershocks for each earthquake i. This is done by fitting a parameterized seismicity model to the data $\{t_i,m_i,x_i,y_i,z_i\}$; we here further develop previous works based on this approach; see *Zhuang et al*. [2011] for a review. We define λ the modeled rate density of earthquakes, i.e., the expected number per unit time and unit volume, as

$$\lambda(x,y,z,t) = \mu + \sum_{i/t_i<t} \nu_i(x,y,z,t) \tag{1}$$

with μ the background rate density, and

$$\nu_i(x,y,z,t) = K_i e^{\alpha m_i}(t+c-t_i)^{-p} f_i(x,y,z)/F_i \tag{2}$$

the interaction term, that incorporates the spatial kernel $f_i(x,y,z)$. While the productivity term $K_i e^{\alpha m_i}$ and the temporal dependence $(t+c-t_i)^{-p}$ are widely accepted as ensemble average laws, there is less consensus on the exact form for the spatial kernel f_i (see *Moradpour et al*. [2014] and *van der Elst and Shaw* [2015] for recent studies on this). We therefore use a histogram estimate for f_i: $f_i(r) = f(r/L_i)/L_i^3$ with $L_i = 10^{0.5(m_i-4)}$ the characteristic rupture length in km, which ensures a magnitude-independent stress drop and is in agreement with previous observations [*Utsu*, 2002; *van der Elst and Shaw*, 2015]. The kernel $f(s)$ is piecewise constant over logarithmically increasing bins and is normalized as $\int_0^\infty ds 4\pi s^2 f(s) = 1$ so that for all earthquake i, $\int_0^\infty dr 4\pi r^2 f_i(r) = 1$ also. To account for the fact that earthquakes can only occur in the half-space $z>0$, where z is depth taken positive, we compute the integral

$$F_i = \iint dx\, dy \int_0^{\infty} dz\, f_i(x, y, z - z_i) \tag{3}$$

for each earthquake i, so that $0 \leq F_i \leq 1$, and $F_i \rightarrow 1$ when $z_i \gg L_i$. This allows to normalize the interaction term v_i as described by equation (2). Without these correction factors F_i, we would introduce a break in the productivity law $N(m)=Ke^{\alpha m}$ at the characteristic magnitude of earthquakes rupturing the whole width of the schizosphere.

This model is inverted with an Expectation-Maximization algorithm that allows the processing of large data sets. We a priori fix the mean probability for an earthquake to be a background earthquake based on the estimate of *Hainzl et al.* [2006] $\frac{E\{t_{i+1}-t_i\}^2}{\mathrm{var}(t_{i+1}-t_i)}$. To facilitate the treatment, we draw at each iteration of the expectation step a realization of the branching structure. The algorithm is fully described in Appendix A. We then count the number of aftershocks N_i for each main shock i, based on the best model. These numbers are integer values and correspond to one random realization of the inverted branching structure.

In a second step, the statistics of K are estimated, by averaging over large numbers of earthquakes. As such, these statistics depend very little on the drawn realization of the branching structure. Integrating equation (1) over space and observation time $[t_i, t_{\max}]$ we find that the expected number of aftershocks N_i for main shock i is

$$N_i = K_i e^{\alpha m_i} X_i \tag{4}$$

with

$$X_i = \frac{(t_{\max} + c - t_i)^{1-p} - c^{1-p}}{1-p} \text{ for } p \neq 1, \text{ or, } X_i = \ln(t_{\max} + c - t_i) - \ln c \text{ for } p = 1 \tag{5}$$

We thus compute the estimated prefactors $\hat{K}_i$ as $\hat{K}_i = \frac{N_i}{e^{\alpha m_i} X_i}$. The terms X_i are correction terms related to the fact that the observed number of aftershocks must decrease for earthquakes occurring late in the catalog, since the observation period, of duration $t_{\max} - t_i$, is shorter after these earthquakes.

To illustrate this treatment, we run it on a synthetic catalog of 10,000 earthquakes generated with $\alpha = 2$, $p = 1.1$, and $c = 10^{-4}$ days, a branching ratio of 0.92, and a (continuous) 3-D spatial kernel $f(s) = \frac{3(\gamma-1)}{4\pi(1+s^3)^{\gamma}}$ with $\gamma = 1.6$; see Figure 1. In the case of this simulated data set, we do not impose a realistic thickness for the schizosphere and let the earthquakes populate a $(1000\ \mathrm{km})^3$ cubic volume. More realistic simulations and exhaustive testing are described in the next section. The simulated, "true" K_i values are drawn from a lognormal distribution with $\frac{\sigma_K}{K} = 1$. Magnitudes are drawn from a Gutenberg-Richter law with $b = 1$, starting at $m = 2.5$ (cutoff magnitude of the Californian data set analyzed in section 4). Our estimated model parameters are $\alpha, \hat{p} = 1.06$, and $\hat{c} = 8.3 \times 10^{-5}$ days, close to the true values. Comparing the estimated $\hat{K}_i$ with the true K_i, we obtain a linear correlation coefficient of 30% that increases to 94% if only considering the 13 largest earthquakes (with $m \geq 5.5$); see Figure 1c. We display in Figure 1d the cumulative distributions of the true and estimated K values, the latter shown for various magnitude cutoffs.

It is clear from this example that the estimated coefficient of variation $\hat{C} = \frac{\sigma_{\hat{K}}}{\hat{K}}$ only becomes a good estimate of the true $\frac{\sigma_K}{K}$ when a high magnitude cutoff is imposed. It can be shown (see Appendix B) that this arises from the fact that N_i are integer values; this adds an extra variability on $\hat{K}_i$ that vanishes when N_i take large values, i.e., for large main shocks. We exploit equation (B1) that describes how $\frac{\sigma^2_{\hat{K}}}{\hat{K}}$ depends on the magnitude cutoff m, by fitting the $\log \frac{\sigma_{\hat{K}}}{\hat{K}}$ versus m curve to yield the estimated $\hat{C}$. Figure 2 shows this fit for the synthetic catalog of Figure 1, for which we find $\hat{C} = 0.92$ in place of the true value $C = 1$; the error bars are here only proportional to the inverse square roots of the weights $\frac{n(m)}{\sigma^2_{\hat{K}}}$ used when fitting the datapoints, where $n(m)$ is the number of earthquakes with $m_i \geq m$, and do not account for the uncertainty on $\sigma_{\hat{K}}$. This method allows to

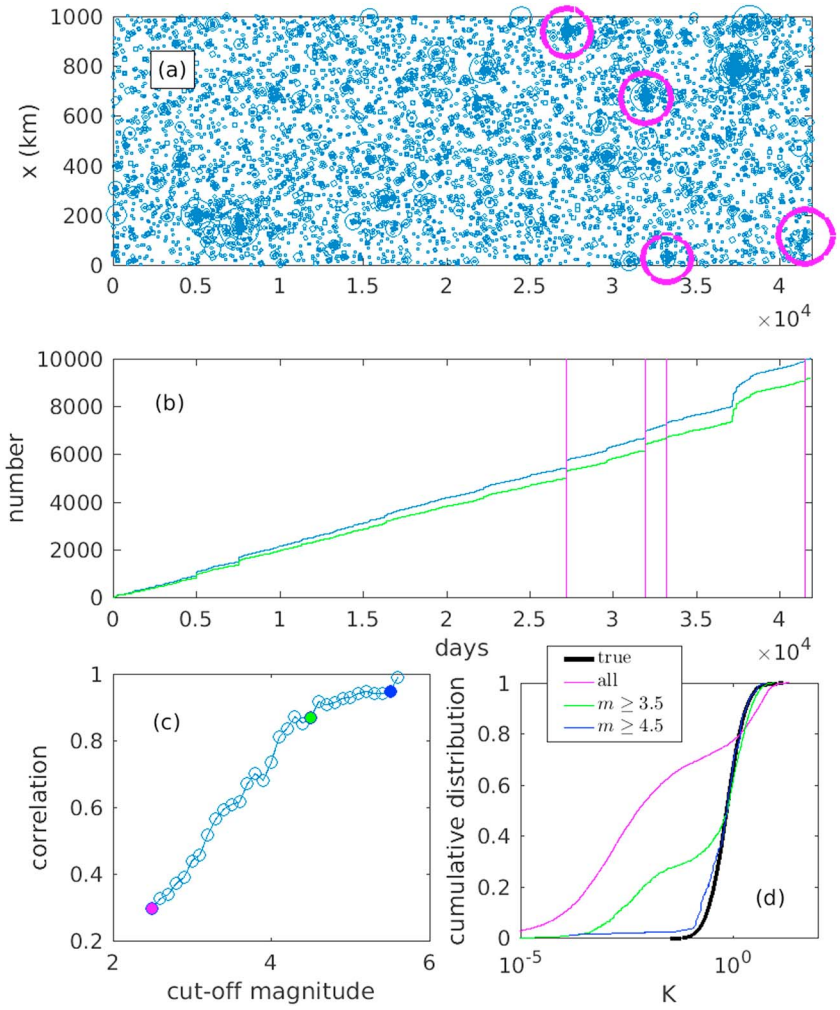

Figure 1. Synthetic catalog with imposed $C = \sigma_K/\overline{K} = 1$ variability. (a) Space-time plot, with space shown along the first direction x. The four $6.1 < m < 6.4$ main shocks removed when testing the influence of time-varying $\mu(t)$ are shown in magenta; see section 3.4. (b) Cumulative time series in blue, with the occurrence times of the four $6.1 < m < 6.4$ main shocks indicated with vertical lines. The green curve is after removing simulated nondetected aftershocks; see section 3.2. (c) Correlation coefficient between the true and inverted prefactors K, function of the magnitude cutoff m. The colored circles refer to the colored distributions of graph (Figure 1d). (d) Cumulative distributions of the prefactors K, for various magnitude cutoffs, compared to the true lognormal distribution, with K normalized by the true mean value.

account for the discretization of N_i but still using all information at hand rather than only considering a small set of large main shocks.

3. Tests

We here investigate the accuracy of the method using synthetic catalogs. The goal is to assess how this treatment behaves in presence of model errors, i.e., when the estimating model (as described in section 2 and Appendix A) differs from the generating model. More specifically, we study how the estimation of K is affected in the presence of (1) model errors related to wrong assumptions about the spatial distribution of earthquakes, namely, the fact that there exists a finite width of the schizosphere and that earthquakes occur on faults with finite (rather than vanishingly small) size. The analyzing model assumes that aftershocks are distributed isotropically around the source hypocenter modeled as a point. This simplifying assumption is required since in general the main shock fault geometry and its position relative to its hypocenter is unknown, although attempts at reconstructing fault networks from seismicity data [e.g., *Ouillon et al.*, 2008] could help tackling this issue; (2) model errors related to a rapidly varying magnitude of completeness

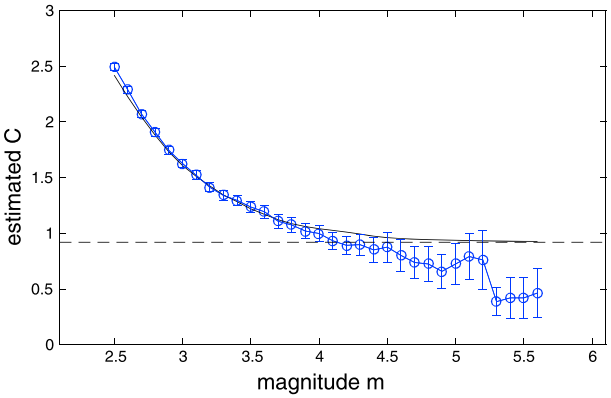

Figure 2. Estimated coefficient of variation $\hat{C} = \frac{\sigma_{\hat{K}}}{\overline{K}}$ function of the magnitude cutoff m, in blue, for the simulated catalog of Figure 1. The error bars are proportional to $\frac{\sigma_{\hat{K}}}{\sqrt{n(m)}}$, with $n(m)$ the number of earthquakes with magnitude above the cutoff magnitude m. The best fit obtained with the model of equation B1 is shown in black; it converges at large m toward $\hat{C} = 0.92$ (dashed line), which is the estimated coefficient for this catalog (the true C value being 1).

following large main shocks, i.e., when small earthquakes that are normally detected are not listed in the catalog due to temporary anomalously high earthquake rates; (3) model errors due to possible fluctuations in the parameters of the Omori-Utsu law, from one main shock to the other; and (4) model errors due to unmodeled temporal and spatial variations in the background rate μ.

Our synthetic catalogs are all generated with the model of equations (1) and (2), with parameter values $\mu = 0.1$, $\alpha = 2$, $p = 1.1$, $c = 10^{-4}$ days, $\overline{K} = 3.24\times10^{-5}$ (giving a branching ratio of 0.92, cf. *Helmstetter et al.* [2003]; here the branching ratio is computed using $\overline{K}$). Magnitudes are drawn from a Gutenberg-Richter law with $b = 1$, starting at $m_c = 2.5$ and with no upper bound. The statistics of the prefactor K differ from catalog to catalog. The estimated $\hat{C}$ values are detailed in Table 2.

3.1. Spatial Distribution

We consider 3 cases for the spatial distribution, use 3 values of C for each case ($C = 0$, 1, or 2), and simulate 10 independent catalogs of 10,000 earthquakes each, for each of the nine scenarios. Table 1 details the characteristics for each scenario. Catalogs with $C = 0$ correspond to a null model, with no variability in K; the estimated $\hat{C}$ must therefore be close to 0 if the method does not create artificial variability.

In case 1, the background earthquakes populate an unrealistic $1000 \times 1000 \times 1000$ km^3 volume, and aftershocks are located isotropically from their parent hypocenter according to a $f(s) = \frac{3(\gamma-1)}{4\pi(1+s^3)^\gamma}$ kernel, with $\gamma = 1.6$. The seismogenic volume is very big compared to the typical rupture lengths: for $b = 1$ and 10,000 earthquakes above $m = 2.5$, the ensemble average maximum magnitude is $\frac{1}{\ln 10}\ln 10^4 + \frac{0.577}{\ln 10} + 2.5 = 6.75$, hence an average maximum rupture length of 23 km, much less than 1000 km. Edge effects are thus very limited, and the generating model can therefore be considered the same as the analyzing model.

In case 2, we limit the seismogenic volume to (1000×1000) km$^2 \times 10$ km (in depth). The positions of the aftershocks relative to their parent hypocenter are determined by first drawing a distance r according to the same spatial kernel $f(s)$ as for case 1, and then the azimuth and takeoff angles are drawn at random; the position is checked to ensure that it lies within the 10 km wide schizosphere, or the azimuth and takeoff angles are redrawn until it does so.

Finally, in case 3, the same procedure is applied, but instead of locating the aftershock relative to the parent hypocenter, we do the same but relative to a point source that is taken at random on the causative fault. For each earthquake, we define a square rupture plane dislocation of size

Table 1. Summary of the Nine Scenarios for the Synthetic Catalogs Used in the Tests[a]

Case	$\sigma_K/\overline{K}$	Seismogenic Volume	Source Geometry
1	0	$1000 \times 1000 \times 1000$ km^3	Point
1	1	Idem	Point
1	2	Idem	Point
2	0	$1000 \times 1000 \times 10$ km^3	Point
2	1	Idem	Point
2	2	Idem	Point
3	0	Idem	Square dislocation
3	1	Idem	Square dislocation
3	2	Idem	Square dislocation

[a]For each scenario, 10 independent catalogs are drawn, containing 10,000 earthquakes each.

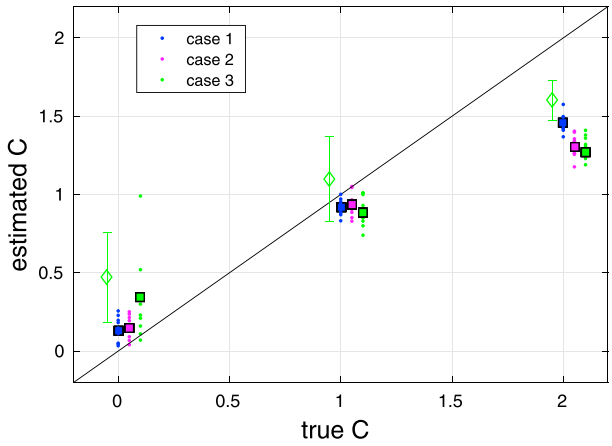

Figure 3. Estimated $\hat{C}$ versus the true coefficient of variation C, for the 90 synthetic catalogs. The dots have been slightly shifted for clarity purposes (the true C is either 0, 1, or 2). The mean of $\hat{C}$ for each case (hence averaged over 10 synthetic catalogs) is shown with the outlined square boxes. The mean values and standard deviations of $\hat{C}$ are detailed in Table 2. The diamonds refer to the mean of the estimated $\hat{C}$ in presence of model errors due to both the finite size of the earthquake rupture (case 3) and fluctuations in the magnitude of completeness after large main shocks.3

$L \times L$, still with $L = 0.01 \times 10^{0.5m}$ (in km), centered on the hypocenter. The strike is drawn uniformly in a 40° interval, and the dip is uniform between 70° and 90°. The rupture cannot extend above $z = 0$ or below $z = 10$ km and is made rectangular if it reaches these limits, keeping a L^2 overall surface. Although reality is far more complex, this case is intended to provide "Californian-like" catalogs. The point source assumption made in cases 1 and 2, and, more importantly, in the analyzing model, is known to significantly affect (Epidemic-Type Aftershock Sequences) ETAS-type inversions [*Hainzl et al.*, 2008], and we thus need to evaluate this specific limitation of the method.

We show in Figure 3 the estimated coefficient of variation $\hat{C} = \frac{\sigma_{\hat{K}}}{\overline{K}}$ versus the true coefficient C. The estimate is reasonably good for $C = 0$ and 1 but becomes significantly too low for $C = 2$. For $C = 0$, the inverted $\hat{C}$ is generally nonzero, since models with variable K are by construction much more flexible than those with constant K. Thus, a non-Dirac distribution of K is effectively obtained, although its dispersion is limited.

Among all model parameters, α is the one suffering from the strongest error. As shown in Figure 4, it is (except for one catalog among the 90 that were generated) always underestimated, the more so in case 3 for which model errors are the strongest. Too low an α value gives too strong a role on small main shocks as triggers: as a limit case, $\alpha = 0$ would imply that all earthquakes trigger on average equal numbers of aftershocks, whatever the size of the main shock. Models with a low α value have a better ability to explain fluctuations in the earthquake rate. This is best explained by considering the opposite situation of $\alpha \gg 1$, so that only the largest

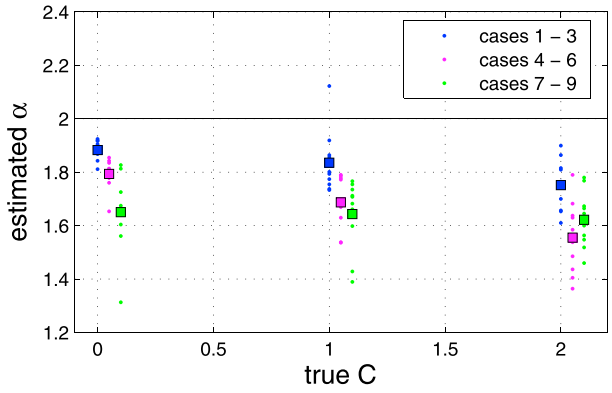

Figure 4. Estimated α value, compared to true value $\alpha = 2$. As in Figure 3, the dots have been slightly shifted to the right for visual clarity (the true C is either 0, 1, or 2). The mean of α for each case is shown with the outlined square boxes.

main shock has significant triggering capacity; any pattern in the time series distinct from an Omori-like decay after this shock would then end up being not modeled. As a consequence, underestimated α values yield flexible models that do not require extra variability in K, thus resulting in an underestimation of the variability of K. This in particular can at least partly explain the low $\hat{C}$ found for $C = 2$.

3.2. Undetected Aftershocks

Helmstetter et al. [2006] found that the magnitude of completeness m_c varies with time t after a main shock of (large) magnitude m, according to

$$m_c(t) = m - \delta - \phi \ln t \tag{6}$$

for t in days, with $\delta = 4.5$ and $\phi = 0.32$ using the Advanced National Seismic System catalog for California earthquakes. A similar relationship was found with the relocated data set of *Hauksson et al.* [2012] by *Hainzl* [2016], albeit with a slightly smaller value for δ; we infer $\delta \simeq 4.2$ by visual inspection of Figure 3 of *Hainzl* [2016].

We apply the $m_c(t)$ cutoff to our 90 synthetic catalogs of section 3.1, using $\phi = 0.32$ and the more constraining $\delta = 4.2$ value (cf. Figure 1b for an example), and rerun the estimation of C. On average, 8821 earthquakes are kept out of the initial 10,000. The estimated $\hat{C}$ is systematically larger than without undetected aftershocks, by on average 0.20; this difference grows roughly as $\sqrt{\hat{C}}$. Interestingly, the estimated α value is not statistically affected by this model error, increasing from 1.75 ± 0.12 on average to 1.82 ± 0.15: because of the limited detection, the observed number of aftershocks is found to decrease uniformly, at first order, between all main shocks regardless of their magnitude. Since periods following the largest earthquakes are more depleted of aftershocks than other periods, this acts to strengthen the observed variability of triggering.

3.3. Variability in the Omori-Utsu Law

Our model only allows parameter K to vary between main shocks. The mean number of aftershocks directly triggered by a main shock of magnitude m is $N(m)=Ke^{\alpha m}X$, cf. equations (4) and (5). Large variations in $N(m)$ between main shocks of equal magnitudes m are, in our approach, accounted for by an appropriate variability in K. However, parameters c and p are also potential sources of variability; we here address how these variabilities interfere with the estimated distribution of K.

We show in Appendix C that a lognormal distribution of c with a coefficient of variation $C_c = \sigma_c/\overline{c}$ causes the estimated $\hat{K}$ (for an analyzing model with constant c) to also follow a lognormal distribution, with a coefficient of variation $C = \sqrt{\left(1 + C_c^2\right)^{(p-1)^2} - 1}$, in the limit of an infinitely long observation time $t_{\max}$. For realistic values of p, i.e., close to 1, this implies only modest C values: for $C_c = 1$, we find $C = 0.083$ (for $p = 1.1$) and $C = 0.167$ (for $p = 1.2$); for $C_c = 2$, i.e., a strong variability in c, we find $C = 0.127$ (for $p = 1.1$) and $C = 0.255$ (for $p = 1.2$). Since there exists a tradeoff between the distributions of K and c, we must keep in mind that in our analysis which assumes a constant c value, a contribution (although likely small) in the estimated $\hat{C}$ actually originates from the unmodeled variability in c. The two effects are impossible to distinguish, at least in the absence of an independent estimate of the variability in c.

To evaluate the impact of a potential variability in p, we generate a new set of synthetic catalogs. Stability of the model requires $p > 1$ [*Zhuang et al.*, 2013]; moreover, even only a slight change in p can significantly impact $N(m)$, as it depends on p as $N(m) \sim \frac{c^{1-p}}{p-1}$. We therefore assumes a p value uniformly distributed between 1 and 1.4. We draw 10 synthetic catalogs of case 1, with $C = 0$. We find that $\hat{C} = 0.25 \pm 0.16$, compared to $\hat{C} = 0.15 \pm 0.14$ without accounting for this model error. This extra variability is only marginal, especially given the relatively large interval of p values.

3.4. Temporal Variations in Background Rate

We reanalyze the 90 synthetic catalogs of section 3.1, after removing all earthquakes with magnitude (i) $6.1 < m < 6.4$ and (ii) $m > 6.4$. The two tests (i) and (ii) are done separately. Removing an earthquake j from the catalog implies that its aftershocks now result from a space and time-varying background rate μ equal to ν_j, i.e., the direct effect of the removed main shock. Equation (1) then becomes

Table 2. Estimated Coefficient of Variation $\hat{C}$, Compared to the True Value C, for the Proposed Three Cases for the Spatial Distribution of Earthquakes (see Table 1)[a]

	$C = 0$	$C = 1$	$C = 2$
Case 1	0.15 ± 0.14	0.94 ± 0.07	1.31 ± 0.04
	0.29 ± 0.20	*1.17 ± 0.11*	*1.53 ± 0.15*
	0.16 ± 0.14	*0.98 ± 0.08*	*1.34 ± 0.06*
	0.27 ± 0.31	*1.03 ± 0.15*	*1.38 ± 0.06*
Case 2	0.13 ± 0.07	0.91 ± 0.06	1.45 ± 0.06
	0.18 ± 0.17	*1.19 ± 0.17*	*1.73±0.12*
	0.21 ± 0.18	*1.03 ± 0.07*	*1.44 ± 0.05*
	0.17 ± 0.14	*1.08 ± 0.18*	*1.50 ± 0.10*
Case 3	0.34 ± 0.25	0.91 ± 0.10	1.30 ± 0.05
	0.47 ± 0.29	*1.10 ± 0.27*	*1.60 ± 0.13*
	0.37 ± 0.29	*0.93 ± 0.10*	*1.31 ± 0.06*
	0.37 ± 0.34	*0.94 ± 0.13*	*1.31 ± 0.06*
	0.31 ± 0.18	*0.98 ± 0.10*	*1.34 ± 0.07*

[a]First value: with no other model errors. Second value: with undetected aftershocks, according to the $m_c(t)$ relationship of equation (6). Third value: after removing $6.1 < m < 6.4$ main shocks to simulate a space-time varying background rate μ. Fourth value: after removing $m > 6.4$ main shocks. Fifth value: using a spatially nonuniform background density. Table entries in italics give values corresponding to model errors.

$$\lambda(x, y, z, t) = \mu(x, y, z, t) + \sum_{i/t_i<t} \nu_i(x, y, z, t) \tag{7}$$

with

$$\mu(x, y, z, t) = \mu + \sum_{j/t_j<t} \nu_j(x, y, z, t), \tag{8}$$

where the summation in equation (8) is now done on the removed main shocks j. We therefore simulate slow slip events equivalent to magnitudes 6.1–6.4 (tests (i)) or greater than 6.4 (tests (ii)) earthquakes. On average, there are (i) 1.22 and (ii) 1.26 such simulated slow slip events per catalog. The average total number of aftershocks of the removed main shocks is 130 and 329, respectively, for tests (i) and (ii), which represents 1.3% and 3.29% of the total seismicity. There are in effect more earthquakes belonging to these simulated bursts of seismicity, as we here only count directly triggered aftershocks of removed main shocks, hence without the second (and more) generations of aftershocks. This compares well with the estimated 2% of the regional seismicity found by *Vidale et al.* [2006] for southern California. Our analysis of these simulated catalogs shows that ignoring temporal variations in background rate in the analyzing model has little effect on the estimated variability of K; see Table 2.

3.5. Spatial Variations in Background Rate

The analyzing model assumes that μ is independent of location (cf. equation (1)). However, the background density is effectively nonuniform; it can be directly computed and mapped as a by-product of our method, by smoothing ω_{0j}, the probabilities to be a background earthquake. We here use a smoothing length of 10 km, and an exponential smoothing kernel. As explained in section 4, the average ω_0 of ω_{0j} over all 42,865 analyzed southern California earthquakes is constrained to equal 0.29. Based on this nonuniform background density, we simulate 10 catalogs of 10,000 earthquakes, using a case 3 triggering kernel, for three values of C (0, 1, and 2). As shown in Figure 3 and Table 2, this does not affect the distribution of $\hat{C}$.

We conclude these tests by emphasizing the fact that while model errors indeed affect the model parameterization, our method provides a reasonable estimate of the variability of K, at least when $C = 0$ or 1, but underestimates it for stronger variability (e.g., $C = 2$), the underestimation caused by the nonisotropic nature of the earthquake rupture dominating the overall error for this level of variability. It is therefore relatively robust, even though it requires the nontrivial inversion of the full branching structure in presence of variable triggering.

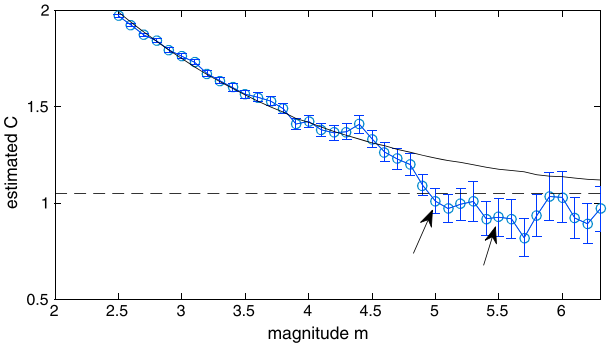

Figure 5. $\frac{\sigma_{\hat{K}}}{\hat{K}}$ versus main shock magnitude cutoff m, for the earthquake catalog of *Hauksson et al.* [2012]. The error bars are proportional to $\frac{\sigma_{\hat{K}}}{\sqrt{n(m)}}$ as in Figure 2. The best fit with equation (B1) is shown with the black line and yield an estimated $\hat{C} = 1.05$ (dashed lines). The distributions of K corresponding to the two C values computed at $m = 5$ and $m = 5.5$ (arrows) are shown in Figure 6.

4. Analysis of California Earthquakes and Relationship With Stress Drops

We analyze the data set of earthquakes in California, 1981–2016, by *Hauksson et al.* [2012]. Inspecting the magnitude-frequency relationship, we find a completeness magnitude of $m_c = 2.5$ for the 35 year long period. We thus analyze the 42,865 earthquakes with $m \geq 2.5$. The estimator of *Hainzl et al.* [2006] gives a mean probability to be a background earthquake of $\omega_0 = 0.29$.

We apply our method to find $\hat{C} = 1.05$, cf. Figure 5. We vary ω_0, also testing $\omega_0 = 0.2$ and $\omega_0 = 0.4$, as well as halving or doubling the characteristic lengths L, to measure the sensitivity of $\hat{C}$ to these parameters. This yields a range of $0.99 \leq \hat{C} \leq 1.11$. As the tests have shown, this value is possibly underestimated, although the bias cannot be known. The mean of $\hat{K}$ is 0.0082, giving an estimated branching ratio of 0.74; we moreover obtain $p = 1.24$ and $c = 0.039$ days.

Figure 6 suggests that at large magnitudes, the distribution of K can be modeled as lognormal, above the censoring value K_0. A Kolmogorov-Smirnov test cannot reject the null hypothesis that this distribution is lognormal with censoring, at the 95% significance level. A similar result is found for a chi-square test with discretization intervals counting at least five main shocks. However, given the size of the samples (120 and 38 main shocks, respectively), it is difficult to assess whether the distributions are indeed lognormal or could be better fitted by other functional forms. As demonstrated by the tests of section 3 (cf. Figure 1), censoring, especially at small magnitudes, causes the distribution of K to deviate from a pure, uncensored lognormal distribution. Moreover, even though a censored lognormal form describes the distribution of K well, at least at first order for $m \geq 5$ and $m \geq 5.5$ main shocks, we note that these two distributions give underestimated $\hat{C}$ values: for $m \geq 5$, we obtain that $\frac{\sigma_{\hat{K}}}{\hat{K}} = 1.01$, while $\frac{\sigma_{\hat{K}}}{\hat{K}} = 0.93$ for $m \geq 5.5$. This demonstrates the need to fit the $\frac{\sigma_{\hat{K}}}{\hat{K}}$ versus m curve as a whole (as we do, cf. Appendix B), rather than only considering values at specific magnitude cutoffs.

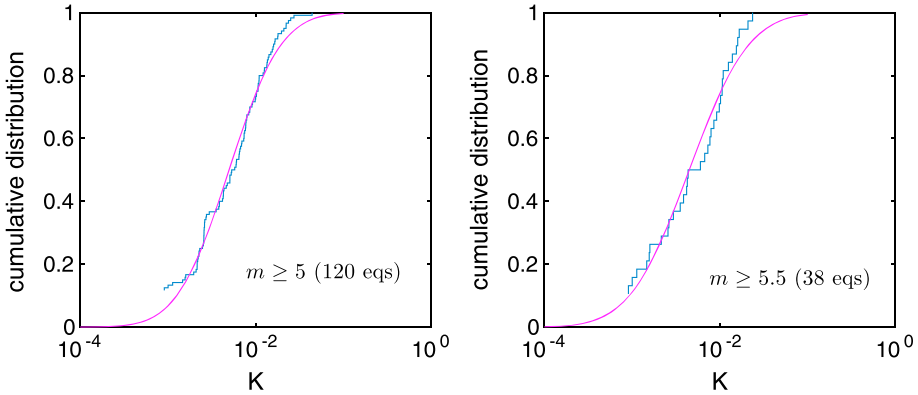

Figure 6. Distribution of main shocks with magnitude (left) $m \geq 5$ and (right) $m \geq 5.5$, which $\hat{C}$ values are indicated with arrows in Figure 5. The best lognormal fits are shown in magenta and yield $\frac{\sigma_{\hat{K}}}{\hat{K}} = 1.01$ and $\frac{\sigma_{\hat{K}}}{\hat{K}} = 0.93$, respectively, showing that a direct estimation of C based on the lognormal distribution of K for a single main shock magnitude cutoff can be unstable and little accurate.

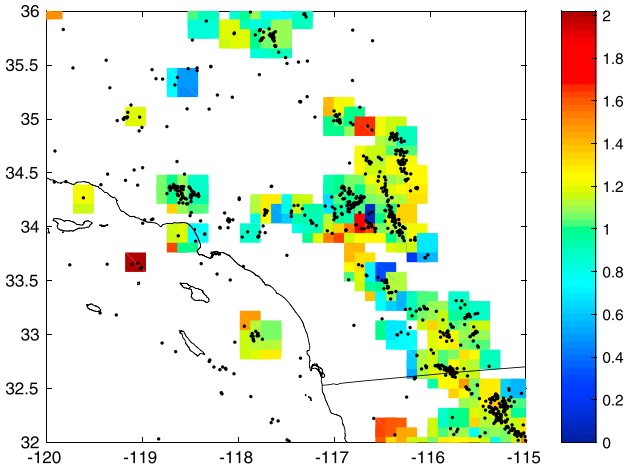

Figure 7. Estimated coefficient of variation $\hat{C}$ of the triggering prefactor K for individual square areas of 20 km length that count at least 100 earthquakes. We define a grid with 10 km spacing, so that cells overlap; we thus map the cells with 10 km instead of 20 km length, even though they really are 20×20 km^2 cells. The black dots are all earthquakes with $m \geq 4$.

We investigate possible spatial dependences of the variability of K, by running the estimation provided by equation (B1) on individual square areas with 20 km length, only keeping those cells that are populated by at least 100 earthquakes; see Figure 7. We do not find any clear pattern, as most areas exhibit a $\hat{C}$ value in the 0.9–1.2 interval with a mean of 1.09, hinting at a limited control by tectonic factors on this local variability. Most notably, the variability of K does not appear to be particularly high in the Salton Trough area, known for its high surface heat flow and its swarm-like seismicity. Heat flow is known to correlate with both the characteristics of aftershock triggering (in numbers, cf. *Boettcher and Jordan* [2004] and *McGuire et al.* [2005], and in temporal dependence, e.g., the p value of the Omori-Utsu law, cf. *Kisslinger and Jones* [1991]) and the average stress drop [*Oth*, 2013; *Hauksson*, 2015]. Moreover, large variations in stress drop have been observed in geothermal areas, likely linked to variations in pore pressure [*Chen and Shearer*, 2011; *Lengliné et al.*, 2014], that could also affect the triggering of aftershocks. Productivity variability displays a clear dependence with depth; see Figure 8, with little variability at shallow depth ($z < 5$ km). We did not find any such dependence in the synthetic catalogs (cases 2 and 3 that are limited to 0–10 km in depth). This feature is not simply due to a lack of resolution at these depths, as the number of earthquakes rapidly increase after the first 2 km, and similar low numbers of earthquakes at greater depths ($z \geq 14$ km) yield much larger variability coefficients $\hat{C}$.

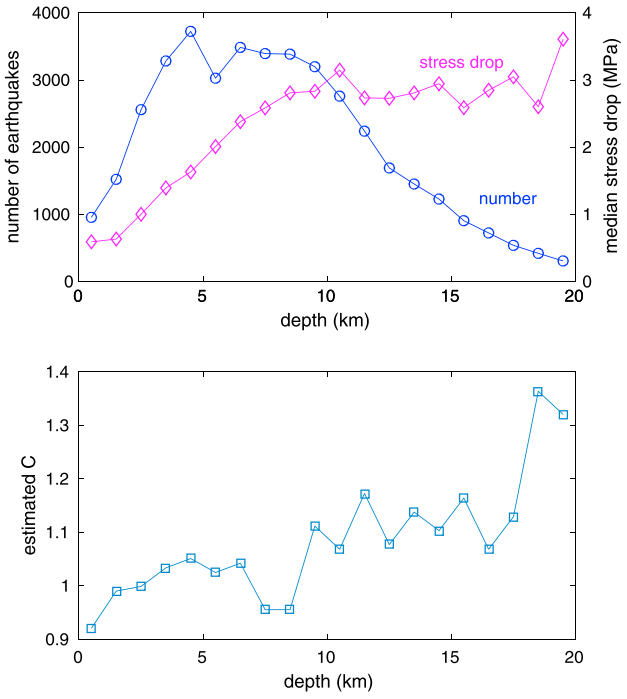

Figure 8. Dependence on depth of (top) the number of earthquakes and the median stress drop, and (bottom) the coefficient of variation $\hat{C}$. The fit provided by equation (B1) is run on successive 1 km-wide intervals.

It can be expected that the number of aftershocks depends on the main shock stress drop, which we here denote τ. We now address the question as to how much the variability of τ could contribute to the variability of K. Rate-and-state friction models predict that on the long term, the total number of earthquakes

triggered by a stress step is proportional to the stress step, even when it is accompanied by a change in stressing rate as expected during seismic swarms [*Helmstetter and Shaw*, 2009]. For a dislocation of size L and slip u, the stress change is proportional to u/L and affects a volume that scales with L. We assume a fractal distribution of target faults around the dislocation, such that their number grows as L^D. The value of D was found to be 1.65 by both *Marsan and Lengliné* [2010] and *Moradpour et al.* [2014], for which we infer this value from the $r^{-1.35}$ decay of the linear density on the intermediate regime (Figure 3 of *Moradpour et al.* [2014], for 100 m $< r <$ 10 km, for $3 < m < 4$ main shocks), in both cases for relocated California earthquake data sets. Alternatively, *Ouillon and Sornette* [2011] find $D = 1.85$ for the fault network activated during the 1986 Mount Lewis sequence in northern California, while *Hainzl et al.* [2014] suggest that $D = 1.8$ for the distribution of faults in southern California. The number of aftershocks N thus depends on stress drop τ and fault length L as $N \propto \tau L^D$. We now proceed to express N as a function of the two variables τ and M_0 instead of τ and L, so to address how it depends on stress drop and magnitude. The seismic moment scales as $M_0 \propto uL^2$ [*Aki*, 1967], and $\tau \propto u/L$, thus $L \propto \left(\frac{M_0}{\tau}\right)^{1/3}$. This implies that $N \propto \tau L^D \propto \tau^{1-D/3} M_0^{D/3}$. The relationship between seismic moment M_0 and local magnitude m is of the form $M_0 \propto 10^{\gamma m}$, with γ typically ranging from 1 to 1.5 [*Prieto et al.*, 2004; *Shearer et al.*, 2006]. We observe that $N \propto e^{\alpha m}$ at constant stress drop, hence $N \propto M_0^{\alpha/\gamma \ln 10}$, i.e., $D = \frac{3\alpha}{\gamma \ln 10}$. The dependence of K on τ must therefore be of the form $K \propto \tau^{1 - \alpha/\gamma \ln 10}$. For $1.65 \leq D \leq 1.85$ and $1 \leq \gamma \leq 1.5$, we find that $1.26 \leq \alpha \leq 2.12$. Our estimated α is 1.38, although we have shown in section 3 that the estimated α is generally significantly lower than the true value (cf. Figure 4). Published maximum likelihood estimates of α for California typically range from 1.1 to 2.0 [*Marsan and Lengliné*, 2008; *Werner et al.*, 2011; *Woessner et al.*, 2011; *Hainzl et al.*, 2013; *Schoenberg*, 2013], depending on the method, the underlying assumptions, and the data set analyzed. The expected $1.26 \leq \alpha \leq 2.12$ interval is therefore coherent with independent estimates of α. Based on this argument, we obtain that K scales as τ^a, with $0.38 \leq a \leq 0.45$.

An alternative interpretation of our results comes from the observation that the spatial clustering of earthquakes is multifractal rather than monofractal [*Hirata and Imoto*, 1991; *Hirabayashi et al.*, 1992; *Hooge et al.*, 1994; *Legrand et al.*, 1996; *Pasten and Comte*, 2014], in which case our previous argument implies that a distribution of fractal dimensions D should give a distribution of α parameters. Local α values are thus expected; the productivity law for any earthquake of index i is then $N_i = K_i e^{\alpha m_i}$. Assuming a constant α value as we do in our analysis would then artificially increase the estimated variability of K. A model in which K and α are both variable is however not tractable: the two parameters cannot be separated on the sole basis of counting aftershocks. We thus must be cautious that our model with variable K and constant α is only an efficient representation of a possibly more complex reality.

Stress drop variability for California earthquakes has been studied by *Shearer et al.* [2006], for magnitude 1.5 to 3.1 events occurring in 1989–2011. As is very generally the case [e.g., *Allmann and Shearer*, 2007, 2009; *Baltay et al.*, 2011; *Oth et al.*, 2010], a lognormal distribution of τ was observed, here with a standard deviation $\sigma_{\ln\tau} = 1.52$, but reducing to $\sigma_{\ln\tau} = 1.10$ for the best recorded shocks (i.e., with at least 20 readings). The latter criterion resulted in removing outliers, hence a better fit with a lognormal law, and a reduction of the dispersion. Taking $\sigma_{\ln\tau} = 1.10$ and $K \propto \tau^a$ with $a = 0.4$ implies that $\sigma_{\ln K} = 0.44$ if the variability of K was solely due to the variability in stress drop. As $\sigma_{\ln K} = \sqrt{\ln\left(1 + C^2\right)}$ for lognormal laws, $\hat{C} \simeq 1.05$ translates into $\sigma_{\ln K} \simeq 0.86$. The contribution of stress drop variability to the overall variability of triggering would thus amount to about 50%, according to our analysis. A direct relationship between individual values of τ and K is not investigated here: for the $M_L \leq 3.1$ magnitude interval explored in *Shearer et al.* [2006], the individual K values are badly resolved, cf. Figures 1c and 1d, owing to the censoring effect, cf. Appendix B. Only the overall level of variability, measured with the coefficient of variation C, can be correctly estimated even in the presence of this censoring, as demonstrated by Figure 1 and more generally the tests of section 3.

Finally, we note that the dependence of median stress drop with depth of *Shearer et al.* [2006] bears some similarity with the observed dependence of $\hat{C}$ with depth; the correlation between the two quantities is 57%, cf. Figure 8. *Shearer et al.* [2006] commented that the increase of median stress drop with depth, at shallow depth ($z \leq 8$ km), is a robust feature of their analysis. Earthquakes occurring at shallow depths were found to be relatively depleted in high frequencies, which could potentially impact their capacity to trigger large numbers of aftershocks, hence reducing the variability of aftershock productivity.

5. Discussion and Conclusions

The estimation of K for individual earthquakes is a delicate issue, as relaxing the K = constant hypothesis classically done in ETAS modeling implies that the model becomes very flexible, and possibly nonrobust. Our approach was here to decipher the branching structure of triggering from the data, using a 3-D space-time ETAS model. Other models could be used to do so [*Marsan and Lengliné*, 2008; *Zaliapin and Ben-Zion*, 2013]. Trade-offs between model parameters are observed when analyzing synthetic data sets, especially when model errors are introduced, leading to an underestimation of the variability of K. The fact that $\hat{C}$ is actually below the real C value (see Figure 3) is a positive feature, as it implies that the model does not overfit the data, i.e., generating an ad hoc and trivial representation of the earthquake time series.

The estimated variability of K exhibits robust features: K values are distributed lognormally above censoring, when this censoring for small main shocks is not strong, and the variability is at first-order independent of location, although a correlation with depth is noted. A coefficient of variation C of 1.05, possibly increasing up to 1.5 if accounting for errors in the modeling of the earthquake time series, is found. Aftershock productivity does thus vary substantially from one main shock to the other (of equal magnitudes): for $C = 1.5$ and a lognormal distribution of K, 58% of main shocks trigger less than half or more than twice the mean number of direct aftershocks; this probability is still 6% for gains in numbers of aftershocks greater than 10 times or less than 1/10 of the mean.

The fact that both K and stress drop follow lognormal distributions, along with intuitive and physical arguments, point to a significant correlation between the two. A simple model predicts that 40% to 80% of the variability of K could be caused by stress drop variability.

Spurious apparent variability of K could be caused by using a triggering model with spatially and temporally uniform parameters. If any one of the model parameters (in particular α, p, and μ in our case) depends on the location and/or is subject to temporal fluctuations, the estimated variability of K will automatically increase, which is the only model parameter that can vary from one main shock to the other. While α values are difficult to estimate, and their actual geographical or temporal variability is therefore little known (see however *Wang et al.* [2010], for large-scale variations, and *Enescu et al.* [2009] and *Hainzl et al.* [2013] for its dependence on heat flow), p and μ are known to exhibit significant variabilities. Previous analyses suggest that the p value could depend on the faulting style [*Tahir and Grasso*, 2015], on the surface heat flow [*Kisslinger and Jones*, 1991], on the magnitude of the main shock [*Ouillon and Sornette*, 2005; *Hainzl and Marsan*, 2008], on depth (P. Shebalin, personnal communication), and even possibly on rupture frictional heating [*Wiemer and Katsumata*, 1999]; c has also been reported by *Narteau et al.* [2009] to vary with the tectonic regime, which could impact the estimated p values as both parameter estimates are strongly correlated. A theoretical dependence of c on main shock and aftershock magnitudes is also argued by *Davidsen and Baiesi* [2016] to be required in order to preserve the self-similar scaling of earthquake triggering laws. The μ parameter is also known to exhibit spatial variability at all scales [e.g., *Zhuang et al.*, 2002; *Helmstetter and Werner*, 2012], as well as large temporal fluctuations, especially during swarms, as it models the tectonic as well as the potentially rapidly varying aseismic forcing [*Reverso et al.*, 2015]. Swarm activity exists in California [*Vidale and Shearer*, 2006], and so this could affect our analysis. Given the difficulty to invert the branching structure for a K-variable model, we do not attempt at further complexifying it and therefore do not explore even more flexible models with nonconstant α, p, or μ parameters that could yield a weaker variability of the productivity K factor. Since the actual variability of these parameters is unknown, we cannot infer at this stage how they would affect our results. We emphasize that our estimation of the variability of K for southern California earthquakes relies on the assumption of an otherwise (i.e., apart from K) homogeneous and stationary model.

Appendix A: Computation of the Numbers of Aftershocks N_i

Given the earthquake data set $\{t_i, m_i, x_i, y_i, z_i\}$, the goal is here to estimate, for each earthquake i, the number N_i of its aftershocks. This is a delicate issue, especially as we here relax the hypothesis that the prefactor K is the same for all earthquakes. Doing so, we add as many unknowns as there are earthquakes in the catalog,

making this problem highly nontrivial. We show in sections 2 and 3 that while the individual values of K are generally badly estimated, its overall distribution can be recovered, at least when considering the largest main shocks. Moreover, the coefficient of variation characteristic of this distribution can be relatively well estimated.

We model seismicity with equations (1) and (2). We start with an initial guess on the model parameters $\{\alpha, p, c\}$, and on the spatial kernel $f(s)$. The prefactors K_i are initially assumed to be all equal, the chosen value $\overline{K}$ mattering not. The algorithm consists in iterating three steps: (a) given the values $k_i = K_i/\overline{K}$, we draw a realization of the branching structure; (b) we optimize the model parameters for this drawn structure; and (c) we finally update the prefactors $k_i = K_i/\overline{K}$.

Step A. We first need to compute the probabilities ω_{ij} that earthquake i triggered earthquake j, defined as $\omega_{ij} = \dfrac{v_{ij}}{\mu + \sum_{k<j} v_{kj}}$, where $v_{ij} = v_i(x_j, y_j, z_j, t_j)$, cf. equation (2), for all pairs $\{i,j\}$. The direct estimation of μ would consist in dividing the number of background earthquakes by the duration of the catalog and by the seismogenic volume, i.e., the total volume in which earthquakes can occur. This latter quantity is difficult to evaluate, and we therefore follow an approach that avoids the determination of this volume: instead of estimating μ, we compute the ratio $\mu/\overline{K}$. The probabilities ω_{ij} can then be rewritten as

$$\omega_{ij} = \frac{v_{ij}'}{\mu/\overline{K} + \sum_{k<j} v_{kj}'}$$

with

$$v_{ij}' = \frac{K_i}{\overline{K}} e^{\alpha m_i} \left(t_j + c - t_i\right)^{-p} f_i\left(x_j, y_j, z_j\right)$$

Similarly, the probability that earthquake j is a background earthquake is

$$\omega_{0j} = \frac{\mu/\overline{K}}{\mu/\overline{K} + \sum_{k<j} v_{kj}'}$$

so that $\omega_{0j} + \sum_{i<j} \omega_{ij} = 1$. We invert the ratio $\dfrac{\mu}{\overline{K}}$ by constraining

$\omega_0 = \dfrac{1}{N}\sum_j \omega_{0j} = \dfrac{E\{t_{i+1} - t_i\}^2}{\mathrm{var}(t_{i+1} - t_i)}$ [*Hainzl et al.*, 2006]. This then permits the full knowledge of all ω_{ij} values.

To speed up the processing, we draw a random realization of the branching structure, by attributing one single parent to all earthquakes. For each earthquake j, a unique parent of index i (with $1 \le i < j$), or $i = 0$ if j is background, is drawn, using the probabilities ω_{ij} and ω_{0j}.

Step B. The model parameters $\{a,p,c\}$, and $f(s)$, are optimized. Parameters p and c are found by fitting the density $f_t(t) = \frac{p-1}{c^{1-p}}(t+c)^{-p}$ to the time lags $\delta t_j = t_j - t_i$ between earthquake j and its parent i. We here do not account for the finite observation time; although the estimates of p and c are improved when using finite observation times, the final estimate of K_i and its distribution is only very marginally affected, while the computation time is significantly reduced in the infinite observation time limit we adopt here. Maximizing the likelihood $\prod_j f_t(\delta t_j)$ is then equivalent to solving the equation in c:

$$c\, E\left\{\frac{1}{\delta t_j + c}\right\}\left(\ln c - 1 - E\left\{\ln\left(\delta t_j + c\right)\right\}\right) + 1 = 0.$$

The estimate of c is then used to compute the estimate of p as follows:

$$p = \frac{1}{1 - cE\left\{\frac{1}{\delta t_j + c}\right\}}.$$

Given the drawn branching structure, the number of aftershocks N_i is known for each earthquake, as is the expected number $\widetilde{N}_i = K_i e^{\alpha m_i} X_i$, where $X_i = \frac{(T+c-t_i)^{1-p} - c^{1-p}}{1-p}$ and $K_i = \overline{K} k_i$. Maximizing the likelihood $\prod_i e^{-\widetilde{N}_i} \frac{\widetilde{N}_i^{N_i}}{N_i!}$ with respect to α and $\overline{K}$ yields that parameter α must be such that

$$\sum_i N_i m_i \times \sum_i K_i e^{\alpha m_i} X_i = \sum_i N_i \times \sum_i m_i K_i e^{\alpha m_i} X_i.$$

We numerically solve this equation to obtain α. We emphasize that parameter $\overline{K}$ can also be optimized along α, although it is not required in this treatment: in step (A), only the values $k_i = K_i / \overline{K}$ are needed, independently of $\overline{K}$.

Finally, we fit a piecewise-constant spatial kernel $f(s)$ to the data. The kernel is defined as $f(s) = f_k$ for $S_k \le s < S_{k+1}$, with $S_k = S_1 \left(\frac{S_{n+1}}{S_1}\right)^{k-1/n}$, for $1 \le k < n+1$. This defines n nonoverlapping intervals, ranging from S_1 to S_{n+1}. The distances r_{ij} from earthquake j to its parent i are normalized by the rupture lengths $L_i = 0.01 \times 10^{0.5 m_i}$ (in km): $s_{ij} = r_{ij}/L_i$. We then count the number of s_{ij} falling into each interval $S_k \le s < S_{k+1}$ and obtain the density by normalizing with the volume $\frac{4}{3}\pi\left(S_{k+1}^3 - S_k^3\right)$. We take $n = 20$ and define S_1 as $\min_{ij}\{s_{ij} > 0\}$ and $S_{n+1} = \max_{ij}\{s_{ij}\}$.

Step C. Knowing the numbers of aftershocks N_i and the model parameters, the prefactors K_i are estimated as $K_i = \frac{N_i}{e^{\alpha m_i} X_i}$.

Steps (A) and (B) are iterated until the parameter estimates converge. Step (C) is then processed. The full loop is then iterated; we found that three iterations are sufficient to obtain a stable distribution of K for catalogs made of a few tens of thousands earthquakes.

Appendix B: Estimator of the Coefficient of Variation

We here describe the estimator of the coefficient of variation, and how this estimation is performed, in an ideal case where there are no model errors, i.e., the estimated model parameters are exact, so that the exact number of aftershocks N_i is known for all main shocks i.

For a given earthquake i characterized by $\{t_i, m_i\}$, the expected number of triggered aftershocks is $\Lambda_i = K_i e^{\alpha m_i} X_i$ with $X_i = \frac{(t_{\max}+c-t_i)^{1-p} - c^{1-p}}{1-p}$ for $p \neq 1$, or $X_i = \ln(t_{\max} + c - t_i) - \ln c$ for $p = 1$. Instead of Λ_i, we observe a realization N_i of a Poisson law with mean Λ_i [*Feller*, 1968]: $N_i = \text{Poisson}(\Lambda_i)$. Our estimated $\hat{K}_i$ is defined as $\hat{K}_i = \frac{N_i}{e^{\alpha m_i} X_i}$.

Parameter K_i is a random law with mean $\overline{K}$ and standard deviation σ_K. This makes Λ_i random. We distinguish two types of averages:

1. $E\{.\}$ the sample mean, i.e., averaging over the earthquakes i;
2. $E\langle . \rangle$ the ensemble average, which correspond to averaging over independent realizations of $\{K_1, K_2, \ldots, K_n\}$, keeping $\{t_i, m_i\}$, hence $e^{\alpha m_i}$ and X_i, fixed.

In practical situations, only E type averages can be computed. We thus define estimators of $\overline{K}$ and σ_K using such $E\langle . \rangle$ type averages and calculate their biases with $E\{.\}$ type averages, both types of average being conditioned on the minimum magnitude.

First-order moment:

The sample mean of $\hat{K}_i$ is $E\left\langle \hat{K}_i \;\middle|\; m_i \ge m \right\rangle = \frac{1}{n(m)} \sum_{i/m_i \ge m} \frac{N_i}{e^{\alpha m_i} X_i}$, with $n(m)$ the number of earthquakes with $m_i \ge m$. Its ensemble average is $E\left\{E\left\langle \hat{K}_i \;\middle|\; m_i \ge m \right\rangle\right\} = \frac{1}{n(m)} \sum_{i/m_i \ge m} \frac{E\{N_i\}}{e^{\alpha m_i} X_i}$. Since $E\{N_i\} = E\{\Lambda_i\} = \overline{K} e^{\alpha m_i} X_i$,

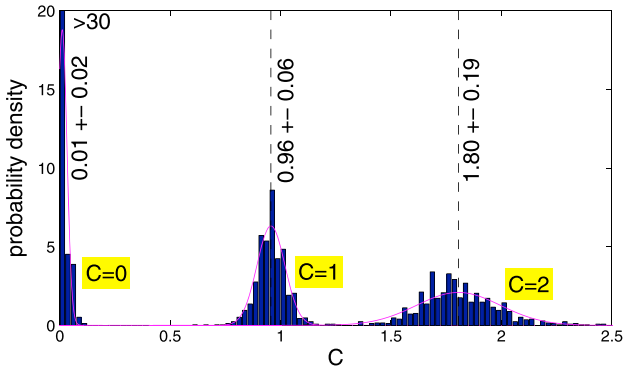

Figure B1. Estimated coefficient of variation $\hat{C}$ for 1000 realizations of a synthetic data set, for three values of $C = 0$, 1, and 2. The best Gaussian fits are shown in magenta, and the means are indicated with the dashed lines.

we simply find that $E\left\{E\left\langle \hat{K}_i \;\middle|\; m_i \geq m \right\rangle\right\} = \overline{K}$, i.e., $E\left\langle \hat{K}_i \;\middle|\; m_i \geq m \right\rangle$ is an unbiased estimator of $\overline{K}$, which random fluctuations around $\overline{K}$ decrease as $n(m)$ increase (i.e., as m decreases).

Second-order moment:

As with the first-order moment, it is straightforward to show that $E\left\{E\left\langle \hat{K}_i^{\,2} \;\middle|\; m_i \geq m \right\rangle\right\} = \frac{1}{n(m)} \sum_{i/m_i \geq m} \frac{E\{N_i^2\}}{e^{2\alpha m_i} X_i^2}$. Since N_i is a Poisson law, we get that $E\{N_i^2\} = E\{\Lambda_i^2\} + E\{\Lambda_i\} = e^{2\alpha m_i} X_i^2 \overline{K}^2 + e^{\alpha m_i} X_i \overline{K}$. Therefore, $E\left\{E\left\langle \hat{K}_i^{\,2} \;\middle|\; m_i \geq m \right\rangle\right\} = \overline{K}^2 + \overline{K} E\left\langle (e^{\alpha m_i} X_i)^{-1} \;\middle|\; m_i \geq m \right\rangle$. For $n(m) \gg 1$, the variability of E averages becomes small, and we approximate $E\left\{\frac{\sigma_{\hat{K}}^2}{\overline{K}^2} | m_i \geq m\right\}$ with $\frac{\sigma_{\hat{K}}^2}{\overline{K}^2} + \frac{1}{\overline{K}} E\left\langle (e^{\alpha m_i} X_i)^{-1} \;\middle|\; m_i \geq m \right\rangle$. The estimator $\frac{\sigma_{\hat{K}}^2}{\overline{K}}$ is thus biased, but with a bias which ensemble average can be approximated. The method we follow is thus to compute the conditioned $\frac{\sigma_{\hat{K}}^2}{\overline{K}}$ for various cutoff magnitudes m and fit the resulting $\frac{\sigma_{\hat{K}}^2}{\overline{K}}$ versus m curve (in log scale) with

$$\frac{\sigma_{\hat{K}}^2}{\overline{K}} = \hat{C}^2 + \frac{1}{\overline{K}} E\left\langle (e^{\alpha m_i} X_i)^{-1} \;\middle|\; m_i \geq m \right\rangle, \tag{B1}$$

to obtain the estimated coefficient of variation $\hat{C}$. We weight the $\frac{\sigma_{\hat{K}}^2}{\overline{K}}$ values with $\frac{n(m)}{\sigma_{\hat{K}}^2}$ to account for the decaying error at large $n(m)$.

We illustrate this method in a simplistic case with $\alpha = 2$, $p = 1.1$, and $c = 10^{-4}$, and the magnitudes m_i drawn from a Gutenberg-Richter law with $b = 1$ starting at $m = 0$. Only the main shocks are simulated, occurring at times t_i uniformly distributed between 0 and 1000. The prefactors K_i are drawn from a lognormal law with $C = \frac{\sigma_K}{\overline{K}}$ equal to 0, 1, or 2, depending on the run. We perform 1000 realizations of the data set for each value of C, each data set counting 1000 main shocks. As already indicated at the beginning of this Appendix, the estimated $\hat{N}_i$ is here taken equal to the true N_i (no model errors), i.e., we assume a perfectly accurate estimation of the numbers of aftershocks. For each catalog, we then calculate $\frac{\sigma_{\hat{K}}}{\overline{K}}$ and compute the best fit with the model of equation (B1) in log scale, i.e., we fit $\log\left(\frac{\sigma_{\hat{K}}}{\hat{K}}\right)$ with $\log\left(\sqrt{\hat{C}^2 + \frac{1}{\overline{K}} E\left\langle (e^{\alpha m_i} X_i)^{-1} \;\middle|\; m_i \geq m \right\rangle}\right)$; a L1-norm is used for this fit, as experiencing with this norm as well as with the more classical L2-norm gave marginally better results for our tests as described in section 3. We show in Figure B1 the resulting distribution of $\hat{C}$, proving that the estimation is robust.

Appendix C: Variability in *K* Caused by Fluctuations in *c*

We here investigate how a variable *c* parameter affects the estimation of the variability of *K*. We recall that $N(m)=Ke^{\alpha m}X$ (equation (4)) and that $X=\frac{c^{1-p}}{p-1}$ in the limit of $t_{\max}\to\infty$ and for $p\neq 1$. Fluctuations in *c*, for all parameters fixed, thus generate variability in *N*. Our analyzing model does not admit any variability in *c* but does so for *K*. The estimated $\hat{K}$ then depends on *c* as $\hat{K}\sim c^{p-1}$. We assume *c* to be lognormally distributed: $c=e^x$ with *x* a normal law with mean μ_x and standard deviation σ_x. The coefficient of variation C_c of *c* is then $C_c=\sqrt{e^{\sigma_x^2}-1}$. Since $\hat{K}\sim c^{p-1}$, this causes $\hat{K}$ to be lognormally distributed too: $\hat{K}=e^y$, with *y* a normal law of mean $(p-1)\mu_x$ and standard deviation $(p-1)\sigma_x$. The coefficient of variation *C* of $\hat{K}$ is $C=\sqrt{e^{\sigma_y^2}-1}=\sqrt{e^{(p-1)^2\sigma_x^2}-1}$, and thus, we finally obtain that $C=\sqrt{\left(1+C_c^2\right)^{(p-1)^2}-1}$.

Acknowledgments

We would like to thank three anonymous reviewers and the Associate Editor for their constructive reviews. The earthquake data used in this study can be downloaded at http://web.gps.caltech.edu/~wenzheng/YHS_2011_focal_catalog.html and http://scedc.caltech.edu/research-tools/alt-2011-dd-hauksson-yang-shearer.html.

References

Aki, K. (1967), Scaling law of seismic spectrum, *J. Geophys. Res.*, *72*, 1217–1231, doi:10.1029/JZ072i004p01217.

Allmann, B. P., and P. M. Shearer (2007), Spatial and temporal stress drop variations in small earthquakes near Parkfield, California, *J. Geophys. Res.*, *112*, B04305, doi:10.1029/2006JB004395.

Allmann, B. P., and P. M. Shearer (2009), Global variations of stress drop for moderate to large earthquakes, *J. Geophys. Res.*, *114*, B01310, doi:10.1029/2008JB005821.

Baltay, A., S. Ide, G. Prieto, and G. Beroza (2011), Variability in earthquake stress drop and apparent stress, *Geophys. Res. Lett.*, *38*, L06303, doi:10.1029/2011GL046698.

Boettcher, M. S., and T. H. Jordan (2004), Earthquake scaling relations for mid-ocean ridge transform faults, *J. Geophys. Res.*, *109*, B12302, doi:10.1029/2004JB003110.

Bouchon, M., and H. Karabulut (2008), The aftershock signature of supershear earthquakes, *Science*, *320*, 1323–1325, doi:10.1126/science.1155030.

Chen, X., and P. M. Shearer (2011), Comprehensive analysis of earthquake source spectra and swarms in the Salton Trough, California, *J. Geophys. Res.*, *116*, B09309, doi:10.1029/2011JB008263.

Davidsen, J., and M. Baiesi (2016), Self-similar aftershock rates, *Phys. Rev. E*, *94*, 022314.

Delahaye, E. J., J. Townend, M. E. Reyners, and G. Rogers (2009), Microseismicity but no tremor accompanying slow slip in the Hikurangi subduction zone, New Zealand, *Earth Planet. Sci. Lett.*, *277*, doi:10.1016/j.epsl.2008.09.038.

Enescu, B., S. Hainzl, and Y. Ben-Zion (2009), Correlations of seismicity patterns in Southern California with surface heat flow data, *Bull. Seismol. Soc. Am.*, *99*, 3114–3123, doi:10.1785/0120080038.

Feller, W. (1968), *An Introduction to Probability Theory and Its Applications*, 3rd ed., pp. 146–173, John Wiley, New York.

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.

Hainzl, S. (2016), Rate-dependent incompleteness of earthquake catalogs, *Seismol. Res. Lett.*, *87*, 337–344.

Hainzl, S., and D. Marsan (2008), Dependence of the Omori-Utsu law parameters on main shock magnitude: Observations and modeling, *J. Geophys. Res.*, *113*, B10309, doi:10.1029/2007JB005492.

Hainzl, S., F. Scherbaum, and C. Beauval (2006), Estimating background activity based on interevent-time distribution, *Bull. Seismol. Soc. Am.*, *96*, 313–320, doi:10.1785/0120050053.

Hainzl, S., A. Christophersen, and B. Enescu (2008), Impact of earthquake rupture extensions on parameter estimations of point-process models, *Bull. Seismol. Soc. Am.*, *98*, 2066–2072, doi:10.1785/0120070256.

Hainzl, S., O. Zakharova, and D. Marsan (2013), Impact of aseismic transients on the estimation of aftershock productivity parameters, *Bull. Seismol. Soc. Am.*, *103*, 1723–1732, doi:10.1785/0120120247.

Hainzl, S., J. Moradpour, and J. Davidsen (2014), Static stress triggering explains the empirical aftershock distance decay, *Geophys. Res. Lett.*, *41*, 8818–8824, doi:10.1002/2014GL061975.

Hauksson, E. (2015), Average stress drops of Southern California earthquakes in the context of crustal geophysics: Implications for fault zone healing, *Pure Appl. Geophys.*, *172*, 1359–1370.

Hauksson, E., W. Z. Yang, and P. M. Shearer (2012), Waveform relocated earthquake catalog for Southern California (1981 to June 2011), *Bull. Seismol. Soc. Am.*, *102*, 2239–2244.

Helmstetter, A., and B. E. Shaw (2009), Afterslip and aftershocks in the rate-and-state friction law, *J. Geophys. Res.*, *114*, B01308, doi:10.7916/D8P55ZH5.

Helmstetter, A., and M. J. Werner (2012), Adaptive spatiotemporal smoothing of seismicity for long-term earthquake forecasts in California, *Bull. Seismol. Soc. Am.*, *102*, 2518–2529.

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.

Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2006), Comparison of short-term and time-independent earthquake forecast models for southern California, *Bull. Seismol. Soc. Am.*, *96*, 90–106.

Hirabayashi, T., K. Ito, and T. Yoshii (1992), Multifractal analysis of earthquakes, *Pure Appl. Geophys.*, *138*, 591–610, doi:10.1007/BF00876340.

Hirata T., and M. Imoto (1991), Multifractal analysis of spatial distribution of microearthquakes in the Kanto region, *Geophys. J. Int.*, *107*, 155–162.

Hirose, H., H. Kimura, B. Enescu, and S. Aoi (2012), Recurrent slow slip event likely hastened by the 2011 Tohoku earthquake, *Proc. Natl. Acad. Sci. U.S.A.*, *109*, 15,157–15,161, doi:10.1073/pnas.1202709109.

Hooge, C., S. Lovejoy, D. Schertzer, S. Pecknold, J.-F. Malouin, and F. Schmitt (1994), Mulifractal phase transitions: The origin of self-organized criticality in earthquakes, *Nonlinear Processes Geophys.*, *1*, 191–197.

Kisslinger, C., and L. M. Jones (1991), Properties of aftershock sequences in Southern California, *J. Geophys. Res.*, *96*, 11,947–11,958, doi:10.1029/91JB01200.

Lange, D., J. R. Bedford, M. Moreno, F. Tilmann, J. C. Baez, M. Bevis, and F. Krueger (2014), Comparison of postseismic afterslip models with aftershock seismicity for three subduction-zone earthquakes: Nias 2005, Maule 2010 and Tohoku 2011, *Geophys. J. Int.*, *199*, 784–799, doi:10.1093/gji/ggu292.

Legrand, D., A. Cisternas, and L. Dorbath (1996), Multifractal analysis of the 1992 Erzincan Aftershock Sequence, *Geophys. Res. Lett.*, *23*, 933–936.

Lengliné, O., L. Lamourette, L. Vivin, N. Cuenot, and J. Schmittbuhl (2014), Fluid-induced earthquakes with variable stress drop, *J. Geophys. Res. Solid Earth*, *119*, 8900–8913, doi:10.1002/2014JB011282.

Lin, J., and R. S. Stein (2004), Stress triggering in thrust and subduction earthquakes and stress interaction between the southern San Andreas and nearby thrust and strike-slip faults, *J. Geophys. Res.*, *109*, B02303, doi:10.1029/2003JB002607.

Marsan, D., and O. Lengliné (2008), Extending earthquakes' reach through cascading, *Science*, *319*, 1076–1079, doi:10.1126/science.1148783.

Marsan, D., and O. Lengliné (2010), A new estimation of the decay of aftershock density with distance to the mainshock, *J. Geophys. Res.*, *115*, B09302, doi:10.1029/2009JB007119.

Marsan, D., A. Helmstetter, M. Bouchon, and P. Dublanchet (2014), Foreshock activity related to enhanced aftershock production, *Geophys. Res. Lett.*, *41*, 6652–6658, doi:10.1002/2014GL061219.

McGuire, J. J., M. S. Boettcher, and T. H. Jordan (2005), Foreshock sequences and short-term earthquake predictability on East Pacific Rise transform faults, *Nature*, *434*, 457–461, doi:10.1038/nature03377.

Moradpour, J., S. Hainzl, and J. Davidsen (2014), Nontrivial decay of aftershock density with distance in Southern California, *J. Geophys. Res. Solid Earth*, *119*, 5518–5535, doi:10.1002/2014JB010940.

Narteau, C., S. Byrdina, P. Shebalin, and D. Schorlemmer (2009), Common dependence on stress for the two fundamental laws of statistical seismology, *Nature*, *462*, 642–645, doi:10.1038/nature08553.

Oth, A. (2013), On the characteristics of earthquake stress release variations in Japan, *Earth Planet. Sci. Lett.*, *377-378*, 132–141.

Oth, A., D. Bindi, S. Parolai, and D. Di Giacomo (2010), Earthquake scaling characteristics and the scale-(in)dependence of seismic energy-to-moment ratio: Insights from KiK-net data in Japan, *Geophys. Res. Lett.*, *37*, L19304, doi:10.1029/2010GL044572.

Ouillon, G., and D. Sornette (2005), Magnitude-dependent Omori law: Theory and empirical study, *J. Geophys. Res.*, *110*, B04306, doi:10.1029/2004JB003311.

Ouillon, G., and D. Sornette (2011), Segmentation of fault networks determined from spatial clustering of earthquakes, *J. Geophys. Res.*, *116*, B02306, doi:10.1029/2010JB007752.

Ouillon, G., C. Ducorbier, and D. Sornette (2008), Automatic reconstruction of fault networks from seismicity catalogs: Three-dimensional optimal anisotropic dynamic clustering, *J. Geophys. Res.*, *113*, B01306, doi:10.1029/2007JB005032.

Pasten, D., and D. Comte (2014), Multifractal analysis of three large earthquakes in Chile: Antofagasta 1995, Valparaiso 1985, and Maule 2010, *J. Seismol.*, *18*, 707–713, doi:10.1007/s10950-014-9432-5.

Perfettini, H., and J. P. Avouac (2007), Modeling afterslip and aftershocks following the 1992 Landers earthquake, *J. Geophys. Res.*, *112*, B07409, doi:10.1029/2006JB004399.

Pollitz, F. F., and M. J. S. Johnston (2006), Direct test of static stress versus dynamic stress triggering of aftershocks, *Geophys. Res. Lett.*, *33*, L15318, doi:10.1029/2006GL026764.

Prieto, G. A., P. M. Shearer, F. L. Vernon, and D. Kilb (2004), Earthquake source scaling and self-similarity estimation from stacking *P* and *S* spectra, *J. Geophys. Res.*, *108*, B08310, doi:10.1029/2004JB003084.

Reverso, T., D. Marsan, and A. Helmstetter (2015), Detection and characterization of transient forcing episodes affecting earthquake activity in the Aleutian Arc system, *Earth Planet. Sci., Lett.*, *412*, 25–34.

Schoenberg, F. P. (2013), Facilitated estimation of ETAS, *Bull. Seismol. Soc. Am.*, *103*, 601–605, doi:10.1785/0120120146.

Shearer, P. M., G. A. Prieto, and E. Hauksson (2006), Comprehensive analysis of earthquake source spectra in southern California, *J. Geophys. Res.*, *111*, B06303, doi:10.1029/2005JB003979.

Tahir, M., and J. R. Grasso (2014), Aftershock patterns of $M_s > 7$ earthquakes in the India-Asia collision belt: Anomalous results from the Muzaffarabad Earthquake Sequence, Kashmir, 2005, *Bull. Seismol. Soc. Am.*, *104*, 1–23, doi:10.1785/0120120158.

Tahir, M., and J. R. Grasso (2015), Faulting style controls for the space-time aftershock patterns, *Bull. Seismol. Soc. Am.*, *105*, 2480–2497, doi:10.1785/0120140336

Utsu, T. (2002), Statistical features of seismicity, in *Earthquake and Engineering Seismology*, edited by W. H. K. Lee et al., pp. 719–732, Academic Press, Amsterdam.

van der Elst, N. J., and B. E. Shaw (2015), Larger aftershocks happen farther away: Nonseparability of magnitude and spatial distributions of aftershocks, *Geophys. Res. Lett.*, *42*, 5771–5778, doi:10.1002/2015GL064734.

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.

Vidale, J. E., K. L. Boyle, and P. M. Shearer (2006), Crustal earthquake bursts in California and Japan: Their patterns and relation to volcanoes, *Geophys. Res. Lett.*, *33*, L20313, doi:10.1029/2006GL027723.

Wang, Q., F. P. Schoenberg, and D. D. Jackson (2010), Standard errors of parameter estimates in the ETAS model, *Bull. Seismol. Soc. Am.*, *100*, 1989–2001, doi:10.1785/0120100001.

Werner, M. J., and D. Sornette (2008), Magnitude uncertainties impact seismic rate estimates, forecasts, and predictability experiments, *J. Geophys. Res.*, *113*, B08302, doi:10.1029/2007JB005427.

Werner, M. J., A. Helmstetter, D. D. Jackson, and Y. Y. Kagan (2011), High-resolution long-term and short-term earthquake forecasts for California, *Bull. Seismol. Soc. Am.*, *101*, 1630–1648, doi:10.1785/0120090340.

Wiemer, S., and K. Katsumata (1999), Spatial variability of seismicity parameters in aftershock zones, *J. Geophys. Res.*, *104*, 13,135–13,151, doi:10.1029/1999JB900032.

Woessner, J., S. Hainzl, W. Marzocchi, M. J. Werner, A. M. Lombardi, F. Catalli, B. Enescu, M. Cocco, M. C. Gerstenberger, and S. Wiemer (2011), A retrospective comparative forecast test on the 1992 Landers sequence, *J. Geophys. Res.*, *116*, B05305, doi:10.1029/2010JB007846.

Yang, W., and Y. Ben-Zion (2009), Observational analysis of correlations between aftershock productivities and regional conditions in the context of a damage rheology model, *Geophys. J. Int.*, *177*, 481–490, doi:10.1111/j.1365-246X.2009.04145.x.

Zaliapin, I., and Y. Ben-Zion (2013), Earthquake clusters in southern California I: Identification and stability, *J. Geophys. Res. Solid Earth*, *118*, 2847–2864, doi:10.1002/jgrb.50179.

Zhuang, J., Y. Ogata, and D. Vere-Jones (2002), Stochastic declustering of space-time earthquake occurrences, *J. Am. Stat. Assoc.*, *97*, 369–380, doi:10.1198/016214502760046925.

Zhuang, J., M. J. Werner, S. Hainzl, D. Harte, and S. Zhou (2011), Basic models of seismicity: Spatiotemporal models, Community Online Resource for Statistical Seismicity Analysis, doi:10.5078/corssa-07487583. [Available at http://www.corssa.org.]

Zhuang, J. C., M. J. Werner, and D. S. Harte (2013), Stability of earthquake clustering models: Criticality and branching ratios, *Phys. Rev. E*, *88*, 062109, doi:10.1103/PhysRevE.88.062109.