Fitting aftershocks decay: method paper

Intro:

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 (magnitude 8) 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 into the so-called modified Omori law (or Omori-Utsu law). The frequency of aftershocks per unit time interval (one hour, one day, one month, etc.) is well represented by the modified Omori formula (Utsu, 1961), or called Omori-Utsu law, which is formulated as:

$$n(t) = \frac{K}{(t+c)^p} \tag{1}$$

where $n(t)$ is the aftershock occurrence (number counts) within a given magnitude range, t the lapse time from the occurrence of the mainshock (the event triggering aftershocks), K the productivity of the aftershock sequence which depends on the magnitude of the mainshock and the lower bound of the magnitude of aftershocks counted, p the power law exponent, and c the time delay before the onset of the power-law aftershock decay rate. There are three unknown parameters here I want to estimate, i.e., K, c, and p. The typical type of Omori-Utsu law is like Figure 1 and Figure 2, where the number of aftershocks decay as a power law after September 28, 2004 Parkfield Earthquake.

According to Utsu (1969), for the case of Japanese aftershock sequences, the constant c was small, 1 day at most, and p varied in the range 0.9-1.9 with values between 1.0 and 1.4 most frequently. The value p was thought to reflect mechanical conditions of the earth's crust. For example, Mogi (1962) demonstrated a certain systematic regional variation of p value in Japan which could be attributed to regional variation of surface heat-flow values. He thought that aftershock activity decays faster, namely the stress relaxes faster, in regions of higher crustal temperature. Later, a similar correlation was also discussed by Kisslinger and Jones (1991).

Dataset:

The dataset of this class project is acquired from Northern California Earthquake Center (Waldhauser and Schaff, 2008; Waldhauser, 2009). I further select a subset of this dataset so that it covers earthquakes record (time, location, magnitude, etc.) from 1984–present within some radius distance from the location of some intermediate-size well-recorded earthquakes, including: 1985 M6.1 Kettleman Hills earthquake, 1989 M6.9 Loma Prieta earthquake, 2004 M6.0 Parkfield earthquake, 2011 M2.7 Prague earthquake, and 2019 M7.1 Ridgecrest earthquake. The earthquake catalogs are truncated such that it contains earthquakes only after the mainshock all the way to the latest records in 2020. For the quality control, we consider a minimum magnitude M1.5 which accounts for the completeness of the dataset. More details are in the method part.

Method:

Since there already exits a parametric form of the model (Equation 1), the rest of the problem is simply how to fit the model to the data and find the best-fit model parameters. Traditionally, estimates of the parameter p had been obtained since Utsu (1961) in the following way. Plot n(t) versus the lapse time t on a log-log scaled plane and then fit an asymptotic straight line; the slope of the line is an estimate for p. The values of c can be determined by a naive empirical technique. Such analysis was based on the time series of counted numbers of aftershocks, while directly based on the occurrence times of aftershocks. On the other hand, Ogata (1983), proposed the following method to estimate the three parameters of the modified Omori law.

Consider occurrence times of aftershock sequence $\{t_1, t_2, \dots, t_N\}$ in a time interval [S, T], in which the origin of the time axis, t = 0, corresponds to the occurrence time of the mainshock. Then assume that the aftershock sequence is distributed according to a nonstationary Poisson process with the intensity function:

$$\lambda(t, \theta) = \frac{K}{(t + c)^p} \tag{2}$$

where

$$\theta = (c, K, p)$$

Now this looks just like the Omori law in Equation (1). Then the log-likelihood function of this intensity function (i.e., the aftershock sequence occurrence) can be written by:

$$logL(c, K, p) = NlogK - p \sum_{i=1}^{N} \log(t_i + c) - K\Lambda(c, p) \tag{3}$$

where N is the total number of aftershocks within the evaluated time window, t_i corresponds to occurrence time of a particular aftershock after the mainshock (in days) and

$$\Lambda(c, p) = \begin{cases} \dfrac{(T + c)^{1-p} - (S + c)^{1-p}}{1 - p} & for\ p \neq 1 \\ \log(T + c) - \log(S + c), & for\ p = 1 \end{cases} \tag{4}$$

Maximizing the log-likelihood function here with respect to the parameters $\theta = (c, K, p)$ will give us the maximum likelihood estimates $\hat{\theta} = (\hat{c}, \hat{K}, \hat{p})$. This maximum likelihood procedure also has the advantage of providing us estimates for the standard errors of the MLE as a byproduct. Specifically, the inverse of the Fisher information matrix $J(\hat{\theta}, S, T)^{-1}$ provides the covariance matrix of the errors of the MLE. So, the uncertainty is evaluated by asymptotic statistics using the Fischer information matrix.

Nevertheless, it is also claimed that (Holschneider et al., 2012) the standard errors of the MLE 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. It only quantifies the variability of the estimation process, instead of giving the credibility regions in the space of parameters. This is crucial, particular for aftershock sequences with a small number of events.

Therefore, in order to get the proper estimations on the uncertainties of the parameters $\theta = (c, K, p)$, I first use the log-likelihood function formulated by Ogata (1983) and apply the MLE method using a regular `python` minimizer `scipy.optimize.minimize`, then use the best-fit solution to run an ensemble of Markov Chain Monte Carlo samplers using an open-source python package `emcee`. `emcee` is an Affine Invariant Markov chain Monte Carlo (MCMC) Ensemble sampler that can allow one to sample the parameters space and provides results of posterior probability distribution, best-fit solution (peak of the posterior probability distribution), as well as the uncertainties represented by the confidence intervals (1 sigma, 2 sigma, etc) on the model parameters.

My first step is to write the formulated Omori law log-likelihood function as a negative form, so that the minimizer can work on that and give me an initial solution which I will feed into the MCMC ensemble samplers. From original likelihood function in Equation (3), the negative log-likelihood is written as:

$$-logL(c, K, p) = -NlogK + p\sum_{i=1}^{N} \log(t_i + c) + K\Lambda(c, p) \tag{5}$$

The occurrence time of all the earthquakes after the corresponding mainshock is t_i in the above formula. I downloaded the earthquake catalogs (contains the information about the occurrence times) for the several aftershock datasets in the US within some radius distance from the mainshock location, including: 1985 M6.1 Kettleman Hills earthquake, 1989 M6.9 Loma Prieta earthquake, 2004 M6.0 Parkfield earthquake, 2011 M2.7 Prague earthquake, and 2019 M7.1 Ridgecrest earthquake (Waldhauser and Schaff, 2008; Waldhauser, 2009). Except the 2011 Prague earthquake was from Oklahoma, other earthquakes are all within California. The earthquake catalogs are truncated such that it contains earthquakes only after the mainshock all the way to the latest records in 2020. For the quality control, people usually need to consider a minimum magnitude which accounts for the completeness of the dataset. This concept is just to account for incomplete detections of the smaller earthquakes (the ability of the seismographs to detect earthquakes decays under a certain magnitude, aka, magnitude of completeness). Generally, the estimated magnitude of completeness in California is ~1.5. So, we only use earthquakes larger than M1.5 for our parameter estimations.

After acquiring the occurrence times, the `scipy.optimize.minimize` minimizer was used to minimize the negative log-likelihood with a given initial guess {c, K, p} = {0.5, 50, 1.1}. Eventually, I got a best-fit solution $\hat{\theta} = (\hat{c}, \hat{K}, \hat{p})$ with the maximum likelihood (the red and blue

curved in Figure 2). Here, the Sequential Least Squares Programming (SLSQP) was selected as the minimizing method, and the bounds for the parameters are assigned as follow: $c \in [10^{-4},\ 2]$, $K \in [2, 10^4]$, $p \in [0.2,\ 2]$.

Then, this best-fit solution from the minimizer is fed into the MCMC ensemble sampler to further explore and sample the posterior probability distribution of the parameters space. The posterior probability function is written as the Bayesian framework as below:

$$P\left(\theta \middle| \vec{t} = \{\mathrm{t}_1, \mathrm{t}_2, \dots\ ,\ \mathrm{t}_N\}\right) = \frac{P\left(\vec{t} = \{\mathrm{t}_1, \mathrm{t}_2, \dots\ ,\ \mathrm{t}_N\} \middle| \theta, I\right) \cdot P(\theta|I)}{P\left(\vec{t} \middle| I\right)} \tag{6}$$

where $P\left(\theta \middle| \vec{t}\right)$ is the posterior probability function which MCMC will sample, $P\left(\vec{t} \middle| \theta, I\right)$ is the likelihood function that was formulated earlier in Equation (3) (without taking the log), $P(\theta|I)$ is the prior information about the parameters, and $P\left(\vec{t} \middle| I\right)$ is the Bayesian evidence term, which is not necessary in this problem. The prior is specified as a uniform prior within the given bounds, i.e., $c \in [10^{-4},\ 2]$, $K \in [2, 10^4]$, $p \in [0.2,\ 2]$. We take the log of the entire Equation (6) to use the log posterior probability for sampling.

There are three parameters to sample. In the `emcee` package, 32 paralleled walkers were used to run for 5000 steps respectively. Different walkers were spread or perturbed around of the best-fit solution acquired from the minimizer, and the perturbation was defined by a 3D Gaussian sphere (corresponds to the three parameters {c, K, p}) with a mean of 10^{-3} for each parameter.

After the ensemble sampling from the 32 walkers, we checked the trace plots (Figure 3) for each parameter, and the burn-in is pretty small and negligible. All 32 walkers successfully converge within 5000 steps for all aftershock datasets. The autocorrelation times are around 30 to 40 steps to walk through the distribution of the parameter space for all aftershock datasets. We still choose to discard the first 100 steps as burn-in and thin by about a half the autocorrelation time (15 steps) and flatten the chain so that we have a flat list of samples. We then make the corner plot for MCMC results and identified the peak (MCMC best-fit solutions) and marked the 1-sigma uncertainties (68% confidence interval) on the histograms, marked the 1-sigma, 2-sigma, and 3-sigma contours on the 2-dimentional posterior probability distribution plots (Figure 4).

Last step, we plot the aftershocks decay with time from the observational dataset and compare our MCMC best-fit solution with observations (Figure 5). Generally, the model fittings all accomplish a good job in capturing the aftershock decay with time. Thus, we conclude that the MLE and MCMC sampling workflow proposed in this class project is appropriate to estimate the model parameters as well as the corresponding uncertainties of the Omori-Utsu law.

Figures:

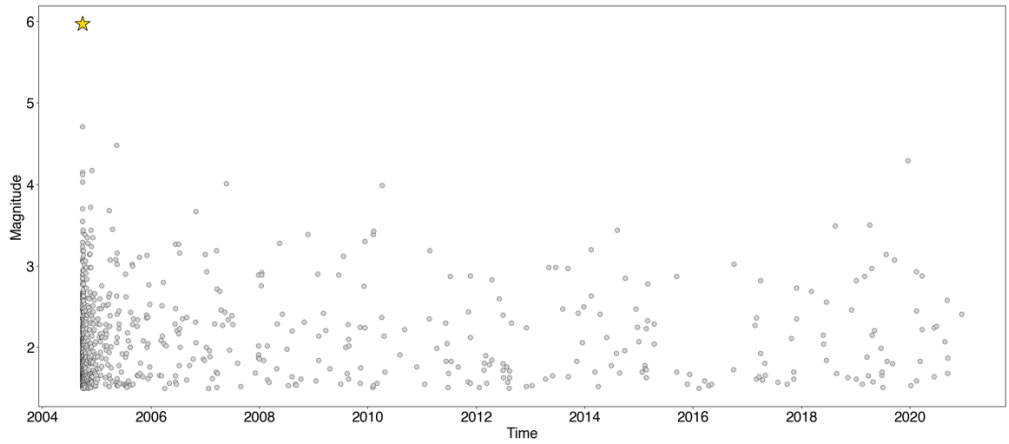

Figure 1. Aftershock datasets after the 2004 M6.0 Parkfield earthquake (occurred on Sep 28, 2004). The yellow star is the M6.0 mainshock. The grey dots are the aftershocks and background events. The vertical axis is the magnitude of earthquakes. Aftershocks concentrated during the first year and gradually decrease with time.

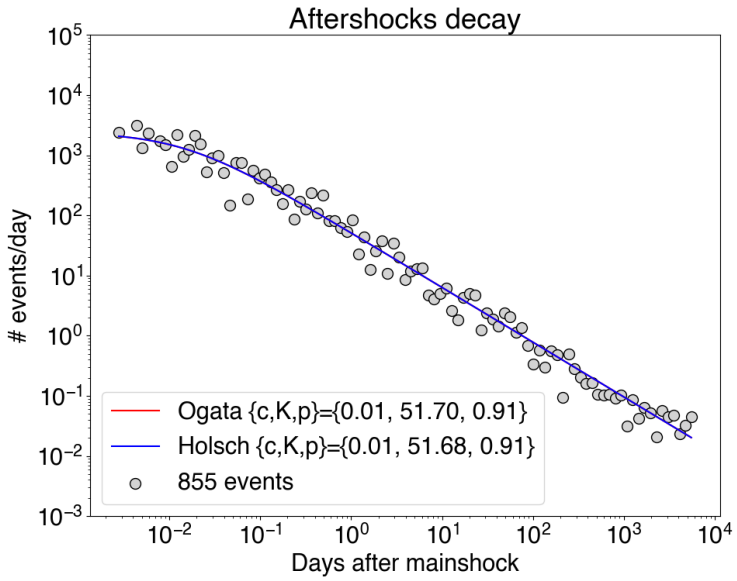

Figure 2. The way to visualize the decay of 2004 Parkfield aftershocks with time. We bin the time, then count the number of earthquakes and calculate the earthquake occurrence rate (# events per day) within each time window. The binning is in log time scale. The x axis is the time lapse from the mainshock (days), the y axis is the earthquake occurrence rate (# events per day). The results are shown in a log-log plot. The grey dots are the 2004 Parkfield aftershocks, which demonstrate a clear Omori law decay. The red and the blue curves are the Omori law models to fit this dataset, parameters are shown in the legend. Two models differ from the formulation of the likelihood function. In this project, the best-fit solution from the Ogata's formulation is further used for MCMC sampling.

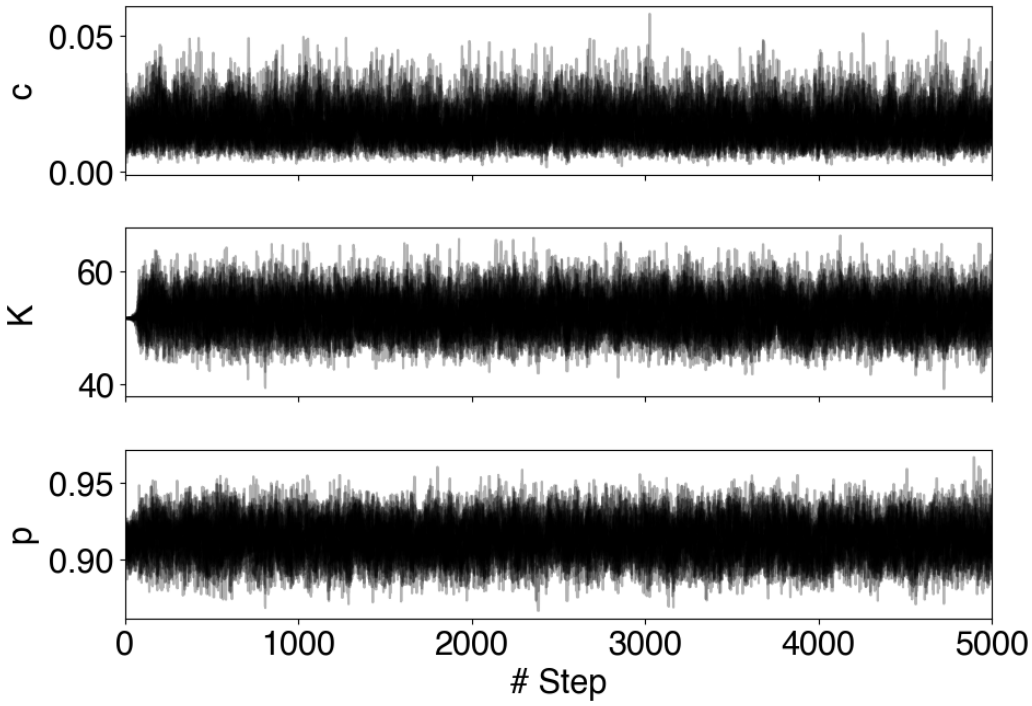

Figure 3. Trace plots from the MCMC ensemble sampling for all three parameters {c, K, p}. Samples are flattened and plotted on x axis from 0 to 5000 steps. The burn-in phase is really short, we only trim the first 100 steps. The final acceptance rates are around 30% for all three parameters. The autocorrelation times are around 30 to 40 steps for all three parameters.

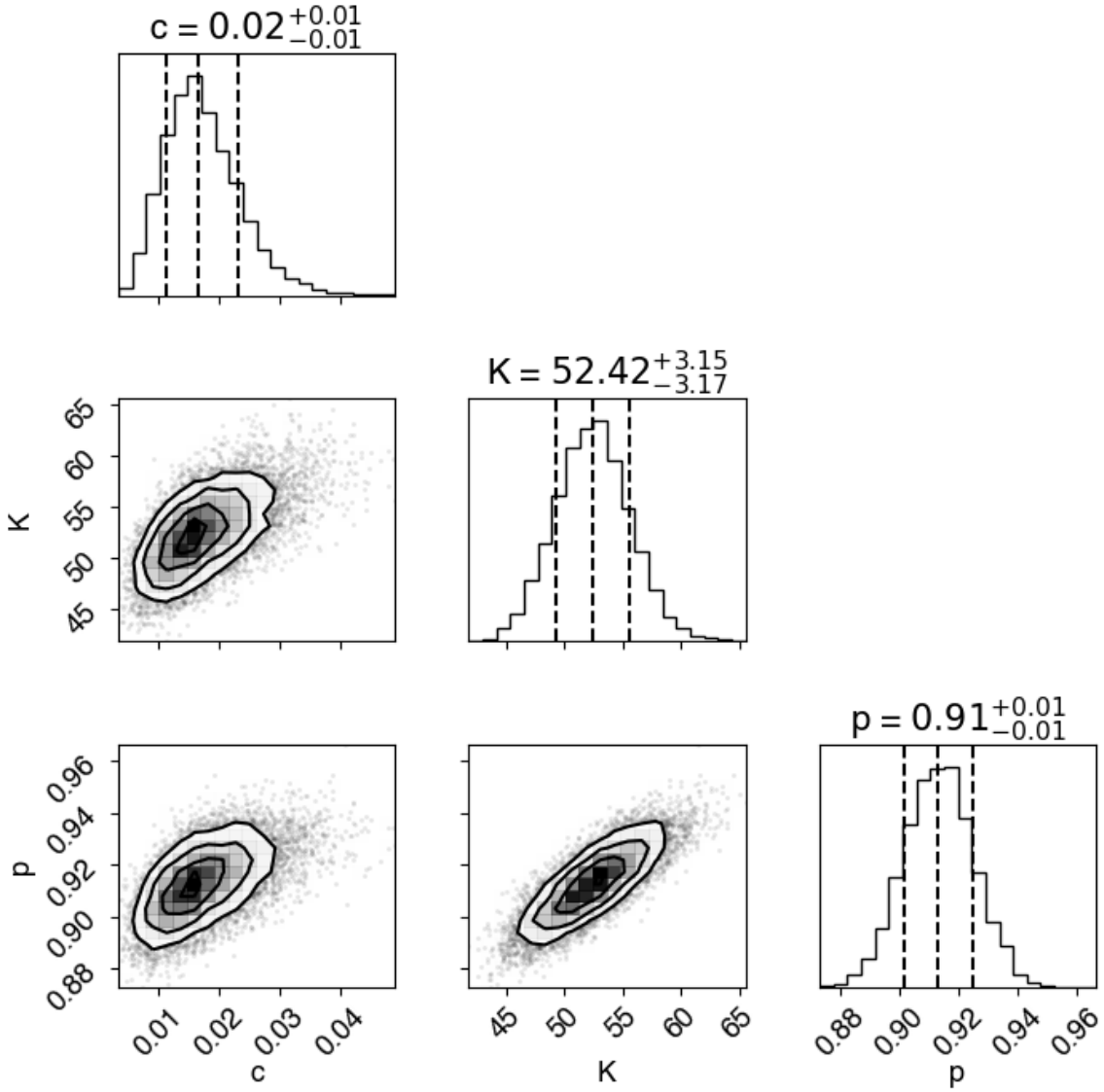

Figure 4. The corner plot for the three parameters {c, K, p} after MCMC ensemble sampling using the package `emcee`. The peak value of c, K, and p are marked by the central dashed lines in the histograms. The ±1 sigma interval (68% confidence interval) for each parameter is also marked by the dashed lines in the histograms. In this 2004 Parkfield dataset, we obtained the best-fit solution {c, K, p} = {0.02, 52.42, 0.91}. The uncertainty for K is much larger than uncertainties for c and p. But in terms of the resolution to consider geophysical problems, these uncertainties are acceptable.

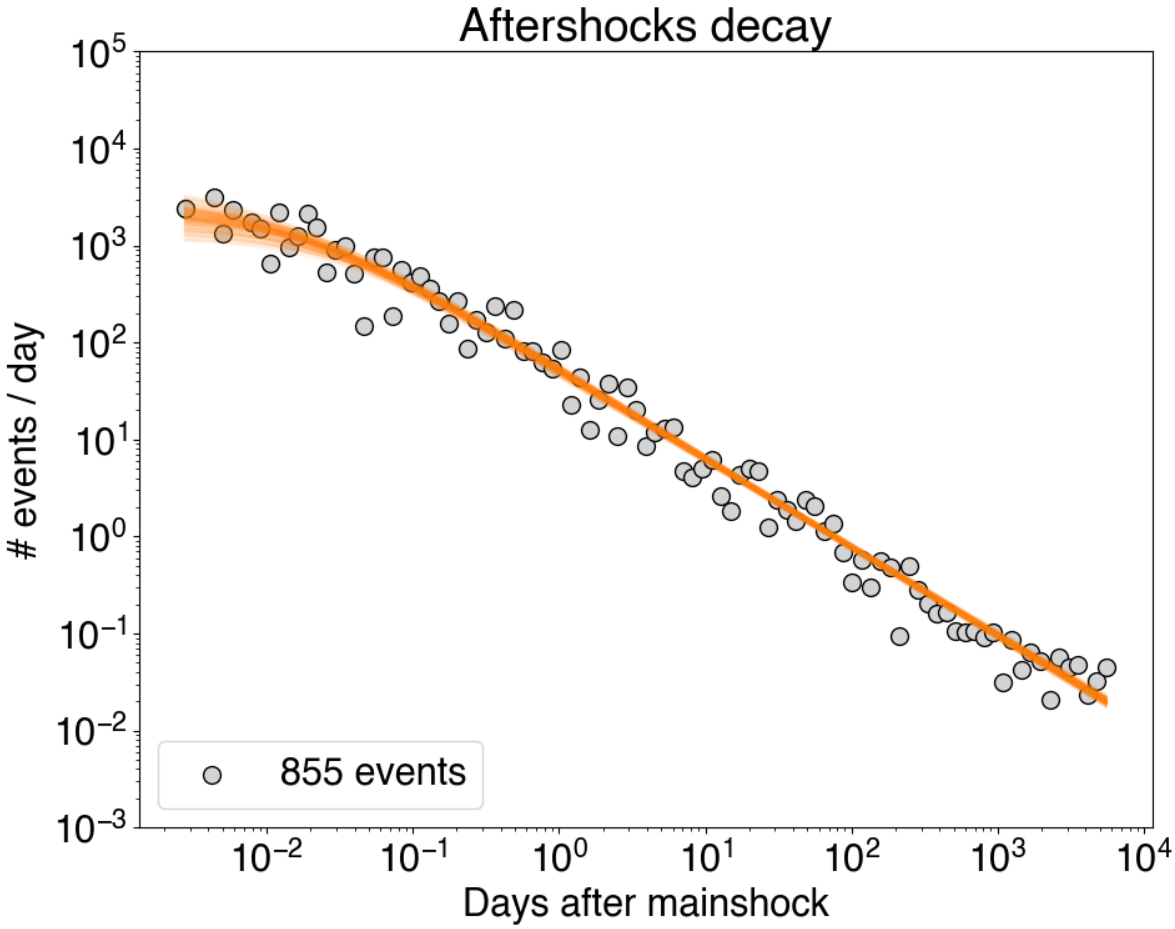

Figure 5. The dataset (grey dots) compared with the MCMC ensemble models (orange curves). Overall, the ensemble models describe very the aftershocks decay very well. The spread of the models is acceptable.

Reference for methods and dataset:

Holschneider, M., Narteau, C., Shebalin, P., Peng, Z., & Schorlemmer, D. (2012). Bayesian analysis of the modified Omori law. Journal of Geophysical Research: Solid Earth, 117(B6).

Kisslinger, C., & Jones, L. M. (1991). Properties of aftershock sequences in southern California. Journal of Geophysical Research: Solid Earth, 96(B7), 11947-11958.

Mogi, K. (1962). On the time distribution of aftershocks accompanying the recent major earthquakes in and near Japan. Bulletin of the Earthquake Research Institute, University of Tokyo, 40(1), 107-124.

Omori, F. (1894). On the after-shocks of earthquakes. J. Coll. Sci., Imp. Univ., Japan, 7, 111-200.

Ogata, Y. (1983). Estimation of the parameters in the modified Omori formula for aftershock frequencies by the maximum likelihood procedure. Journal of Physics of the Earth, 31(2), 115-124.

Utsu, T. (1961). A statistical study on the occurrence of aftershocks. Geophys. Mag., 30, 521-605.

Utsu, T., & Hirota, T. (1969). A note on the statistical nature of energy and strain release in earthquake sequences. Journal of the Faculty of Science, Hokkaido University. Series 7, Geophysics, 3(2), 49-64.

Waldhauser, F. (2009). Near-real-time double-difference event location using long-term seismic archives, with application to Northern California, Bull. Seism. Soc. Am., 99, 2736-2848, doi:10.1785/0120080294

Waldhauser, F. and D.P. Schaff (2008). Large-scale relocation of two decades of Northern California seismicity using cross-correlation and double-difference methods, J. Geophys. Res.,113, B08311, doi:10.1029/2007JB005479