

## Bayesian analysis of the modified Omori law

M. Holschneider,<sup>1</sup> C. Narteau,<sup>2</sup> P. Shebalin,<sup>2,3</sup> Z. Peng,<sup>4</sup> and D. Schorlemmer<sup>5</sup>

Received 23 November 2011; revised 11 May 2012; accepted 11 May 2012; published 30 June 2012.

[1] In order to examine variations in aftershock decay rate, we propose a Bayesian framework to estimate the  $\{K, c, p\}$ -values of the modified Omori law (MOL),  $\lambda(t) = K(c + t)^{-p}$ . The Bayesian setting allows not only to produce a point estimator of these three parameters but also to assess their uncertainties and posterior dependencies with respect to the observed aftershock sequences. Using a new parametrization of the MOL, we identify the trade-off between the  $c$  and  $p$ -value estimates and discuss its dependence on the number of aftershocks. Then, we analyze the influence of the catalog completeness interval  $[t_{\text{start}}, t_{\text{stop}}]$  on the various estimates. To test this Bayesian approach on natural aftershock sequences, we use two independent and non-overlapping aftershock catalogs of the same earthquakes in Japan. Taking into account the posterior uncertainties, we show that both the handpicked (short times) and the instrumental (long times) catalogs predict the same ranges of parameter values. We therefore conclude that the same MOL may be valid over short and long times.

**Citation:** Holschneider, M., C. Narteau, P. Shebalin, Z. Peng, and D. Schorlemmer (2012), Bayesian analysis of the modified Omori law, *J. Geophys. Res.*, 117, B06317, doi:10.1029/2011JB009054.

### 1. Introduction

[2] Aftershocks are the most prominent expression of the global relaxation process induced by abrupt perturbations of the state of stress in the neighborhood of seismic ruptures. More than 100 years ago, *Omori* [1894] provided the first quantitative description of an aftershock decay rate, documenting the number of earthquakes triggered by the M8 Nobi earthquake (18 October 1891, Honshu, Japan). To more accurately model the diversity of aftershock decay rates that had been reported later, *Utsu* [1961] converted the hyperbolic behavior observed by *Omori* [1894] into the so-called modified Omori law (MOL):

$$\lambda(t) = \frac{K}{(t + c)^p}, \quad (1)$$

where  $\lambda$  is the aftershock frequency within a given magnitude range,  $t$  the time from the triggering event (the so-called main shock),  $K$  the productivity of the aftershock

sequence,  $p$  the power law exponent, and  $c$  the time delay before the onset of the power-law aftershock decay rate. For more than a century of instrumental data, this empirical power-law relationship has been successfully used from the centimeter-scale of laboratory experiments [*Scholz*, 1968; *Ojala et al.*, 2004] to the scale of plate tectonics to reproduce aftershock frequencies over a wide range of time scales [*Utsu et al.*, 1995; *Stein and Liu*, 2009].

[3] Even though other models have been proposed [*Ogata*, 1988; *Otsuka*, 1985; *Kisslinger*, 1993; *Narteau et al.*, 2002; *Gasperini and Lolli*, 2009], the MOL still remains the simplest and the most widely used formula to examine the major temporal properties of aftershock sequences. Usually  $\{K, c, p\}$ -values of the MOL are obtained with the maximum likelihood point estimator method and their variability are evaluated by asymptotic statistics using the Fischer information matrix [*Ogata*, 1983]. Nevertheless, these variability cannot be considered as parameter uncertainties but rather as the variability of the estimator under surrogate data as obtained by the bootstrapping process. If only one single realization is available, bootstrapping is a common technique to simulate the sampling from the unknown underlying data generating distribution. In any case the quantification of the variability of the estimator answers the question: “by how much is the estimate likely to change, when an independent data set becomes available?”. Another concept of uncertainty would rather try to answer the following question: “how many models can explain my data equally well?”. The latter is the parameter uncertainty as provided by the Bayesian calculus.

[4] In the last decade,  $p$ -values have been estimated and discussed with respect to different sources of heterogeneity in aftershock zones [*Kisslinger and Jones*, 1991; *Nanjo et al.*, 1998; *Wiemer and Katsumata*, 1999; *Ouillon and Sornette*,

<sup>1</sup>Institute of Mathematics, Center for the Dynamics of Complex Systems, Universität Potsdam, Potsdam, Germany.

<sup>2</sup>Equipe de Dynamique des Fluides Géologiques, Institut de Physique du Globe de Paris, Sorbonne Paris Cité, UMR 7154 CNRS, Université Paris Diderot, Paris, France.

<sup>3</sup>International Institute of Earthquake Prediction Theory and Mathematical Geophysics, Moscow, Russia.

<sup>4</sup>School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, Georgia, USA.

<sup>5</sup>GFZ German Research Center for Geosciences, Potsdam, Germany.

Corresponding author: M. Holschneider, Institute of Mathematics, Center for the Dynamics of Complex Systems, Universität Potsdam, PO Box 601553, D-14115 Potsdam, Germany. (hols@math.uni-potsdam.de)

2005; Helmstetter and Shaw, 2006]. In addition, *Hauksson and Jones* [1988] proposed that the  $K$ -value is related to the stress drop of the main shock, matching the observation that the productivity is directly related to the magnitude of the main shock [*Hainzl and Marsan*, 2008]. Others have suggested that such a productivity may be inversely proportional to sediment thickness and heat flow [*Yang and Ben-Zion*, 2009; *Enescu et al.*, 2009]. However, less attention has been devoted to explaining the variability of the  $c$ -values. This is certainly because  $c$ -value estimates are very sensitive to catalogue completeness in the immediate time after the main shock [*Kagan*, 2004]. Consequently, data completeness artifacts can be introduced by overlapping seismograms during coda waves, saturation of acquisition methods, and absence or malfunction of stations [*Vidale et al.*, 2004; *Peng et al.*, 2006]. Nevertheless, analyzing catalogues of seismicity above completeness magnitude thresholds, several studies have pointed out that the  $c$ -value may have a physical origin [*Narteau et al.*, 2002; *Shcherbakov et al.*, 2004]. For example, *Narteau et al.* [2009] suggest that the  $c$ -value can be described as a characteristic time for failure and that, in this case, early aftershocks may be used as a source of information about the state of stress of the brittle crust along active fault zones.

[5] In rate-and-state friction, damage mechanics, and static fatigue models, the same physical ingredient is responsible for a non power-law aftershock decay rate over short time [*Scholz*, 1968; *Dieterich*, 1994; *Narteau et al.*, 2002; *Shcherbakov and Turcotte*, 2004; *Ben-Zion and Lyakhovskiy*, 2006]. In all cases, the amplitude of the step-like perturbation of the stress field is inversely proportional to the time delay before the onset of the power-law regime. Then, considering that this stress perturbation depends not only on the size of the event but also on the initial loading along the entire fault zone [*Atkinson*, 1991], the characteristic time delay before the power-law regime can be used to infer the pre-existing level of stress at the length scale of the main shock rupture [*Narteau et al.*, 2005, 2008]. Clearly, this characteristic time is short ( $\ll$  day) and decreases with the magnitude of the main shock. For this reason, it is necessary to use a specific range of small magnitude main shocks to reduce the influence of the co-seismic stress perturbation and limit instrumental artifacts (i.e., overlapping seismograms). Thus, using the largest aftershocks of small magnitude main shocks, the characteristic time before the power-law aftershock decay rate may be estimated using parametric [*Narteau et al.*, 2009] or non-parametric models [*Shebalin et al.*, 2011]. Most importantly to our present concern, the variability of these estimations and the mutual dependence between this time delay  $c$  and the power-law exponent  $p$  has also to be evaluated.

[6] The Bayesian approach can be described as an extension of binary logic (true/false) to a quantitative expression of the strength of belief [*Cox*, 1946]. In fact, instead of extracting from the data a single set of model parameter values (i.e., the best-fitting values), Bayesian statistics is designed to measure the degree of belief of each set of values that span the model parameter space in light of the observations. Theoretically, Bayesian calculus is the only possible extension of binary logic to continuous truth values [*Jaynes*, 2005]. However, because this approach requires to explicitly consider a priori knowledge about the studied

system, any solution has to be considered in view of this subjective prior distribution.

[7] To learn from data using a model, frequentist and Bayesian approaches are complementary to one another. The frequentist approach is based on an estimator, for example the maximum likelihood estimator, producing a value from a random realization of the process. This value should be close to the true value of the parameter and the quality of the estimator is quantified in terms of its bias (i.e., its deviation from the true value averaged from many realizations) and its variance (i.e., the dispersion of the estimation averaged from many realizations). The frequentist approach is therefore based on the fictive idea that many independent experiments can be performed, even though there is only one realization. The Bayesian approach is only based on this unique realization. The entire knowledge about the system is represented by an ensemble of models together with a distribution of credibility. Then, observations change this a priori belief into a modified posterior belief giving more credence to some models and less to others. The uncertainties in the Bayesian thinking are simply the width of the posterior distribution. This can be quantified in terms of its variance. Nevertheless, it is always better to work with the full posterior distribution instead of reducing it to a Gaussian like object that may be quantified in terms of mean and variance. In a Bayesian setting, the posterior distribution determines the range of models that explain the observations. In a frequentist setting the variance of the estimator determines by which amount the proposed value for the parameter is fluctuating from one realization to another. In the limit of a large number of independent observations both approaches become asymptotically equivalent since the posterior distribution and the distribution of the maximum likelihood estimator converge to the same Gaussian distribution. However, for a small number of observations, the two uncertainties are not equal and have different meaning (e.g., *Leonard and Hsu* [1999] for a general introduction to Bayesian formalism and, e.g., *Le Cam* [1986] for asymptotic properties). In conclusion, Bayesian techniques do not provide a better or more precise way of estimating the unknown parameters. They constitute an alternative way to deal with uncertainties in a self-consistent way. For this reason, there is no point of comparing the quality of the estimates provided by frequentist and Bayesian approaches [see, e.g., *Berger et al.*, 1997].

[8] Concerning temporal properties of aftershock sequences, there is still not a common Bayesian framework to complement the classical maximum likelihood approach [*Ogata*, 1983]. Hence, we present here a new parametrization of the MOL to separate the aftershock productivity from the shape of the aftershock decay rate, especially over short time. Then, we apply Bayesian statistics to this new parametrization to produce the Bayesian posterior distribution of the MOL parameters as well as a Bayesian assessment of their uncertainty across the entire parameter space of the model. Using such a posterior knowledge, we can investigate the systematic dependency between estimates of the power exponent  $p$  and the time delay  $c$ . Finally, we illustrate our approach with the analysis of an aftershock catalogue that combines hand-picked and instrumental data in Japan. Thus, we can discuss the Bayesian approach and aftershock

properties from two independent data sets covering different time periods of the same aftershock sequences.

## 2. The Modified Omori Law in a Bayesian Setting

[9] The MOL (equation (1)) can be rewritten as

$$\lambda(t) = \frac{\tilde{K}}{(1+t/c)^p}, \quad t > 0. \quad (2)$$

where  $\tilde{K} = K/c^p$  is the productivity with units of frequency. In this parametrization, all parameters are connected in the sense that any information about one of them changes the knowledge that we have about the others. As a consequence, the three parameters  $\{\tilde{K}, c, p\}$  may not be estimated independently. In what follows, we propose a new parametrization in which we can separate two groups of parameters, which carry independent information.

### 2.1. Reparametrization of the Modified Omori Law

[10] When the MOL is applied to real data, we cannot actually assume that it holds for all times due to observational shortcomings. We rather have to consider a time interval  $[t_{\text{start}}, t_{\text{stop}}]$  over which the MOL describes the observed data. The lower bound is due to catalogue incompleteness shortly after the main shock, essentially because of overlapping seismograms. The upper bound has also to be considered because eventually aftershocks cannot be distinguished from background seismicity anymore. Here, we do not estimate this time interval, but we suppose that it is known prior to the analysis of the three-parameter family of the MOL. For a given  $[t_{\text{start}}, t_{\text{stop}}]$ , we may reparametrize the MOL for  $p \neq 1$ ,

$$\lambda(t) = \Lambda \frac{D(c, p)}{(1+t/c)^p}, \quad t > 0, \quad \text{with} \\ D(c, p) = \frac{(1-p)/c}{(1+t_{\text{stop}}/c)^{1-p} - (1+t_{\text{start}}/c)^{1-p}}, \quad (3)$$

and for the special case  $p = 1$ ,

$$\lambda(t) = \Lambda \frac{D(c, p=1)}{1+t/c}, \quad t > 0, \quad \text{with} \\ D(c, p=1) = \frac{1/c}{\log(1+t_{\text{stop}}/c) - \log(1+t_{\text{start}}/c)}, \quad (4)$$

where  $\log$  denotes the natural logarithm to the base  $e$ . In these parametrizations,  $\Lambda$  is the Poissonian rate of the total number of aftershocks in the observation interval  $[t_{\text{start}}, t_{\text{stop}}]$ , i.e., the expected number of events during this time interval. Therefore,  $\Lambda$  is dimensionless. The constant  $D(c, p)$  ensures that for  $t \in [t_{\text{start}}, t_{\text{stop}}]$ , the two-parameter family

$$\tilde{\lambda}_{c,p}(t) = \frac{D(c, p)}{(1+t/c)^p} \quad (5)$$

is a probability density function (pdf) of the event times  $t$  since the main shock in the observation interval. The mapping between the  $\{\tilde{K}, c, p\}$  parametrization and the new one  $\{\Lambda, c, p\}$  is

$$\Lambda = \mathbb{E}(\text{number of aftershocks in } [t_{\text{start}}, t_{\text{stop}}]) = \frac{\tilde{K}}{D(c, p)}, \quad (6)$$

so that  $c$  and  $p$  maintain their meaning. The advantage of this parametrization is that the productivity is now decoupled from the other parameters  $c$  and  $p$  as we shall see below. The parameters  $c$  and  $p$  describe the shape of the event time distribution, whereas the productivity  $\Lambda$  describes the expected number of events in the observation interval. Meanwhile, the parameters  $c$  and  $\Lambda$  may still depend on lower cutoff magnitude as it is the case for the MOL. The drawback of this new parametrization is that the productivity is now defined with respect to a fixed time interval. This is not a parameter of the model but an observational constraint that has to be determined carefully with respect to the catalogue completeness. Note, however, that for fixed  $c$  and  $p$ -values the productivity  $\Lambda'$  of another time interval  $[t'_{\text{start}}, t'_{\text{stop}}]$  is simply given by

$$\Lambda' = \Lambda \frac{D'(c, p)}{D(c, p)}, \quad (7)$$

where  $D'$  is computed using  $t'_{\text{start}}$  and  $t'_{\text{stop}}$ .

### 2.2. The Likelihood Function of the Parameters of the Modified Omori Law

[11] Suppose we have observed  $n$  occurrences of aftershocks at times  $t_{\text{start}} \leq t_1 < t_2 < \dots < t_n \leq t_{\text{stop}}$  in the observation window. The likelihood of the parameters in view of these observations, i.e., the probability to observe  $n$  aftershocks at time  $\{t_i\}_{i=1,2,\dots,n}$ , factorizes as follows:

$$L(\Lambda, c, p | \{t_i\}_{i=1,\dots,n}) = \mathbb{P}(\{t_i\}_{i=1,\dots,n} | \Lambda, c, p) = \mathbb{P}(\{t_i\} | c, p, n) \mathbb{P}(n | \Lambda), \quad (8)$$

where  $\{t_i\}$  is any finite sequence of events and  $\mathbb{P}(n | \Lambda)$  the Poisson distribution of frequency  $\Lambda$ . Accordingly, to generate a random sequence of events one first draws the number  $n$  of events according to a Poisson distribution with rate  $\Lambda$  and then draws  $n$  independent trials from the distribution of times with pdf  $\tilde{\lambda}_{c,p}$  (see equation (5)). Therefore, we may write

$$L(\Lambda, c, p | \{t_i\}_{i=1,\dots,n}) = \frac{\Lambda^n e^{-\Lambda}}{n!} \times D(c, p)^n \prod_{i=1}^n \frac{1}{(1+t_i/c)^p}. \quad (9)$$

The first factor is the Poissonian probability to observe  $n$  aftershocks given a rate of  $\Lambda$ . The second factor is the likelihood factor, given the fact that we have  $n$  events distributed according to the shape distribution of parameters  $\{c, p\}$  in the time interval  $[t_{\text{start}}, t_{\text{stop}}]$ . By the Fischer-Neyman factorization theorem this implies that the number  $n$  of observed events is a sufficient statistics for  $\Lambda$  [Scharf, 1991]. This means that all information about this parameter is contained in the mere number of observed events. The precise times of these aftershocks do not carry any information about the productivity  $\Lambda$ . However, the exact observed times carry all the information about the shape parameters  $c$  and  $p$ . Note that in the traditional parametrization there would be no sufficient statistics, and the times would also influence the estimation of  $K$ . Then, it is important to emphasize that this reparametrization is useful even in a frequentist setting, where maximum likelihood point estimators are considered.

[12] On a log scale the part of the likelihood function that carries information about  $c$  and  $p$  reads

$$-\log L(c, p | \{t_i\}_{i=1, \dots, n}) = -n \log D(c, p) + p \sum_{i=1}^n \log(1 + t_i/c). \quad (10)$$

The minimum of this function yields the maximum-likelihood estimator  $c^*$  and  $p^*$  for  $c$  and  $p$ , respectively.

### 2.3. Asymptotic Behaviors

[13] For later use we discuss now the asymptotic behavior of the likelihood function in various regions of the parameter space. This discussion will include an inventory of all pdfs that are limit points of the MOL family. The following subfamilies are limit cases of the MOL's time distributions  $\tilde{\lambda}_{p,c}$  in  $[t_{\text{start}}, t_{\text{stop}}]$ :

- [14] 1. **Region I:** uniform distribution  $\sim 1$
- [15] 2. **Region II:** exponential distribution  $\sim e^{-\alpha t}$
- [16] 3. **Region III:** Dirac distribution  $\sim \delta_{t_{\text{start}}}$
- [17] 4. **Region IV:** power law distribution  $\sim t^{-p}$

These different regions correspond in essence to the behavior of the direction  $\alpha = p/c$ . The reason is that we may write

$$(1 + t/c)^{-p} = ((1 + t/c)^c)^{-\alpha} = \left( (1 + t/(p/\alpha))^{p/\alpha} \right)^{-\alpha}. \quad (11)$$

Now we may use the well-known expression

$$(1 + t/u)^u \rightarrow e^t, \quad (u \rightarrow \infty) \quad (12)$$

to understand the various limit behaviors in the  $\{c, p\}$  space.

#### 2.3.1. Region I

[18] Any path along which we have

$$\left( \frac{1 + t_{\text{stop}}/c}{1 + t_{\text{start}}/c} \right)^p \rightarrow 1 \quad (13)$$

the MOL tends towards a uniform distribution in the observation interval:

$$\tilde{\lambda}_{c,p}(t) \rightarrow \frac{1}{t_{\text{stop}} - t_{\text{start}}} \chi_{[t_{\text{start}}, t_{\text{stop}}]}(t), \quad (14)$$

where  $\chi_{[t_{\text{start}}, t_{\text{stop}}]}$  denotes the characteristic function of the interval  $[t_{\text{start}}, t_{\text{stop}}]$ , 1 inside and 0 outside. This is precisely the condition under which the probability density at the two endpoints becomes equal. From the monotonicity of the law over this interval, the limit is necessarily a uniform distribution.

[19] A sufficient condition is that  $p \rightarrow 0$  while  $c$  is behaving arbitrary. Another possible collection of path is  $c \rightarrow \infty$  and  $p \rightarrow \infty$  but not faster growing than  $o(c)$  so that  $\alpha = p/c \rightarrow 0$ . In Region I we have a zero-parameter family; all members are the same.

#### 2.3.2. Region II

[20] Any path  $c \rightarrow \infty$ ,  $p \rightarrow \infty$  in such a way that  $\alpha = p/c$  converges to some limit  $\alpha_\infty$  we obtain an exponential distribution

$$\tilde{\lambda}_{c,p}(t) \rightarrow F_1 e^{-\alpha_\infty t}. \quad (15)$$

The normalization constant  $F_1$  makes this a pdf over  $[t_{\text{start}}, t_{\text{stop}}]$ . In this region we have a one-parameter family. Note that Region I can be formally obtained as the limit  $\alpha_\infty \rightarrow 0$  in which case the exponential distribution becomes uniform.

#### 2.3.3. Region III

[21] Along any path for which for all  $t > t_{\text{start}}$  we have

$$\left( \frac{c + t_{\text{start}}}{c + t} \right)^p \rightarrow \infty, \quad (16)$$

the MOL distribution of time points tends towards a delta function at  $t_{\text{start}}$

$$\tilde{\lambda}_{c,p}(t) \rightarrow \delta_{t_{\text{start}}}(t). \quad (17)$$

This is equivalent to  $p \rightarrow \infty$  and  $\alpha = p/c \rightarrow \infty$ . Note again that this is a formal end-member case of Region II.

#### 2.3.4. Region IV

[22] Along any path  $c \rightarrow 0$  and  $p$  converging to some limit, the distributions become a power law

$$\tilde{\lambda}_{c,p}(t) \rightarrow t^{-p}. \quad (18)$$

This one-parameter family of distributions cannot be obtained as a formal limit of Region II. However, Region II contains Region IV as an end-member case as  $p \rightarrow \infty$ .

[23] As a consequence, in Regions I, II and IV, the likelihood function will tend to a finite value  $\neq 0$ . This limiting value is the likelihood of the limit distribution. All models close to the limiting point have the same positive credibility and cannot be distinguished from the data alone. On the other hand, in Region III, the likelihood of the limit function is 0 and the likelihood function decays to 0.

[24] These asymptotic behavior are important to understand the entire range of submodels encompassed in the MOL. For example, the MOL can be used to fit arbitrarily well an exponential decay rate (see Region II).

### 2.4. Bayesian Posterior Distribution of the Modified Omori Law Parameters

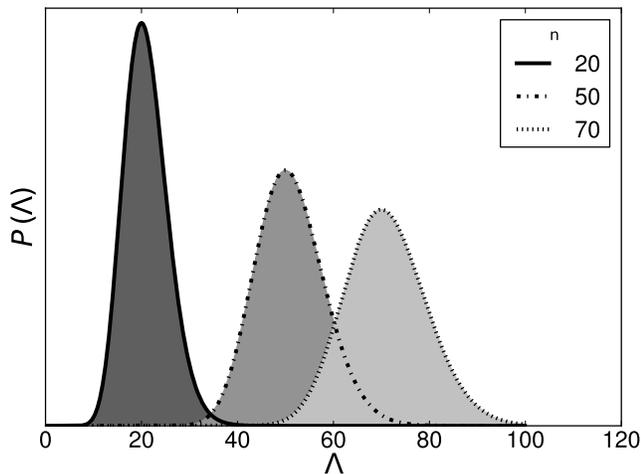
[25] Suppose we have some prior beliefs about the parameters  $\{\Lambda, c, p\}$ . This is expressed by a prior distribution  $\mathbb{P}_0(\Lambda, c, p)$  of the parameters. We assume that our prior belief may be factorized as

$$\mathbb{P}_0(\Lambda, c, p) = \mathbb{P}_0(\Lambda) \mathbb{P}_0(c, p) = \mathbb{P}_0(\Lambda) \mathbb{P}_0(c) \mathbb{P}_0(p). \quad (19)$$

This means that our a prior beliefs about  $\Lambda$  and  $\{c, p\}$  are independent. After making the observations  $\{t_i\}_{i=1, \dots, n}$ , our knowledge about the parameters has changed. The posterior distribution of parameter values is given by Bayes' theorem as

$$\begin{aligned} \mathbb{P}(\Lambda, c, p | \{t_i\}_{i=1, \dots, n}) &= F_2 \frac{\Lambda^n e^{-\Lambda}}{n!} \mathbb{P}_0(\Lambda) \prod_{i=1}^n \frac{D(c, p)}{(1 + t_i/c)^p} \mathbb{P}_0(c) \mathbb{P}_0(p) \\ &= \mathbb{P}(\Lambda | n) \mathbb{P}(c, p | \{t_i\}_{i=1, \dots, n}). \end{aligned} \quad (20)$$

The normalization constant  $F_2$  ensures that the integral over  $d\Lambda dc dp$  is unity so that this corresponds to a probability distribution in the parameter space. As expected, the posterior factorizes into two independent contributions.



**Figure 1.** Posterior density of  $\Lambda$  for different number of aftershocks. This parameter represents the productivity of an aftershock sequence and its estimation depends only on the number of events.

#### 2.4.1. Posterior Distribution of $\Lambda$

[26] Treating  $c$  and  $p$  as nuisance parameters and integrating them out, we obtain the posterior distribution of  $\Lambda$ . Thanks to the chosen parametrization, the posterior information about  $\Lambda$  takes the following simple form:

$$\mathbb{P}(\Lambda|n) = \int \int \mathbb{P}(\Lambda, c, p | \{t_i\}_{i=1, \dots, n}) dc dp = F_3 \Lambda^n e^{-\Lambda} \mathbb{P}_0(\Lambda), \quad (21)$$

where  $F_3$  is another normalization constant. From this equation, we see that only the number  $n$  of observed aftershocks influences the gain of knowledge about  $\Lambda$ . Then,  $n$  is a sufficient statistics for  $\Lambda$ . No information about  $c$  or  $p$  influences our knowledge about  $\Lambda$ . If nothing is known a priori about  $\Lambda$ , we may use a flat prior  $\mathbb{P}_0(\Lambda) = 1$ , or a scale invariant Jeffreys prior  $\mathbb{P}_0(\Lambda) = 1/\Lambda^{1/2}$ . In this latter case, the posterior is a Gamma distribution with parameter  $n$ :

$$\mathbb{P}(\Lambda|n) \sim \frac{\Lambda^{n-1/2} e^{-\Lambda}}{\Gamma(n+1/2)}. \quad (22)$$

Its mean value is  $n$  and its variance is  $n+1/2$  (Figure 1).

#### 2.4.2. Posterior Distribution of $\{c, p\}$

[27] The posterior distribution of  $c$  and  $p$  poses slightly more problems. From equation (20), we see that it can be written as

$$\mathbb{P}(c, p | \{t_i\}_{i=1, \dots, n}) = F_4 \prod_{i=1}^n \frac{D(c, p)}{(1+t_i/c)^p} \mathbb{P}_0(c) \mathbb{P}_0(p). \quad (23)$$

The problem is that a flat, uninformative prior in  $c$  and  $p$  is not possible because this would lead to a posterior which is not normalizable. As we have shown in section 2.3, the MOL tends to regular limit distributions as  $p$  and  $c$  get large while  $\alpha = p/c$  tends to some finite value. This implies, that

the likelihood function for a fixed set of observations tends towards a constant, which is different from 0 along any such path. In Region III, however, the limit distribution is singular, which gives 0 credibility to the data and the likelihood function decays rapidly. For  $p \rightarrow \infty$  and  $c \rightarrow \infty$ , such a behavior impedes a posterior based on a flat prior for  $c$  and  $p$  to be normalizable, which however is necessary to make it a probability distribution. A similar problem arises in the truncated Gutenberg-Richter distribution [Holschneider et al., 2011]. The point is that, in the light of finite data, all models close to one of the end-member cases look all the same. They cannot be distinguished from the finite sample available. Taking a flat prior for  $c$  and  $p$  would give infinite weight to essentially always the same model. We therefore need to introduce some prior information upon which we base our analysis. Obviously, the posterior will depend on this prior information. However, over the last century of reported data, we can determine bounds on  $p$  and  $c$  values for which the dependency is rather mild. Furthermore, in the limit of large  $n$  the influence of the prior information will tend to zero.

[28] We suggest using an informative prior, which puts some upper bound on  $c$  so that

$$\mathbb{P}_0(c) = \frac{1}{C} \chi_{[0, C]}(c), \quad (24)$$

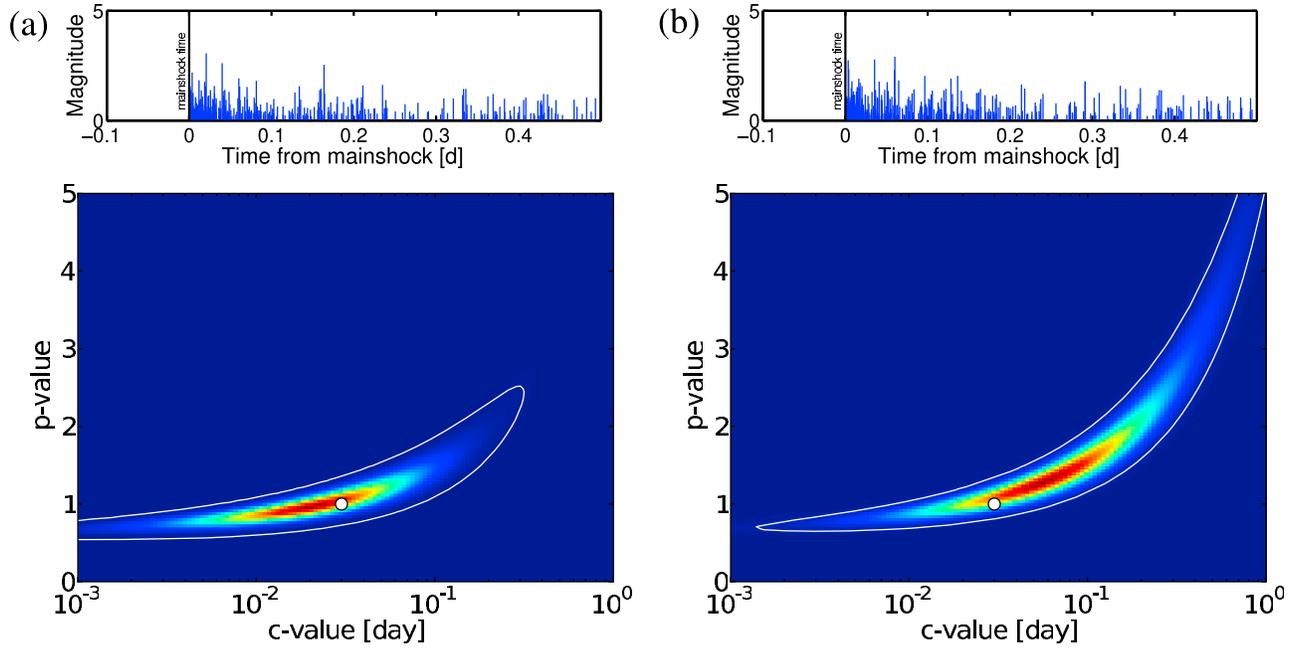
where  $\chi_{[0, C]}$  is the characteristic function of the interval  $[0, C]$ , i.e., 1 inside and 0 outside. In this case the posterior is normalizable because the likelihood function decays as  $p \rightarrow \infty$  (Region III) and we may use a flat prior for  $p$ . However, in practical applications, we will also use some a priori upper bound for  $p$ . Thus, we work in some a priori uniform density in a rectangular parameter space  $\{c, p\}$ . Finally, given our flat prior information on  $\{c, p\}$ , the maximum of the posterior density distribution (i.e., the maximum posterior point estimator) is located at the estimates provided by the maximum likelihood method, at least as long as it is inside the prior box and not on its border.

#### 2.5. Asymptotic Posterior for Large $n$

[29] In Figure 2 we have plotted the posterior distribution for two synthetic time series generated for a MOL with parameters  $p = 1$ ,  $c = 100$  s,  $t_{\text{start}} = 100$  s and  $t_{\text{stop}} = 0.5$  d. The actual posterior information we can draw from the samples is a random function itself since it depends on the random realization of time points. Nevertheless, in the limit of a large number  $n$  of events, the posterior distributions will be close to a common limit function, as we shall see now.

[30] Moreover, Figure 2 shows that there is a strong posterior dependency between the two parameters. Imposing one of them, will influence the estimation of the second. This posterior dependency can be understood in the large sample limit, where the posterior covariance may be expressed in terms of the Kullback-Leibler distance between the different model distributions. Asymptotically the posterior distribution becomes Gaussian with the covariance function given by the Fischer Information matrix, which is the Hessian of the Kullback-Leibler distance function.

[31] Let us consider the limit, as the number of aftershocks approaches infinity. Suppose, the true data generating model



**Figure 2.** The posterior density of  $\{c, p\}$  using a logarithmic scale for  $c$ . (a, b) The sequences corresponding to different realizations generated from 300 events drawn randomly from the MOL with  $p = 1$  and  $c = 0.02$  d (the white dot in the image) for  $10^{-4}$  d  $< t < 1$  d. The white contour line delimits the 95% isoline. In this figure, as in all the following figures showing posterior density distributions, the color bar is linear between the minimum (blue) and the maximum (red) values.

has parameters  $c'$  and  $p'$ . We now consider the posterior distribution  $\mathbb{P}(c, p | \{t_i\}_{i=1, \dots, n})$ . We have

$$\begin{aligned} \frac{1}{n} \log \mathbb{P}(p, c | \{t_i\}_{i=1, \dots, n}) &\sim \log(D(c, p)) - \frac{1}{n} \sum_{i=1}^n p \log(1 + t_i/c) \\ &+ \frac{1}{n} \log \mathbb{P}_0(c, p) \\ &\sim \frac{1}{n} \sum_{i=1}^n \log L(p, c | \{t_i\}_{i=1, \dots, n}) + \frac{1}{n} \log \mathbb{P}_0(c, p). \end{aligned} \quad (25)$$

Since the samples  $\{t_i\}$  are drawn from a MOL distribution  $\tilde{\lambda}_{c', p'}$  with parameters  $\{c', p'\}$ , the sum can be replaced by the weighted mean and, in the average for large  $n$ , we obtain

$$-\frac{1}{n} \log \mathbb{P}(p, c | \{t_i\}_{i=1, \dots, n}) \simeq -\log(D(c, p)) + pD(c', p') - \int_{t_{\text{start}}}^{t_{\text{stop}}} \frac{\log(1 + t/c)}{(1 + t/c')^{p'}} dt. \quad (26)$$

For fixed  $c', p'$  this function of  $c$  and  $p$  is up to a constant that depends on  $c', p'$ , asymptotically equal to the Kullback-Leibler (KL) divergence between the pdf of time points  $\tilde{\lambda}_{c, p}$  and  $\tilde{\lambda}_{c', p'}$

$$D_{KL}(\tilde{\lambda}_{c', p'} || \tilde{\lambda}_{c, p}) = \mathbb{E}_{\tilde{\lambda}_{c', p'}} \log(\tilde{\lambda}_{c', p'} / \tilde{\lambda}_{c, p}) = -\mathbb{E}_{\tilde{\lambda}_{c', p'}} \log \tilde{\lambda}_{c, p} - E_{c', p'}, \quad (27)$$

where  $E_{c', p'} = \mathbb{E}_{\tilde{\lambda}_{c', p'}} \log(\tilde{\lambda}_{c', p'})$  is the entropy of  $\tilde{\lambda}_{c', p'}$  [Kullback and Leibler, 1951; Bishop, 2006]. This quantity  $D_{KL}$  is a

measure for how distinguishable the distribution  $\tilde{\lambda}_{c, p}$  is from the distribution  $\tilde{\lambda}_{c', p'}$  when the sampling times are generated from the latter (Figure 3a).

[32] In the limit of large  $n$ , we may express the posterior distribution in terms of the KL distance as follows:

$$\begin{aligned} \mathbb{P}(p, c | \{t_i\}_{i=1, \dots, n}) &\simeq e^{-nD_{KL}(\tilde{\lambda}_{c', p'} || \tilde{\lambda}_{c, p})} \simeq H(c, p)^n, \\ H(c, p) &= e^{-D_{KL}(\tilde{\lambda}_{c', p'} || \tilde{\lambda}_{c, p})}. \end{aligned} \quad (28)$$

From this, we can see how the number of aftershocks will influence the width of the posterior, at least in the asymptotic situation of large  $n$ . For example, Figure 3b shows the contour lines that corresponds roughly to the width for 1, 10, 100 and 1000 aftershocks.

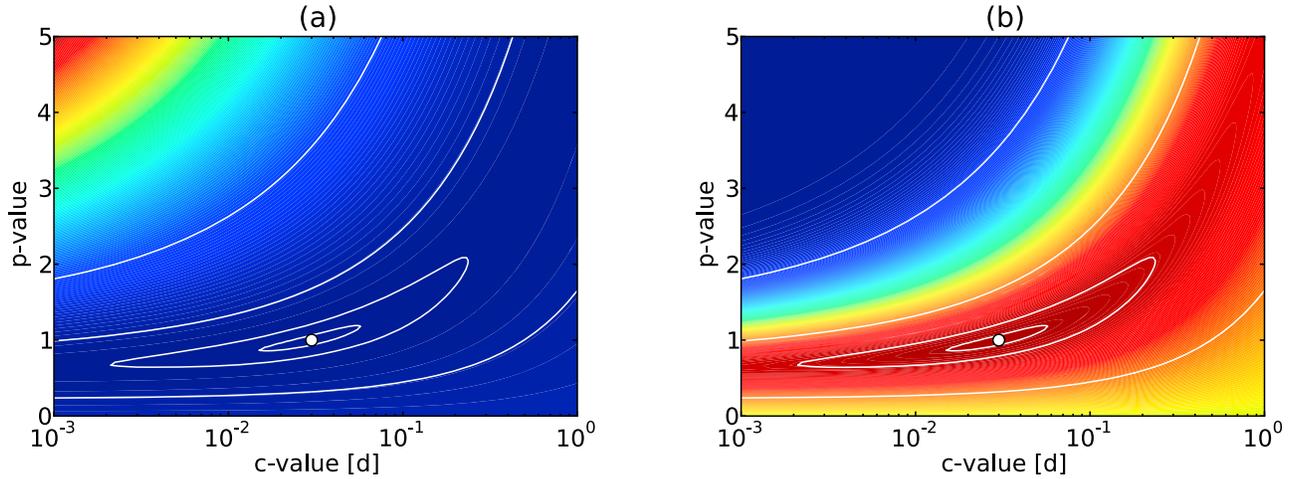
[33] Note that the curved shaped valley corresponds to the straight lines  $\alpha = p/c$  along which the distributions  $\tilde{\lambda}_{c, p}$  tend to an exponential distribution so that the KL divergence tends to some finite value.

### 2.5.1. The Fischer Information

[34] In the limit  $n \rightarrow \infty$  the behavior of the posterior can be fully analyzed. Asymptotically, the posterior distribution becomes Gaussian with expectation  $\{c', p'\}$  and a covariance matrix which is given by the inverse Hessian of the  $D_{KL}$  at that point. More precisely we consider the following rescaled, dimensionless quantities

$$\tau = \sqrt{n} \frac{c - c'}{c'}, \quad \rho = \sqrt{n}(p - p'), \quad \tau_{\text{start}} = \frac{t_{\text{start}}}{c'}, \quad \tau_{\text{stop}} = \frac{t_{\text{stop}}}{c'}. \quad (29)$$

This amounts to measure all times in units of  $c'$ , the characteristic time scale of the data generating a MOL behavior. We



**Figure 3.** The Kullback-Leibler divergence for the MOL with  $p = 1$  and  $c = 0.03$  d (white spot): (a)  $D_{KL}$ , (b)  $e^{-D_{KL}}$ . The contour lines on Figure 3a correspond to the levels  $-\log(0.1)/n$  for  $n = 1, 10, 100, 1000$ . Thus, they indicate where the posterior has its 10% of maximum contour lines for  $n$  samples. On Figure 3b contour lines are for  $0.1^{1/n}$ .

can express the KL distance in terms of  $c'$ ,  $p'$ ,  $\tau$ ,  $\rho$  and  $n$  as follows:

$$D_{KL}^{t_{\text{start}}-t_{\text{stop}}}(c', p' | c, p) = -\log(c') - \log(\tilde{D}(1 + \tau/\sqrt{n}, p' + \rho/\sqrt{n})) + (p' + \rho/\sqrt{n})\tilde{D}(1, p') \cdot \int_{\tau_{\text{start}}}^{\tau_{\text{stop}}} \frac{\log(1 + u/(1 + \tau/\sqrt{n}))}{(1 + u)^{p'}} du \quad (30)$$

$$= -\log(c') + D_{KL}^{\tau_{\text{start}}, \tau_{\text{stop}}}(1, p' | 1 + \tau/\sqrt{n}, p' + \rho/\sqrt{n}), \quad (31)$$

where  $D_{KL}^{t_{\text{start}}, t_{\text{stop}}}(c', p' | c, p)$  is the KL distance as in equation (27) and  $\tilde{D}$  is  $D$  as in equations (3) and (4) with  $t_{\text{start}} = \tau_{\text{start}}$  and  $t_{\text{stop}} = \tau_{\text{stop}}$ . Then, developing the KL around its minimum at  $\{c', p'\}$ , which is equivalent to expanding the above expression around  $\tau \sim 0$  and  $\rho \sim 0$  to second order, yields

$$\mathbb{P}(\tau, \rho | \{t_i\}_{i=1, \dots, n}) \sim G(0, \Sigma), \quad (32)$$

where  $\Sigma$  is minus the inverse of the Hessian of the function  $\Phi : (c, p) \mapsto D_{KL}(\tilde{\lambda}_{c'=1, p'} | \tilde{\lambda}_{c, p})$  evaluated at  $c = c' = 1$  and  $p = p'$  and  $G(0, \Sigma)$  is the pdf of a two dimensional Gaussian with mean 0 and covariance matrix  $\Sigma$ :

$$\Sigma = -\mathcal{H}^{-1}, \quad \mathcal{H} = \begin{bmatrix} \partial_{\tau\tau}^2 \Phi & \partial_{\tau\rho}^2 \Phi \\ \partial_{\rho\tau}^2 \Phi & \partial_{\rho\rho}^2 \Phi \end{bmatrix}. \quad (33)$$

Explicit computation shows that  $\Sigma$  has the following structure:

$$\Sigma = \begin{bmatrix} \Sigma_{\tau\tau} & \Sigma_{\tau\rho} \\ \Sigma_{\rho\tau} & \Sigma_{\rho\rho} \end{bmatrix}, \quad (34)$$

where  $\Sigma_{\tau\rho} = \Sigma_{\rho\tau}$  and the functions depend only on the rescaled quantities  $\tau_{\text{start}}$ ,  $\tau_{\text{stop}}$  and  $p$ . Thus,  $\Sigma_{\tau\tau}$  measures the posterior variance of  $\tau$  which is the relative posterior variance of  $c$ . From this it follows that the marginal posterior

distribution of  $c$  is a Gaussian with mean and variance given by

$$\mathbb{E}(c | \{t_i\}_{i=1, \dots, n}) \simeq c', \quad (35)$$

$$\mathbb{V}(c | \{t_i\}_{i=1, \dots, n}) \simeq \frac{c'^2}{n} \Sigma_{\tau\tau}(t_{\text{start}}/c', t_{\text{stop}}/c', p). \quad (36)$$

The posterior of  $p$  instead looks

$$\mathbb{E}(p | \{t_i\}_{i=1, \dots, n}) \simeq p', \quad (37)$$

$$\mathbb{V}(p | \{t_i\}_{i=1, \dots, n}) \simeq \frac{1}{n} \Sigma_{\rho\rho}(t_{\text{start}}/c', t_{\text{stop}}/c', p). \quad (38)$$

If instead of fixing  $n$ , we rather fix the productivity  $\tilde{K}$ , the posterior uncertainties scale as follows in the limit of large  $\tilde{K}$

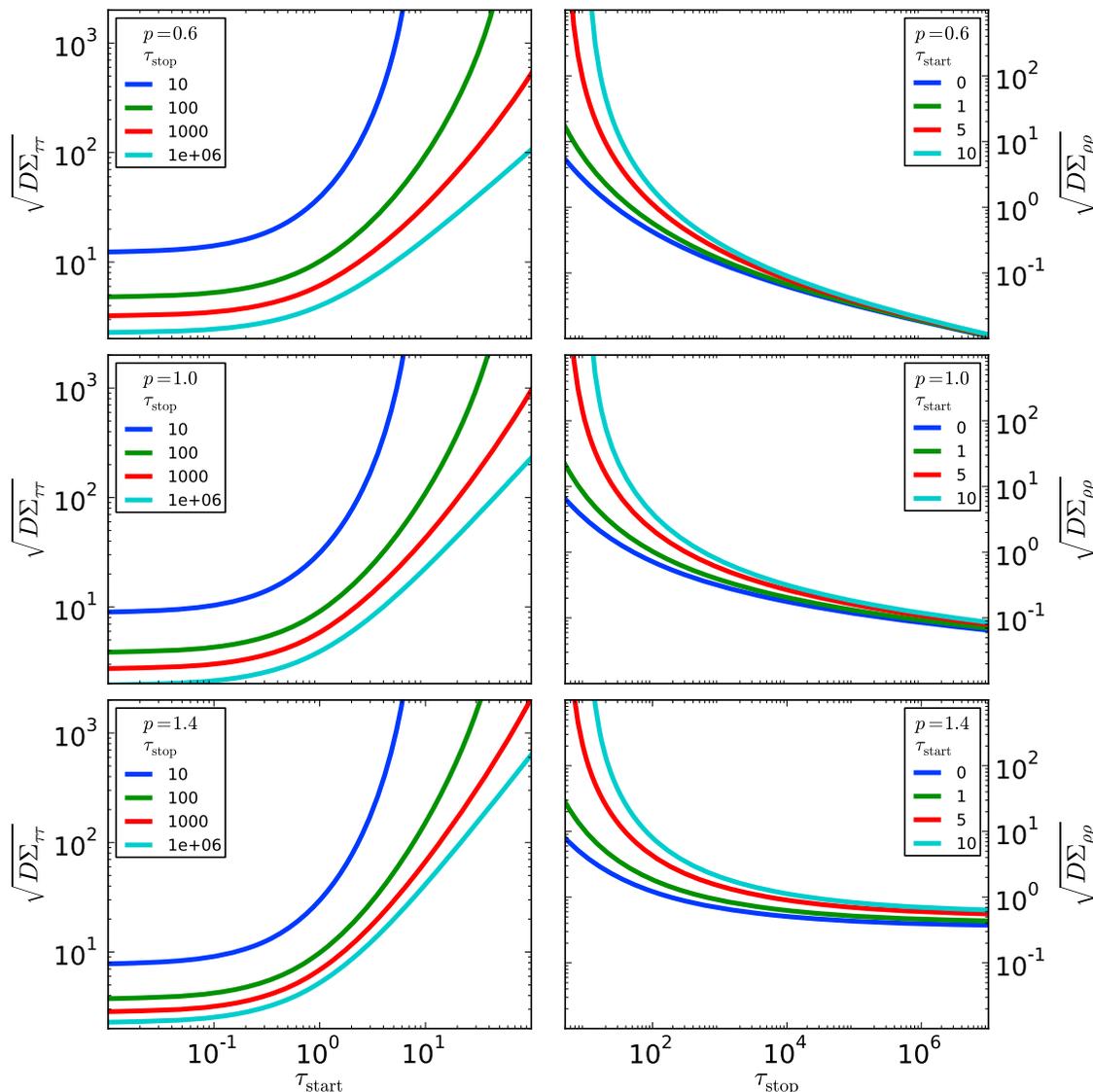
$$\mathbb{V}(c | \{t_i\}_{i=1, \dots, n}) \simeq \frac{c'^2}{\tilde{K}} D(c', p') \Sigma_{\tau\tau}(t_{\text{start}}/c', t_{\text{stop}}/c', p), \quad (39)$$

$$\mathbb{V}(p | \{t_i\}_{i=1, \dots, n}) \simeq \frac{1}{\tilde{K}} D(c', p') \Sigma_{\rho\rho}(t_{\text{start}}/c', t_{\text{stop}}/c', p). \quad (40)$$

Figure 4 shows these posterior uncertainties for  $p \in \{0.6, 1.0, 1.4\}$  using

$$n \sim \frac{\tilde{K}}{D(c', p')}. \quad (41)$$

Not surprisingly,  $t_{\text{stop}}$  has essentially no influence on the posterior variance of  $c$  as soon as  $t_{\text{stop}} > 100 c$ . Vice versa, as soon as  $t_{\text{start}} < c$ , the posterior uncertainty of  $p$  depends only on  $t_{\text{stop}}$ . On the other hand, if  $t_{\text{stop}}$  becomes small, no estimate of  $p$  is possible. Note also, that even for  $t_{\text{start}} = 10 c$ , reasonable estimates of  $c$  are possible, in



**Figure 4.** Standard deviations of the posterior uncertainties of the rescaled (left)  $c$ -value and (right)  $p$ -value estimates as a function of the rescaled observation limits  $\tau_{\text{start}} = t_{\text{start}}/c'$  and  $\tau_{\text{stop}} = t_{\text{stop}}/c'$  (see equations (39) and (40)). There is only a very weak dependency on the actual  $p$ -value, which ranges from (top) 0.6, (middle) 1.0, and (bottom) 1.4. The influence of  $t_{\text{stop}}$  on the relative resolution of  $c$  is negligible and saturates quickly for  $t_{\text{stop}} > 100c$ . Note that, even for  $t_{\text{start}} \simeq c$ , the resolution of  $c$  is about the same as for  $t_{\text{start}} = 0$ . On the other hand, the influence of  $t_{\text{start}}$  on the resolution of the  $p$ -value is negligible as soon as  $t_{\text{stop}} > 10^3c$ . For  $c$  and  $p$ , there is a limit in resolution, which may only be changed through the number of observations. Under optimal conditions (unlimited observation interval) the limiting relative resolution for  $c$  is about 10 times smaller than the one for  $p$ . However, the resolution will increase with an increasing number  $n$  of events in longer observation time interval, for  $p \leq 1$  actually without bounds.

particular in data-rich situations with more than  $10^3$  events. Indeed, in this situation, we expect a determination of  $c$  with an accuracy of 30%. As a rule of thumb, for  $p \simeq 1$ ,  $n > 100$  (so that we are in the asymptotic regime),  $t_{\text{start}} \leq c$  and  $t_{\text{stop}} \geq 10^4c$ , we find a resolution

$$\frac{\Delta c}{c} \simeq \frac{10}{\sqrt{n}}, \quad \Delta p \simeq \frac{1}{\sqrt{n}}. \quad (42)$$

### 3. Bayesian Analysis of Aftershock Decay Rates

[35] In order to illustrate our Bayesian approach with real data, we analyze seismicity rate immediately after 82 main shocks with magnitude ranging from 3 to 5 using the catalogue of the Japan Meteorological Agency (JMA) and the handpicked catalogue produced by *Peng et al.* [2007]. This catalogue was obtained by high-pass-filtering and manual detection in the envelope of clear double peaks corresponding to  $P$  and  $S$  arrivals. Thus, by recovering the smaller events hidden in overlapping seismograms and coda,

the handpicked catalogue documents 5 times as many aftershocks in the first 200 s as in the JMA catalogues. Here, using the combined catalogue produced by *Peng et al.* [2007], we can compare and concatenate two independent data sets covering two different time periods of the same aftershock sequences (Figure 5a):

[36] 1.  $t_{\text{start}} = 20$  s and  $t_{\text{stop}} = 900$  s for short times,

[37] 2.  $t_{\text{start}} = 900$  s and  $t_{\text{stop}} = 10^7$  s for long times.

To select aftershocks with respect to their magnitude, we convert the magnitude of the handpicked catalogue to the JMA magnitude using the empirical relationships proposed by *Peng et al.* [2007]. Following this work, we also set the minimum magnitude threshold for aftershocks to  $M_{\text{min}}^A = 1.5$  to eliminate all potential artifacts related to catalogue incompleteness over the time interval  $[t_{\text{start}}, t_{\text{stop}}]$ . Finally, we stack all aftershocks according to the main shock time to compensate for the small number of events in each sequence. Thus, we assume that the variations of  $c$  and  $p$ -values from one sequence to another are small.

[38] For the catalogues over short ( $n = 64$  aftershocks) and long times ( $n = 474$  aftershocks), Figure 5 shows the posterior density of the  $\{c, p\}$ -values using our Bayesian analysis of the MOL. Not surprisingly given their time intervals  $[t_{\text{start}}, t_{\text{stop}}]$ , the catalogue over long times (Figure 5c) is more appropriate than the catalogue over short times (Figure 5b) to estimate the  $p$ -value, and this is the inverse regarding the estimation of the  $c$ -value. Nevertheless, it is impossible to distinguish between both catalogues taking into account the albeit rather large uncertainty on these parameter estimates (Figure 5e). Only the concatenated catalogue allows to reduce uncertainty and to isolate a more localized area in the parameter space  $\{c, p\}$  of the MOL (Figure 5d). This zone overlaps with the areas predicted by individual catalogue and we cannot reject the hypothesis that the same MOL works for all data sets. Hence, despite variable uncertainty, all these catalogues can be used alone to correctly, if not necessarily accurately, evaluate  $\{c, p\}$ -values.

[39] For the concatenated catalogue, Figure 6a shows the marginal posterior distributions of  $p$  and  $c$  in linear scales. Using these normal-like distributions, the median and the highest posterior density intervals can be computed to determine the most probable range of  $\{c, p\}$ -values. For a Bayesian credibility interval of 95%, Figure 6b shows that we have  $0.89 < p < 0.94$  and  $8 \text{ s} < c < 116 \text{ s}$ . These results support the hypothesis that there is a characteristic time delay before the onset of a power-law aftershock decay rate (i.e.,  $c \neq 0$ ). Despite larger uncertainties given a smaller number of events, the same conclusion may be reached using larger  $M_{\text{A}}^{\text{min}}$  and  $t_{\text{start}}$ -values. In the future, identifying more aftershocks at earlier times, especially within 20 s of the main shock occurrence time, could help to better constrain the  $c$  value.

[40] To assess the quality of fit, Figure 7 shows the observed aftershock sequences and the best-fit provided by the concatenated catalogue and the two independent catalogues over short and long times. These direct comparisons between the data and the models reveal also the influence of the time interval  $[t_{\text{start}}, t_{\text{stop}}]$  on the evaluation of the  $\{c, p\}$ -values. As shown theoretically in section 2 and numerically in Figure 5, the catalogue over long times is less efficient to predict the early aftershock decay rate by extrapolation (cyan curve in the inset of Figure 7a) because,

as  $t_{\text{start}} > 10 c$ , there is a large uncertainty on the estimation of the  $c$ -value. Inversely, the catalogue over short times is less efficient to predict the long term aftershock decay rate (orange curve in the inset of Figure 7a) because, as  $t_{\text{stop}} < 10 c$ , it is impossible to get an accurate estimate of the  $p$ -value.

[41] Most importantly, Figures 5, 6 and 7 show on real aftershock sequence how the  $\{c, p\}$ -values are related to one another. In fact, as we move from individual catalogues to the concatenated one, we see how the gain of knowledge acquired on one parameter benefits to the other.

#### 4. Concluding Remarks

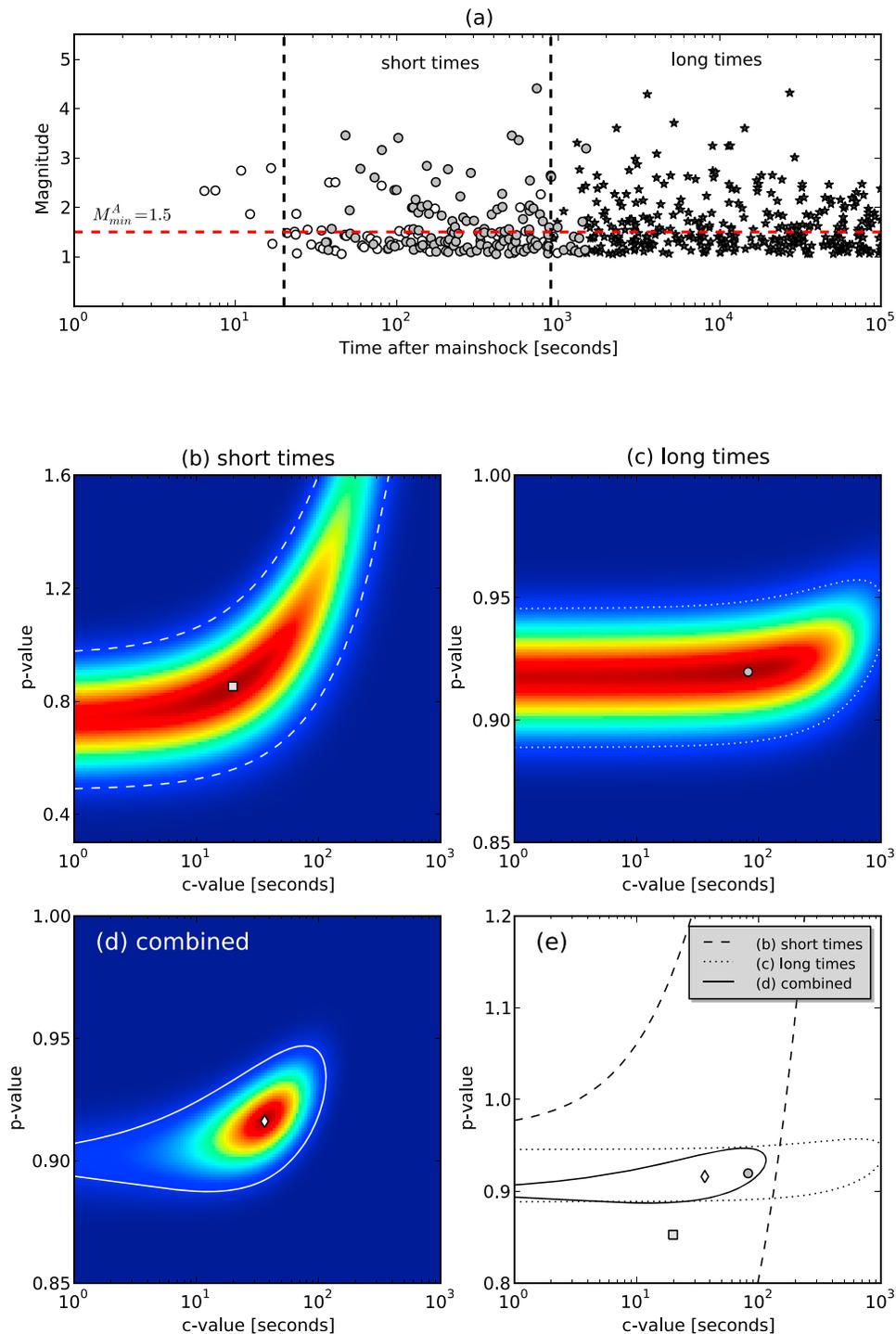
[42] The main objective of this paper is to introduce the Bayesian estimation technique into the study of the aftershock decay rate. In practice, we concentrate on the power-law exponent and the characteristic time delay before the onset of the power-law aftershock decay rate. We could have chosen any aftershock production law that incorporates this two ingredients over short and long times, respectively [e.g., *Dieterich*, 1994; *Narteau et al.*, 2002]. For simplicity but without loss of generality, we have chosen here the MOL.

[43] The main advantage of our Bayesian framework is to provide (1) an estimation of the underlying parameters of the posterior distribution together with (2) a quantification of the uncertainty of our knowledge based only on the amount of information contained in the data. This uncertainty represents an additional source of information that can be exploited together with the variance of the maximum likelihood estimator, which is commonly used to determine the error bars on the estimates of the MOL [*Ogata*, 1983]. The latter one quantifies only the variability of the estimation process, whereas our methods gives credibility regions in the space of parameters. This is crucial, in particular for aftershock sequences with a small number of events where both methods widely differ.

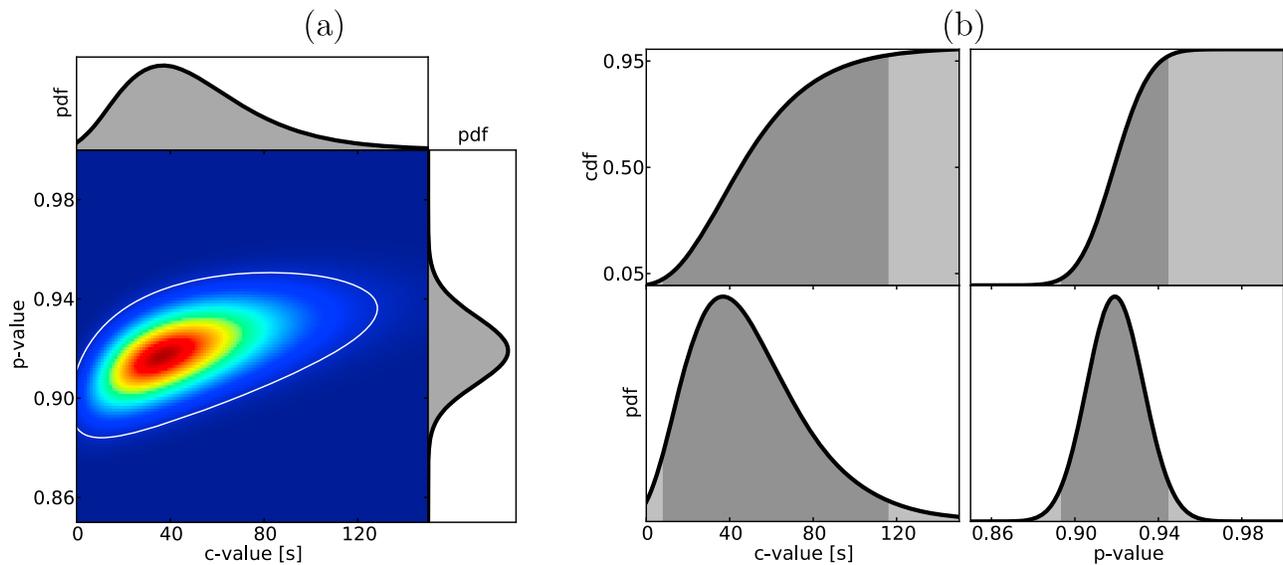
[44] Another advantage of our statistical framework is to dissociate the productivity parameter  $K$  from the shape parameters  $\{c, p\}$ . Thus, we can determine in much greater detail and with more accuracy the shape of the transition toward the power-law regime over short times.

[45] We emphasize the critical role of  $t_{\text{start}}$  and  $t_{\text{stop}}$  and precisely show how these two parameters determine the uncertainty in calculated results. We conclude that these two time limits should be kept constant as much as it is possible for the comparison of different aftershock sequences. Nevertheless, dealing with the variability of aftershock data completeness over short (overlapping seismograms) and long (seismic noise) times, this is not always possible. Our analytical and numerical solutions can then be used to evaluate the parameters of the MOL taking into account the overall resolution of these estimates with respect to the time interval  $[t_{\text{start}}, t_{\text{stop}}]$  of data completeness.

[46] Similarly, our Bayesian approach can also be implemented with a reduced number of events in the limit of  $n \rightarrow 0$ . This is particularly important for the analysis of the statistical properties of the largest aftershocks, which are rare despite their higher level of detectability. In the limit of  $n \rightarrow 0$ , it is clear that the uncertainty on the parameter values of individual sequence may be high. Nevertheless, it



**Figure 5.** The posterior density of  $\{c, p\}$  using two independent time periods of the catalogue produced by Peng *et al.* [2007, online supplementary material]. (a) Magnitude versus logarithm of times for  $M > 1$  events. The open circles mark events picked by hand. The stars denote events listed in the JMA catalogue. The gray circles mark those that are both picked by hands and listed in the JMA catalogue. A random number between  $-0.05$  and  $0.05$  is added to the magnitudes for plotting purposes. The horizontal dashed line indicate the minimum magnitude threshold for aftershocks. The vertical dashed lines indicate the  $t_{\text{start}}$ -values of the two independent time periods under investigation. The posterior density of  $\{c, p\}$  for the catalogues over (b) short times ( $20\text{s} < t < 900\text{s}$ ), (c) long times ( $900\text{s} < t < 10^7\text{s}$ ) and for the (d) concatenated catalogue. In each figure the marker and the line represent the most probable  $\{c, p\}$ -values and the 95% isoline, respectively. (e) All markers and lines are reported for comparison. Note the logarithmic scale for  $c$ .

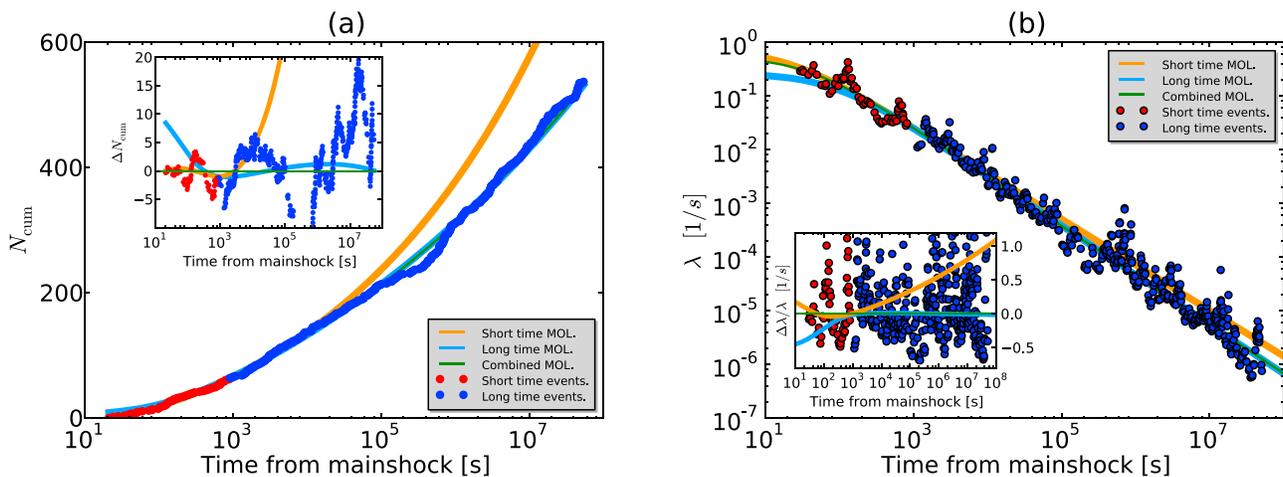


**Figure 6.** Marginal posterior distributions of  $c$  and  $p$ . (a) The posterior density of  $\{c, p\}$  in linear scales and the probability distribution function of the marginal posterior distributions of  $p$  (left) and  $c$  (top). (b) The probability (bottom) and cumulative (top) distribution functions of the marginal posterior distributions of  $c$  (left) and  $p$  (right). The dark gray areas correspond to the 95% Bayesian credibility regions delimited by the 0.025 and 0.975 quantiles, respectively. Thus, the Bayesian posterior credibility for  $8s < c < 116s$ , regardless of the value of  $p$  is 95%. In the same way, the credibility for  $0.89 < p < 0.94$  regardless of  $c$  is again 95%. When considering information about  $p$  and  $c$  simultaneously, it is the white contour line in Figure 6a that delimits 95% credibility in the  $\{c, p\}$  space.

could be compensated by the large number of intermediate magnitude main shocks.

[47] We have investigated the codependence of the parameter  $c$  and  $p$  of the MOL, noting that, for each of them, the precision of their estimates is subordinated to the amount

of information that constrains the other. A fundamental consequence is that the estimation of one parameter should not be isolated from the other. In other words, it is recommended to simultaneously estimate the power-law regime at long times (i.e., the  $p$ -value) and the time delay before the



**Figure 7.** The stacked aftershock sequence after 82 main shocks with magnitude ranging from 3 to 5 in Japan using two independent time periods of the catalogue produced by Peng *et al.* [2007, online supplementary material]. Red and blue dots are for  $M > 1.5$  events in  $20s < t < 900s$  and  $t > 900s$ , respectively. (a) The cumulative number of events with respect to the time from the main shocks. Lines are the best-fit of the concatenated catalogue (green) and of the two independent catalogues over short (orange) and long times (cyan). For the data and the two independent catalogues, the inset shows the difference with the best-fit obtained from the concatenated catalogue. (b) The aftershock rate with respect to the time from the main shocks. Lines are the best-fit of the concatenated catalogue (green) and of the two independent catalogues over short (orange) and long times (cyan). For the data and the two independent catalogues, the inset shows the relative difference with the best-fit obtained from the concatenated catalogue.

onset of this stationary relaxation regime (i.e., the  $c$ -value). In this case, our Bayesian approach is an efficient and effective method to instantaneously capture such a posterior dependences of the model parameters.

[48] From our analytical solutions and our tests on real aftershock sequences, we have shown that, despite variable uncertainties, we can analyze different time periods of an aftershock sequence to efficiently estimate the parameters of the MOL. Most importantly, this indicates that instrumental catalogues may be used to correctly evaluate  $\{c, p\}$ -values. The only condition is to select specific time periods and magnitude ranges that ensure aftershock data completeness and a reasonable resolution on the MOL parameters. In this case, our Bayesian approach may be applied to declustered instrumental catalogues to explore systematic variations of  $\{K, c, p\}$ -values along active fault zones. We infer that it may significantly contribute to the statistical characterization of aftershock sequences.

[49] **Acknowledgments.** The paper was improved by the constructive comments and thoughtful suggestions of I. Main and R. Shcherbakov.

## References

- Atkinson, B. (1991), *Fracture Mechanics of Rocks*, Elsevier, New York.
- Ben-Zion, Y., and V. Lyakhovskiy (2006), Analysis of aftershocks in a lithospheric model with seismogenic zone governed by damage rheology, *Geophys. J. Int.*, *165*, 197–210, doi:10.1111/j.1365-246x.2006.02878.x.
- Berger, J. O., B. Boukai, and Y. Wang (1997), Unified frequentist and Bayesian testing of a precise hypothesis, *Stat. Sci.*, *12*, 133–160.
- Bishop, C. (2006), *Pattern Recognition and Machine Learning*, Springer, New York.
- Cox, R. T. (1946), Probability, frequency and reasonable expectation, *Am. J. Phys.*, *14*, 1–13, doi:10.1119/1.1990764.
- Dieterich, J. (1994), A constitutive law for rate of earthquake production and its application to earthquake clustering, *J. Geophys. Res.*, *99*, 2601–2618.
- 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.
- Gasparini, P., and B. Lohi (2009), An empirical comparison among aftershock decay models, *Phys. Earth Planet. Int.*, *175*, 183–193, doi:10.1016/j.pepi.2009.03.011.
- 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.
- Hauksson, E., and L. Jones (1988), The July 1986 Oceanside  $M_L = 5.3$  earthquake sequence in the continental borderland, Southern California, *Bull. Seismol. Soc. Am.*, *78*, 1885–1906.
- Helmstetter, A., and B. E. Shaw (2006), Relation between stress heterogeneity and aftershock rate in the rate-and-state model, *J. Geophys. Res.*, *111*, B07304, doi:10.1029/2005JB004077.
- Holschneider, M., G. Zöller, and S. Hainzl (2011), Estimation of the maximum possible magnitude in the framework of the doubly-truncated Gutenberg-Richter model, *Bull. Seismol. Soc. Am.*, *112*, 1649–1659, doi:10.1785/0120100289.
- Jaynes, E. (2005), *Probability Theory: The Logic of Science*, Cambridge Univ. Press, Cambridge, U. K.
- Kagan, Y. (2004), Short-term properties of earthquake catalogs and models of earthquake source, *Bull. Seismol. Soc. Am.*, *94*, 1207–1228.
- Kisslinger, C. (1993), The stretched exponential function as an alternative model for aftershock decay rate, *J. Geophys. Res.*, *98*, 2271–2281.
- Kisslinger, C., and L. Jones (1991), Properties of aftershocks in southern California, *J. Geophys. Res.*, *96*, 11,947–11,958.
- Kullback, S., and R. Leibler (1951), On information and sufficiency, *Ann. Math. Stat.*, *22*, 79–86.
- Le Cam, L. (1986), *Asymptotic Methods in Statistical Decision Theory*, Springer, New York.
- Leonard, T., and J. S. Hsu (1999), *Bayesian Methods*, Cambridge Univ. Press, Cambridge, U. K.
- Nanjo, K., H. Nagahama, and M. Satomura (1998), Rates of aftershock decay and the fractal structure of active fault systems, *Tectonophysics*, *287*, 173–186.
- Narteau, C., P. Shebalin, and M. Holschneider (2002), Temporal limits of the power law aftershock decay rate, *J. Geophys. Res.*, *107*(B12), 2359, doi:10.1029/2002JB001868.
- Narteau, C., P. Shebalin, and M. Holschneider (2005), Onset of power law aftershock decay rates in southern California, *Geophys. Res. Lett.*, *32*, L22312, doi:10.1029/2005GL023951.
- Narteau, C., P. Shebalin, and M. Holschneider (2008), Loading rates in California inferred from aftershocks, *Nonlinear Processes Geophys.*, *15*, 245–263.
- 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.
- Ogata, Y. (1983), Estimation of the parameters in the modified Omori formula for aftershocks frequencies by a maximum likelihood procedure, *J. Phys. Earth*, *31*, 115–124.
- Ogata, Y. (1988), Statistical models for earthquake occurrence and residual analysis for point process, *J. Am. Stat. Assoc.*, *83*, 9–27.
- Ojala, I. O., I. G. Main, and B. T. Ngwenya (2004), Strain rate and temperature dependence of Omori law scaling constants of AE data: Implications for earthquake foreshock-aftershock sequences, *Geophys. Res. Lett.*, *31*, L24617, doi:10.1029/2004GL020781.
- Omori, F. (1894), On after-shocks of earthquakes, *J. Coll. Sci. Imp. Univ. Tokyo*, *7*, 111–200.
- Otsuka, M. (1985), Physical interpretation of Omori's formula, *Sci. Rep. Shimabara Volcano Observ.*, *13*, 11–20.
- Ouillon, G., and D. Sornette (2005), Magnitude-dependent Omori law: Theory and empirical study, *J. Geophys. Res.*, *110*, B04306, doi:10.1029/2004JB003311.
- Peng, Z. G., J. E. Vidale, and H. Houston (2006), Anomalous early aftershock decay rate of the 2004  $M_w$  6.0 Parkfield, California, earthquake, *Geophys. Res. Lett.*, *33*, L17307, doi:10.1029/2006GL026744.
- Peng, Z. G., J. E. Vidale, M. Ishii, and A. Helmstetter (2007), Seismicity rate immediately before and after main shock rupture from high frequency waveforms in Japan, *J. Geophys. Res.*, *112*, B03306, doi:10.1029/2006JB004386.
- Scharf, L. (1991), *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Addison-Wesley, Reading, Mass.
- Scholz, C. (1968), Microfractures, aftershocks, and seismicity, *Bull. Seismol. Soc. Am.*, *58*, 1117–1130.
- Shcherbakov, R., and D. L. Turcotte (2004), A damage mechanics model for aftershocks, *Pure Appl. Geophys.*, *161*, 2379–2391.
- Shcherbakov, R., D. L. Turcotte, and J. B. Rundle (2004), A generalized Omori's law for earthquake aftershock decay, *Geophys. Res. Lett.*, *31*, L11613, doi:10.1029/2004GL019808.
- Shebalin, P., C. Narteau, M. Holschneider, and D. Schorlemmer (2011), Short-term earthquake forecasting using early aftershock statistics, *Bull. Seismol. Soc. Am.*, *112*, 297–312, doi:10.1785/0120100119.
- Stein, S., and M. Liu (2009), Long aftershock sequences within continents and implications for earthquake hazard assessment, *Nature*, *462*, 87–89, doi:10.1038/nature08502.
- Utsu, T. (1961), A statistical study on the occurrence of aftershocks, *Geophys. Mag.*, *30*, 521–605.
- Utsu, T., Y. Ogata, and R. Matsu'ura (1995), The centenary of the Omori formula for a decay law of aftershocks activity, *J. Phys. Earth*, *43*, 1–33.
- Vidale, J. E., Z. Peng, and M. Ishii (2004), Anomalous aftershock decay rates in the first hundred seconds revealed from the Hi-net borehole data, *Eos Trans. AGU*, *85*(47), Fall Meet. Suppl., Abstract S23C-07.
- Wiemer, S., and K. Katsumata (1999), Spatial variability of seismicity parameters in aftershock zones, *J. Geophys. Res.*, *104*, 13,135–13,151.
- 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.