

FULL PAPER

Open Access



Bayesian earthquake forecasting approach based on the epidemic type aftershock sequence model

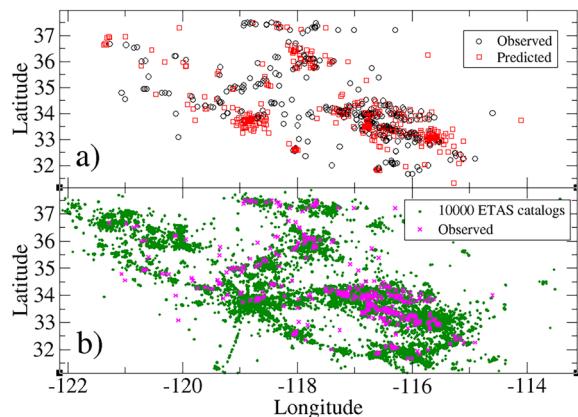
Giuseppe Petrillo^{1*} and Jiancang Zhuang²

Abstract

The epidemic type aftershock sequence (ETAS) model is used as a baseline model both for earthquake clustering and earthquake prediction. In most forecast experiments, the ETAS parameters are estimated based on a short and local catalog, therefore the model parameter optimization carried out by means of a maximum likelihood estimation may be not as robust as expected. We use Bayesian forecast techniques to solve this problem, where non-informative flat prior distributions of the parameters is adopted to perform forecast experiments on 3 mainshocks occurred in Southern California. A Metropolis–Hastings algorithm is employed to sample the model parameters and earthquake events. We also show, through forecast experiments, how the Bayesian inference allows to obtain a probabilistic forecast, differently from one obtained via MLE.

Keywords Statistical seismology, Time series analysis, Probabilistic forecast, Stochastic process

Graphical Abstract



*Correspondence:
Giuseppe Petrillo
g.petrillo@ssmeridionale.it
Full list of author information is available at the end of the article

Introduction

The duration of an earthquake is much smaller than the average temporal distance between two earthquakes. Because of this, it is reasonable to treat an earthquake as a point event in time. This assumption may break down for events with very large magnitudes, but since they are rare, the mathematical approximation of a point process can be adopted for the entire seismic catalog. The epidemic type aftershock sequence (ETAS) model, a kind of point process initially devised to delineate mainshock–aftershock sequences, wherein mainshocks, aftershocks, and foreshocks exhibit analogous behaviors in terms of inducing subsequent events, has gained widespread adoption among researchers as the standard benchmark for assessing hypotheses linked to earthquake clustering (Console et al. 2010; Helmstetter and Sornette 2003; Lombardi and Marzocchi 2010; Ogata 1988, 1998; Petrillo and Lippiello 2023; Petrillo et al. 2023; Spassiani et al. 2024; Zhuang 2011, 2012). In particular, it is possible to obtain the occurrence rate λ of an event with magnitude $m > m_0$, at the position (x, y) and at a time t as:

$$\begin{aligned} \lambda(x, y, t | \mathcal{H}_t) = & \mu(x, y) \\ & + KK_t K_r \sum_{j: t_j < t} 10^{\alpha(m_j - m_0)} v(t - t_j) s(x - x_j, y - y_j), \end{aligned} \quad (1)$$

where the sum considers all previous earthquakes that occurred in a certain region, $\mu(x, y)$ describes the background seismicity and K_t and K_r are constants fixed by imposing the normalization conditions:

$$K_t \int_{t_j}^{\infty} v(t - t_j) dt = 1, \quad (2)$$

and

$$K_r \int_{x_j}^{\infty} \int_{y_j}^{\infty} s(x - x_j, y - y_j) dx dy = 1. \quad (3)$$

The functional form $v(t - t_j | p, c) = \frac{1}{(t - t_j + c)^p}$ is suggested by Ogata (1988), whereas, many proposals have been made in the literature for the description of the spatial component. Among these, one of the most popular is certainly the one proposed by Ogata and Zhuang (2006), where $s(x - x_j, y - y_j | \delta, q) = [(x - x_j)^2 + (y - y_j)^2 + \delta]^{-q}$. The last form of the spatial part can be extended introducing also the information about the ruptured aftershock area $\delta(m) = d10^{\gamma m}$ (Kagan 2002). Hence, implementing the proposed functional forms and the conditions expressed in Eqs. (2, 3) into Eq. (1) we obtain

$$\begin{aligned} \lambda(x, y, t | \mathcal{H}_t) = & \mu + \sum_{j: t_j < t} K 10^{\alpha(m_j - m_0)} \frac{(p-1)c^{p-1}}{(t - t_j + c)^p} \frac{q-1}{\pi} \\ & (\delta(m_j))^{q-1} [(x - x_j)^2 + (y - y_j)^2 + \delta(m_j)]^{-q}. \end{aligned} \quad (4)$$

Therefore, the set of parameters is $\vec{\theta} = (p, \alpha, c, K, \mu, d, q, \gamma)$. It is important to note that we are assuming independence between magnitude among triggered events (Petrillo and Zhuang 2022, 2023).

The fundamental problem is to estimate the parameters of the model that best fit observed experimental data. Hence, in the frequentist statistical framework, the unknown parameter set $\vec{\theta}$ is estimated using likelihood maximization methods (Lippiello et al. 2014; Ogata 1998, 1983; Ogata and Zhuang 2006). As a result of this procedure, once the optimal set of parameters is obtained, it can be implemented in the model and managed for a possible forecast task. However, this procedure can lead to severe errors in correct estimation of the parameters. Indeed, one of the causes may be the insufficiency of observed data due to the shortness of the seismic catalog considered: some parameters may demonstrate substantial fluctuations with minor alterations in log-likelihood, and the standard errors of parameter estimates in the ETAS model are notable and cannot be disregarded. Conventional standard error estimates derived from the Hessian matrix are revealed to be inaccurate when the observed space–time window is small (Wang et al. 2010). Lastly, it is necessary to specify that the numerical methods for the optimization of parameters bring with them major biases, such as: the isotropy of the spatial distribution, the infinite spatial kernel and the short-term aftershock incompleteness. Spatial isotropy often holds for small magnitude earthquakes, however for large magnitude events it can lead to an underestimation of the productivity parameter α (Grimm et al. 2022; Hainzl et al. 2008, 2013). Even the infinite spatial kernel can cause an underestimation of productivity due to an unrealistic far trigger impact of small magnitudes (Grimm et al. 2022). Finally, the obscuring of the short-term aftershock is caused by an overlapping of coda waves immediately after the occurrence of a seismic event and can lead to an overestimation of Omori parameter c and an underestimation of productivity parameter α (Hainzl 2022; Petrillo and Lippiello 2020; Seif et al. 2017).

To overcome these problems, alternative methods to the maximum likelihood estimation (MLE) can be used, such as in Seif et al. (2018) for a systematic study on foreshocks activity or in other epidemic-like models where it is difficult to calculate the likelihood function (Petrillo and Lippiello 2020).

A further approach that is on the same line as MLE is certainly the Bayesian optimization of the model parameters. The Bayesian approach to parameter sampling allows bypassing many of the MLE problems by taking into account the uncertainty of the parameters to be estimated (Omi et al. 2015; Shcherbakov 2014) and avoiding simulation biases. In this case, the parameters are considered as a random variables. In practice, the Bayesian estimation process does not return a vector of parameters θ , but a distribution, given by the training set available Y . Before the observations, one assigns an a priori probability distribution $\pi(\vec{\theta})$ to each parameter θ_i on the basis of previous knowledge. After the observations, a correction of the judgment can be done and one assigns an a posteriori probability distribution for each parameter. In the case that there is no previous knowledge about the prior parameter distribution, it can be chosen as non-informative. In practice, parameters are drawn from a flat distribution within a given range determined by mathematical constraints.

Indicating by $P(\cdot)$ the general operator for some random variable, the “a posteriori” distribution $P(\vec{\theta}|Y)$, i.e., the probability of observing a certain set of parameters $\vec{\theta}$ given a certain observation data Y , can be written, thanks to Bayes’ theorem, as:

$$P(\vec{\theta}|Y) = \frac{P(Y|\vec{\theta})\pi(\vec{\theta})}{P(Y)} \propto P(Y|\vec{\theta})\pi(\vec{\theta}), \quad (5)$$

where $P(Y|\vec{\theta})$ is the probability to observe a set of data Y given a certain set of parameter $\vec{\theta}$, namely, the likelihood function, $\pi(\vec{\theta})$ is the prior distribution and $P(Y)$ is the probability of observing Y without any condition. Therefore, by obtaining the $P(\vec{\theta}|Y)$ distributions of the model parameters, bringing with them all the uncertainty due to their estimation, a forecast protocol can be implemented.

Obviously, from a practical point of view, carrying out this procedure analytically could be complicated. For this reason, it is possible to adopt numerical protocols, as Markov-Chain Monte Carlo (MCMC), to solve the proposed problem (Ross 2021).

Recent approaches to Bayesian parameter estimation have been successfully implemented to perform seismic forecast tests. As an example, in Holschneider et al. (2012) the authors show the validity of the modified Omori law both in the short and long term by means of a Bayesian setting. Moreover, in Ross (2021); Ross and Kolev (2022) Ross optimizes the parameters of the ETAS model with a Bayesian approach through the latent variable method. Another interesting Bayesian approach has also been applied to the Italian catalog demonstrating the methodology by retrospective early forecast of seismicity associated with the 2016 Amatrice seismic sequence

activities in central Italy (Ebrahimian and Jalayer 2017). Furthermore, Molkenthin et al. (2022) introduces a non-parametric representation of the spatially varying ETAS background intensity through a Gaussian process prior with which they are able to obtain the “a posteriori” distributions of the parameters presenting results on L’Aquila (Italy) event as a case study. Finally, a very interesting approach has been used by Shcherbakov et al. (2019) for forecasting the magnitude of the largest expected earthquake by combining the Bayesian method with the extreme value theory.

Although much effort has been made, the Bayesian estimation in statistical seismology remains an approach yet to be fully explored. In fact, the Bayesian estimation in Ross (2021) and consequently the forecast experiment are obtained through the use of latent variables justifying the use of this method raising doubts that the direct procedure through MCMC of the parameters works correctly.

In this paper, we carry out an illustration of Bayesian forecast on the aftershocks of 3 mainshocks with magnitude greater than 7 contained in the Southern California Catalogue (Hauksson et al. 2012). Bayesian parameter optimization is performed directly and not through the use of hidden variables. The Likelihood calculation is carried out with the innovative method proposed by Lippiello et al. (2014) which allows an accurate, stable, and relatively fast parameter inversion. This allows us to bypass the use of latent variables and to avoid bias on the estimation of the parameters due to the incorrect numerical estimation of Likelihood. To strengthen the validity of Bayesian forecast, we plot Molchan diagrams and calculate forecast scores by means of *S*-test and *L*-test. We would like to underline that, by means of this conceptually simple forecast experiment it is possible to confirm and strengthen all the results in literature on Bayesian forecast.

Methods

Likelihood and Bayesian estimation

Defining Σ and $[t_{in}, t_{fin}]$ as the spatial and the temporal region, respectively, the log-likelihood function for the ETAS model (LL) is given by:

$$LL = P(Y|\theta) = \sum_{i \in \Sigma \times [t_{in}, t_{fin}]} \log \lambda(x_i, y_i, t_i) - \int_{\Sigma} d\Sigma \int_{t_{in}}^{t_{fin}} \lambda(x, y, t) dt. \quad (6)$$

There is no difficulty in computing the first term, however, approximations are needed to solve the integral. For example, Schoenberg (2013) assumes that both t_{fin} and the spatial area Σ tend to infinity. As shown in Lippiello et al. (2014), however, these approximations can be too crude leading to bias in the estimation of the parameters.

In fact the limit $t_{fin} \rightarrow \infty$ does not represent a significant advantage from the computational point of view, whereas $\Sigma \rightarrow \infty$ introduces biases for parameter estimation. In particular, it is easy to solve the integral in polar coordinates, but the shape of the region Σ usually breaks the circular symmetry. For this reason, we use the novel method proposed by Lippiello et al. (2014) for the computation of the likelihood in Eq. (6). LL clearly depends on the parameters of the model and in the MLE frequentist approach, if a sufficiently large parametric region is explored, maximization can be achieved. In the Bayesian framework, we are interested in the *a posteriori* distribution $P(\vec{\theta}|Y)$. Defining $\pi(\vec{\theta})$ as a flat prior distribution of a certain parameter, in this study we employ for the MCMC method a component-wise Metropolis–Hastings (MH) algorithm (Hastings 1970; Metropolis et al. 1953). The MH algorithm produces a Markov chain on each parameter of our model through the following rules:

- Choose randomly a parameter of the model θ_i ,
- Extract a random number ϵ such that $\epsilon \sim \mathcal{N}(0, \sigma)$,
- Perform the parameter update $\theta'_i = \theta_i e^\epsilon$.

The decision to accept or reject the value of the new parameter is made computing the so-called Metropolis–Hastings acceptance ratio (MHAR):

$$\rho(\theta'_i, \theta_i) = \left(\frac{P(Y|\theta_1, \dots, \theta'_i, \dots, \theta_{N_p})}{P(Y|\theta_1, \dots, \theta_i, \dots, \theta_{N_p})} \times \frac{\pi(\theta_1, \dots, \theta_i, \dots, \theta_{N_p} | \theta_1, \dots, \theta'_i, \dots, \theta_{N_p})}{\pi(\theta_1, \dots, \theta'_i, \dots, \theta_{N_p} | \theta_1, \dots, \theta_i, \dots, \theta_{N_p})} \right), \quad (7)$$

where $P(Y|\theta_1, \dots, \theta'_i, \dots, \theta_{N_p})$ is the likelihood of the observed sequence and N_p is the total number of the parameters, in this study chosen as $N_p = 8$. The new candidate θ'_i will be accepted with probability $\min(1, \rho(\theta'_i, \theta_i))$, i.e., the algorithm accepts new values θ'_i only when the new likelihood is greater than the previous value, but an important feature is that the new proposed value may be accepted also if the ratio is decreasing, similar to stochastic optimization methods. It can be shown (Beck and Au 2002; Ebrahimian and Jalayer 2017) that, if $\vec{\theta}^n$ is distributed as $P(\cdot|Y)$ also the following sample is distributed as $P(\cdot|Y)$. In fact, the probability density function of the $\vec{\theta}^n$ given Y can be written as:

$$P(\vec{\theta}^n|Y) = \int_{\Omega} P(\vec{\theta}^{n+1}|\vec{\theta}^n)P(\vec{\theta}^n|Y)d\vec{\theta}^n = P(\theta^{n+1}|Y), \quad (8)$$

where the calculation was performed using the equality $\min(1, a/b)b = \min(1, b/a)a$. We empirically tune the standard deviation of ϵ distribution in order to obtain an acceptance rate around the optimal 45%. This

corresponds to $\sigma \simeq 0.1$. The choice of the ϵ parameter can be crucial both because if it is too large only a small part of the proposed test values will be accepted, and because if it is too small, the Markov chain will move very slowly and not all parameter space could be explored. The sampled values exhibit a lack of independence, i.e., each sampled value is correlated with its predecessor. This correlation arises from the proposed distribution being centered at the preceding value whenever a new value is proposed. As long as the prior value influences the likelihood of the proposed values, consecutive values remain dependent. To mitigate this autocorrelation phenomenon, one can plot an autocorrelation plot and determine the number of iterations required to minimize this effect. Therefore, in our sampling process, we never consider subsequent samples, but rather propose a sample only after a lag of 10 samples. Furthermore, we perform a 10% burn-in of the initial random walk.

This procedure is iterated until the desired sample of posterior distribution $P(\theta|Y)$ is obtained. In this study, we perform $n_{mc} = 10^6$ Monte Carlo steps. We would like to emphasize that a series of numerical tests has been conducted to evaluate the stability of parameter optimization, with a particular focus on the influence of the initial values of $\vec{\theta}$. The optimization process appears stable; however, a critical point identified is the distance of the initial parameter values from their optimal values, which slows down the convergence of the algorithm. This phenomenon necessitates an increased computational effort and time to achieve the desired optimization state, underscoring the importance of the careful selection of initial parameters in the optimization process.

Results

Forecast

The Bayesian inversion of the parameters has an advantage in statistical seismology, especially with regard to forecast techniques. The issue of selecting a specific spatio-temporal region for evaluating the log-likelihood through MLE introduces non-negligible bias to parameter estimation. Even if one confines the assessment of λ to the temporal interval $[t_{in}, t_{fin}]$ with $t_{in} > t_0$, events occurring in $[t_0, t_{in}]$ have contributed to triggering subsequent events at times and thus must be incorporated in the evaluation of the LL. Similarly, events occurring outside the finite space Σ are regarded as potential triggering events and are included in the assessment of the LL. The spatial bias, in particular, is discussed by Harte (2012). However, this is not the only issue. In general, in the MLE, the correct calculation of the standard errors of the ETAS parameters through the Hessian matrix of the log-likelihood function is guaranteed to be valid asymptotically for an infinite space-time window (Ogata 1978).

This asymptotic limit cannot be obtained with current seismic catalogs and the errors calculated through the Hessian matrix can be strongly biased. To avoid this problem, numerical approaches for error estimation can be used, for example, by using the root-mean-square of the errors in the parameter estimates for the simulations. However, due to the limited number of simulations performed, the sample from which to extrapolate the standard error can be small and therefore bootstrap procedures are usually applied. All these operations are computationally very demanding (Fox et al. 2016; Wang et al. 2010). Conversely, using the Bayesian inference, it is easy to take into account the intrinsic uncertainty of the model parameters, avoiding being too confident about a seismic forecast.

To perform a forecast with an ETAS model, suppose we have observed earthquakes in a space-time interval $\Sigma^0 \times [t_{in}^0, t_{fin}^0]$ and we want to predict the seismic occurrence in a future space-time interval $\Sigma^F \times [t_{in}^F, t_{fin}^F]$. The Bayesian forecast distribution is obtained marginalizing the distribution of Y^F given θ over the posterior distribution $P(\theta|Y)$, that is

$$P(Y^F|Y) = \int P(Y^F|\theta, Y)P(\theta|Y)d\theta, \quad (9)$$

but using Monte Carlo procedure, one can write

$$P(Y^F|Y) \propto \sum_n P(Y^F|\vec{\theta}^n, Y). \quad (10)$$

In practice, with each set of parameters, $\{\vec{\theta}^n | n \in \mathbb{N}\}$, generated from the posterior distribution by means of the MCMC sampler, it is possible to simulate an ETAS synthetic catalog from Eq. (4). The number of numerical catalogs obtained will be exactly equal to the number of steps performed by the Monte Carlo procedure, namely

- $\lambda(x, y, t|\vec{\theta}^1) \rightarrow \mathcal{C}_1$
- $\lambda(x, y, t|\vec{\theta}^2) \rightarrow \mathcal{C}_2$
- $\lambda(x, y, t|\vec{\theta}^3) \rightarrow \mathcal{C}_3$
- ...
- $\lambda(x, y, t|\vec{\theta}^{n_{mc}}) \rightarrow \mathcal{C}_{n_{mc}}$

where \mathcal{C}_k represents the k th synthetic catalog containing the information about the occurrence time, the location and the magnitude of an event. Since we are interested in predicting the number of earthquakes, we count the number of events produced by each single synthetic catalog in the space-time window of interest:

- $\mathcal{C}_1 \rightarrow N_1^{fore}$
- $\mathcal{C}_2 \rightarrow N_2^{fore}$

- $\mathcal{C}_3 \rightarrow N_3^{fore}$
- ...
- $\mathcal{C}_{n_{mc}} \rightarrow N_{n_{mc}}^{fore}$.

The details of the generation tree algorithm for simulating each individual catalog are described in the Section Methods. Each catalog i produced by each Monte Carlo step will predict a certain number of earthquakes N_i^{fore} in the forecast space-time region of interest.

The statistical test for earthquake number is the most simple conceptually and it is intended to measure how well the earthquake forecast number of events matches the observed one (N^{obs}). In order to understand if N^{obs} fall in one of the tails of the distribution it is possible compute the quantile δ_i^{under} defined as the fraction of N_i^{fore} smaller than observed number of events N^{obs} and δ_i^{over} as the fraction of N_i^{fore} greater than N^{obs} (Schorlemmer et al. 2007): the model is consistent with the observations if N^{obs} is in the center of the distribution N_i^{fore} .

Optimization and forecast results

An earthquake catalog is defined as a list of experimental earthquakes recorded in a certain geographical region and in a particular period. We consider the relocated southern California catalog (from 1981/01/01 to 2011/12/31) (Hauksson et al. 2012). For the optimization of the parameters, it is necessary to choose a period of time. In our algorithm, to reduce the parameters evaluation biases as in Harte (2012), we define three different times: the initial time t_0 , which represents the beginning of time span of the catalog, the learning start time t_{in} , and the end learning time t_{fin} . We clearly have that $t_0 \leq t_{in} < t_{fin}$. For the spatial part, we consider certain region $\Sigma \in \Sigma_0$. In general, for the optimization of the parameters, one can evaluate λ at time $t > t_{in}$ inside a certain region $\Sigma \in \Sigma_0$.

In Fig. 1, we present the correlation matrix among all the parameters of the model. It can be easily seen that the implementation of the burn-in of the first steps of the Markov chain, allows not to have anomalous drifting effects between the parameters. Figure 1 is a reference, in fact, depending on the parameterization used in Eq. (1) it is possible to observe or not observe correlations between the parameters of the model. In fact, for the temporal ETAS, other types of parameterization for the occurrence rate of aftershocks can be $K_0 = K_t K$, leading to $v(t) = K_0(t + c)^{-p} 10^{\alpha m}$ as proposed by Ogata (1998). However, the latter parameterization inherently includes correlations between parameters, which can complicate their estimation when employing MCMC techniques. As an example, in Fig. 2 we show the correlation between the parameters K and p of the model. In choosing to

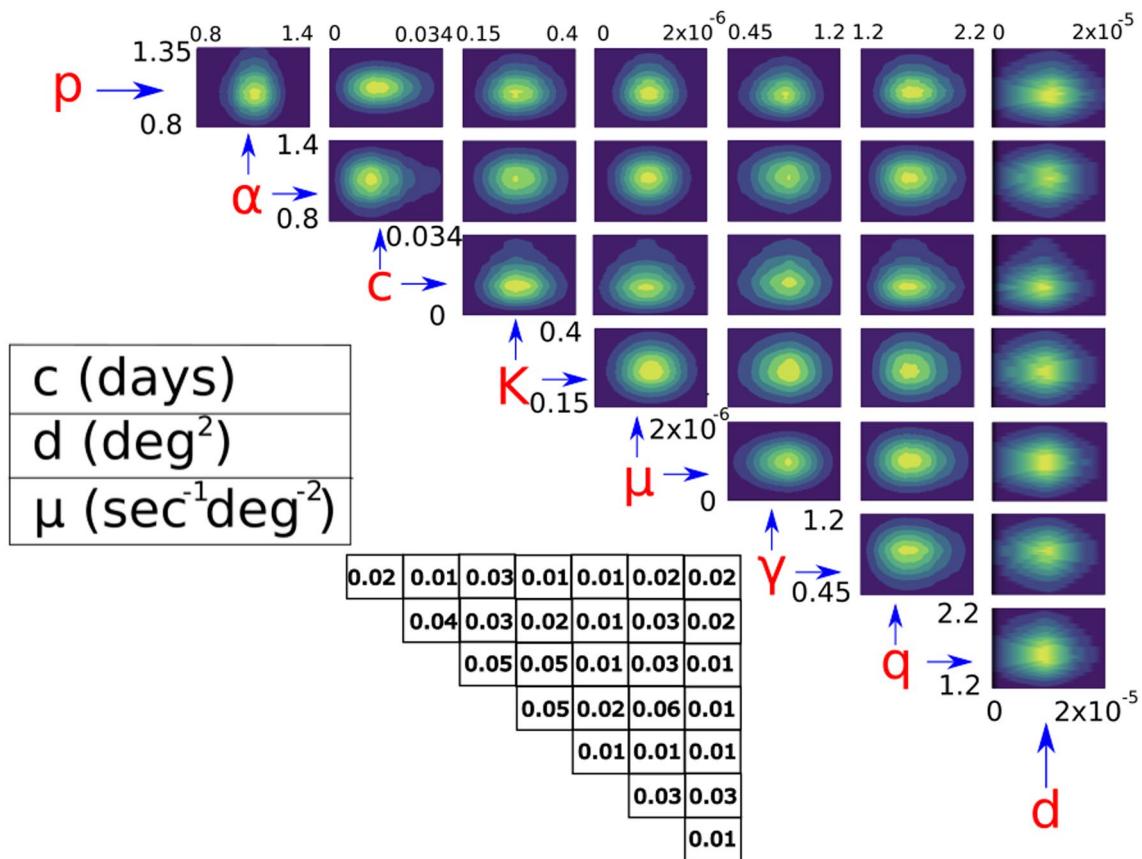


Fig. 1 The correlation matrix (bottom triangular table), together with the matrix contour plot (upper triangular table) among the parameters of the model considered in this study. c is measured in days, d in deg^2 , whereas μ is expressed in $\text{sec}^{-1}\text{deg}^{-2}$

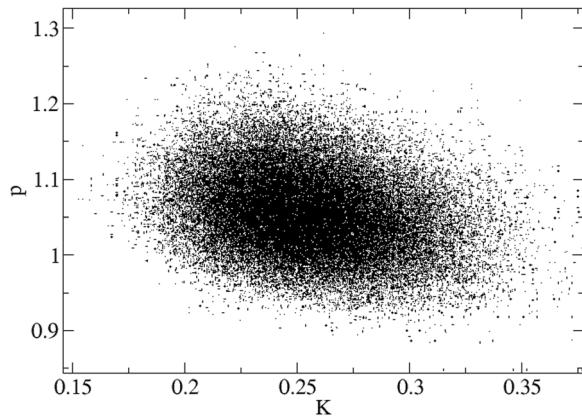


Fig. 2 The correlation between parameters K and p

parameterize the ETAS model with separate variables for K_t and K , we note an absence of correlation between these two parameters. Figure 2 also shows more clearly the non-existence of the starting transient of the random walk that have been eliminated from the chain. In Fig. 3

the posterior distributions of the parameters, including its average and standard deviation are presented: it can be seen that all 8 distributions are gaussian shaped, consistent with what showed in Ross (2021) by means of the latent variables approach. In addition, the MHAR of the algorithm is also shown in the same figure. The point where the curve becomes stationary gives an indication of the burn-in point of the data. The value of the acceptance ratio is particularly important and lot of proposals for the target variance ratio are proposed in the literature (Besag and Green 1993; Besag et al. 1995). We find empirically that, to obtain $\text{MHAR} \simeq 0.45$, σ must be set around 0.1.

In order to evaluate the effectiveness of the model with the parameters estimated by means of Bayesian inference, we carry out a retrospective forecast after the 1992 Landers earthquake, the 1999 Hector Mine earthquake and the El Mayor-Cucapah earthquake (sometimes called “Baja California earthquake”). In particular, the last mainshock occurred on April 4th of the year 2010, it was centered in Baja California and felt also throughout Southern California, Arizona and Nevada. The

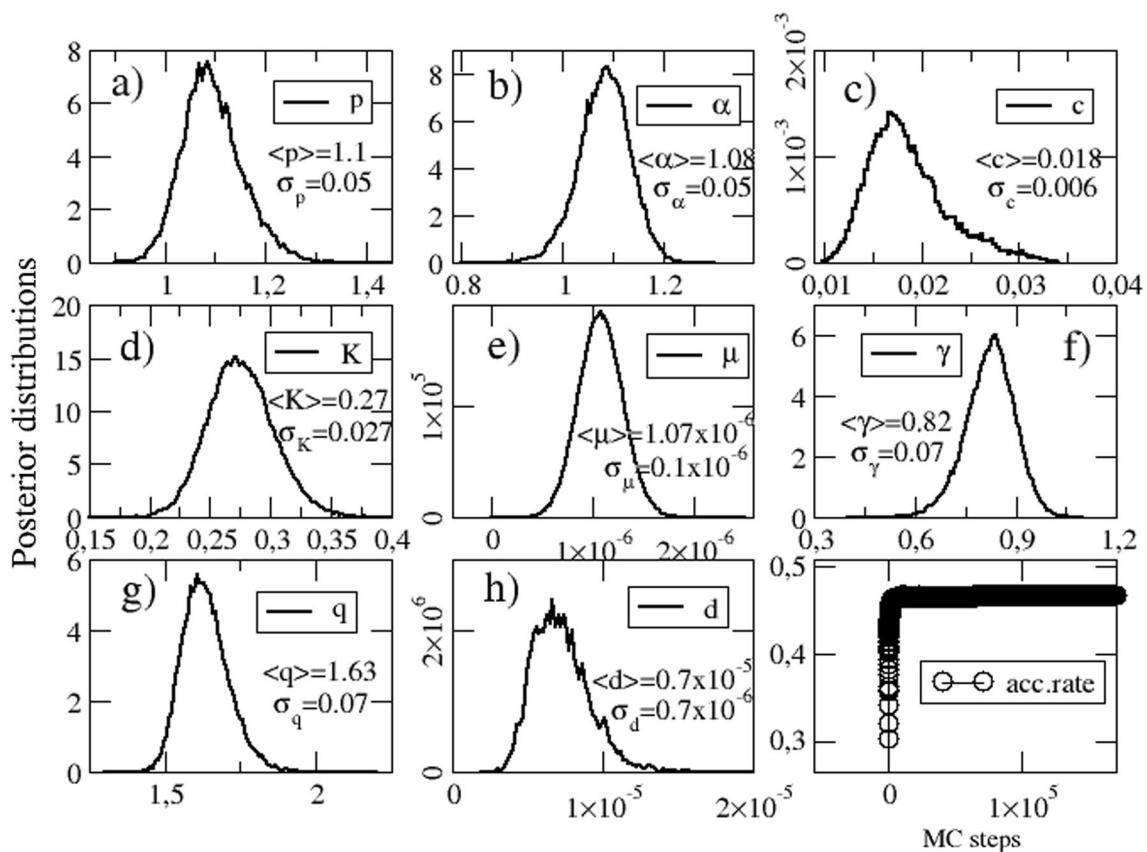


Fig. 3 From **a** to **h**, posterior distributions for the eight parameters of the ETAS model are defined in Eq. (1) obtained simulating 1,000,000 Monte Carlo steps. The parameter c is measured in days, d is measured in deg^2 , whereas μ is expressed in $sec^{-1}deg^{-2}$. The last plot at the bottom right represents the acceptance rate calculated with the MH acceptance ratio

spatial localization of the events, in the time-window of one month after each earthquake, is shown in Fig. 4. The observation history comprises includes one month prior to each event, i.e., we consider a distinct learning period (one month) for each sequence under consideration. Consequently, the posterior distribution of the parameters utilized for Bayesian forecast may vary slightly. From a purely qualitative point of view, in Fig. 5 the instrumental catalogs one month before and after each of the three earthquakes considered in the study is plotted. In order to quantitatively assess the quality of the model, for each set of parameters obtained by Bayesian inference, we count the total number of the observed earthquakes from the catalog produced by the ETAS model, $N_i^{fore,ETAS}$ and from the uniform model $N_i^{fore,UNI}$

The experimental catalogs present missing events immediately after the occurrence of high-magnitude earthquakes. It is now well established that observed incompleteness can be attributed to the overlapping of coda waves which causes a blind time where small events are not detected (de Arcangelis et al. 2018; Hainzl 2016; Helmstetter et al. 2006; Lippiello et al.

2016, 2019). In order to minimize the effect of short-term aftershock incompleteness, we test a forecast restricted to earthquakes with magnitude greater than a minimum magnitude m_L . In Fig. 6a, b, the Bayesian ETAS model number of earthquakes predicted, together with the Poissonian model number and the standard, non-Bayesian ETAS model, is plotted for $m_L = 4$ and for $m_L = 5$, respectively, for the 1992 Landers earthquake. Similar results can be found in Figs. 7, 8 for 1999 Hector Mine and Baja California event, respectively. We did not consider lower magnitudes smaller than 4 for the forecast due to the retrospective very stringent assumptions on the completeness magnitude m_c (the completeness magnitude is assumed $m_c = 3$ for whole sequence). For example, for the Baja California event, this is justified by a lack of seismic stations at south of the border (Hauksson et al. 2010). It can be noted that the forecast distribution by the Bayesian ETAS model is more spread out (less uncertain) than its standard ETAS counterpart since it counts the uncertainty of the model parameters. The forecast of the Poissonian model strongly underestimates the total number of

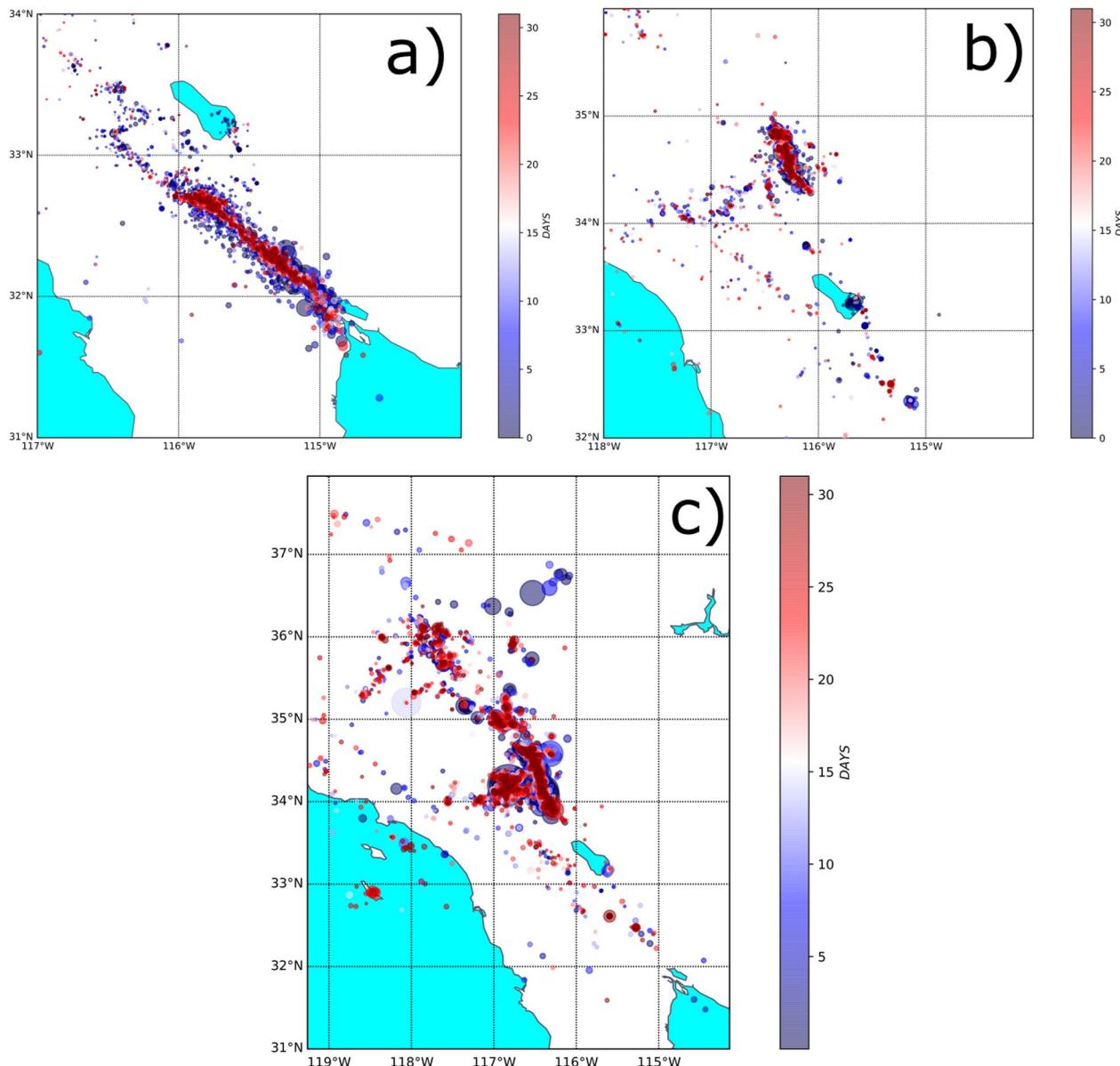


Fig. 4 The location of the earthquakes of the 2010 Baja California **a**, 1999 Hector Mine **b** and 1992 Landers, one month after the event took place. The size of the circles is related to the magnitude of the plotted event, while the color indicates the occurrence time of the events since the mainshock is occurred, as shown in the legend. The plot is obtained using Python from basemap toolkit (Hunter 2007)

earthquakes; conversely, the Bayesian ETAS model is consistent with the true number of earthquakes that occurred. Regarding the comparison with the standard ETAS model, we note that both models equally underestimate the number of aftershocks for the Landers earthquake but underperform compared to the Bayesian ETAS model for the other two mainshocks. This discrepancy may be attributed to the fact that the uncertainty in forecast for the Bayesian model is strictly related to parameter estimation, leading to a more

realistic standard deviation in the distribution. Conversely, the standard ETAS model derives its intrinsic randomness from the model's stochasticity, despite the fixed values of the parameters.

A sanity check of successful optimization of the spatial parameters for the ETAS model is also performed. In Figs. 9, 10, 11a, we show a comparison between the observed events and those predicted by a single ETAS realization, considering only earthquake magnitude greater than 4, always one month after the occurrence

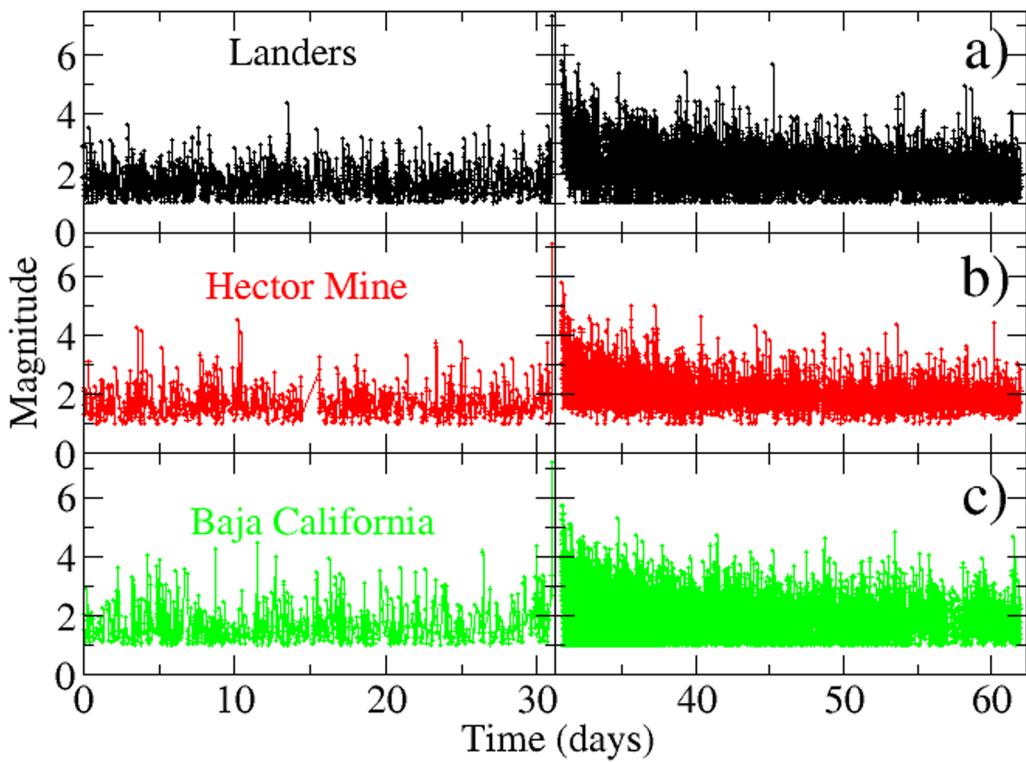


Fig. 5 Instrumental seismic catalogs one month before (left panels) and after (right panels) for **a** 1992 Landers **b** 1999 Hector Mine and **c** 2010 Baja California earthquake

of the each mainshock and in the spatial window considered for each event. Small inconsistencies observed in the spatial distribution of events are due to the non-isotropy of the spatial kernel of the ETAS model. Also, in Figs. 9, 10, 11b, we show a cloud forecast (10000 ETAS catalogs) for the same spatio-temporal period. Regarding the spatial predictability test, we plot the Molchan error diagrams of the location forecasts for the Bayesian ETAS and the non-homogeneous Poisson model in Fig. 12. The Molchan diagram, well defined also in Zechar and Jordan (2008), represents a diagram designed for evaluating earthquake forecast ability and show the fraction of missed earthquakes versus the fraction of space occupied by the observation. In particular, discretizing the target space in J cells of dimension $[0.1, 0.1]$ and calling $N(C_j)$ the number of observed events in the cell j , the alarm rate (AR) is defined as:

$$AR = \frac{\sum_j I(\lambda_j > \lambda_{th})}{\# \text{cells}}, \quad (11)$$

whereas the missing rate (MR) can be written as:

$$MR = \frac{\sum_j I(\lambda_j < \lambda_{th})I(N(C_j) \geq 1)}{\sum_j I(N(C_j) \geq 1)}, \quad (12)$$

where λ_j represents the occurrence rate in the j cell, λ_{th} the alarm threshold and I is the logic function, namely, $I(A) = 1$ if statement A is true and, otherwise $I(A) = 0$.

The line $MR = 1 - AR$ in Fig. 12 represents the unskilled alarm and can be noted that the Bayesian ETAS model outperform the non-homogeneous Poisson model.

S-test and L-test

To evaluate the likelihood of observed events given the forecast, we carry out the *L*-test Zechar (2010). To complete the test, we produce a set of synthetic catalogs $\Omega_k = \{\omega_k(i, j)\}$, where $\omega_k(i, j)$ represents the number of simulated earthquakes in the bin (i, j) in the synthetic catalog k . For each Ω_k , we compute the joint log-likelihood L_k obtaining a distribution. Comparing the observed joint log-likelihood with the simulated joint log-likelihood, one can understand if the observation is consistent with the proposed forecast. Numerically, it is possible to compute the quantile score γ defined as:

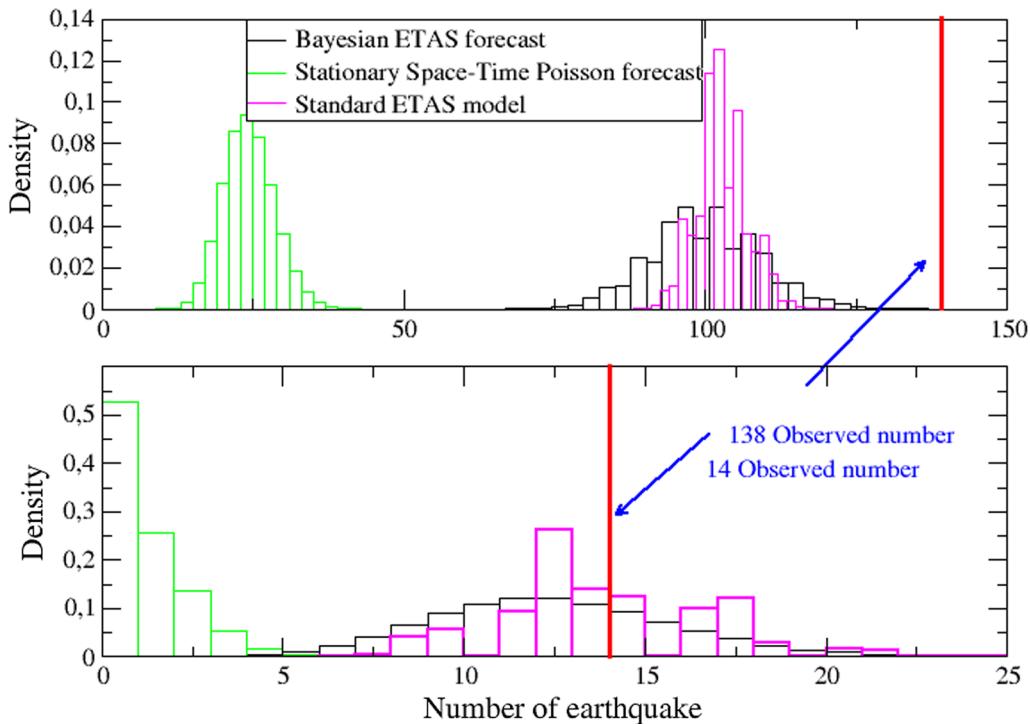


Fig. 6 (Upper panel) Forecast distribution for the number of earthquakes with magnitude $m \geq 4$ and (lower panel) for the number of earthquakes with magnitude $m \geq 5$ occurred 1 month after the 1992 Landers earthquake. The spatial window is $[(-119.26, -114.14), (31, 37.95)]$. Black histogram represents the Bayesian ETAS forecast, green histogram the Stationary space-time Poissonian forecast, the magenta histogram the standard ETAS model, whereas the red line represents the observed number of earthquakes (138 for $m \geq 4$ and 14 for $m \geq 5$)

$$\gamma = \frac{|\{L_k | L_k \leq L\}|}{|\{L\}|}, \quad (13)$$

where $|\cdot|$ represents the number of the elements in the ensemble $\{\cdot\}$. If the quantile score falls in the critical region (defined by α), the observation is inconsistent with the forecast.

Differently, the S -test isolates the spatial component of the forecast and test the consistency of spatial rates with observed events. For this purpose, one has $\Omega = \{\sum_i \omega(i, j)\}$, and $\Lambda = \{\frac{N_{obs}}{N_{syn}} \sum_i \lambda(i, j)\}$, where the sum extends over all magnitudes and N_{obs}, N_{syn} are the number of observed events in the instrumental catalog and synthetic catalog, respectively. As in L -test, one can compute observed joint log-likelihood $S = L(\Omega | \Lambda)$ and compare this value to the distribution of the simulated joint log-likelihood $S_k = L(\Omega_k | \Lambda)$. If the quantile score, ζ ,

$$\phi_i = \frac{\mu(x_i, y_i)}{\mu(x_i, y_i) + \sum_{j: t_j < t} \frac{KK_t K_r 10^{\alpha(m_j - m_0)}}{(t - t_j + c)^p} (\delta(m_j))^{q-1} [(x - x_j)^2 + (y - y_j)^2 + \delta(m_j)]^{-q}} \quad (14)$$

as in Eq. (13) is greater than the significance $\alpha = 0.025$, then the observed spatial distribution seismicity is consistent with the forecast.

In our simulation, both the likelihood of observed events given the forecast and the spatial distribution of the seismicity are consistent with the forecast. The results are presented in Table (1).

Background seismicity

In Fig. 13, the background activity of the instrumental catalog is compared against the one obtained by means of the ETAS model. The probability that an event belongs to the background is calculated by means of the method developed by Zhuang et al. (2002) with which it is possible to compute the probability that a certain event i of the catalog belongs to the background seismicity. In particular,

represents the probability that the i th event is a background event. Employing a similar procedure, it is

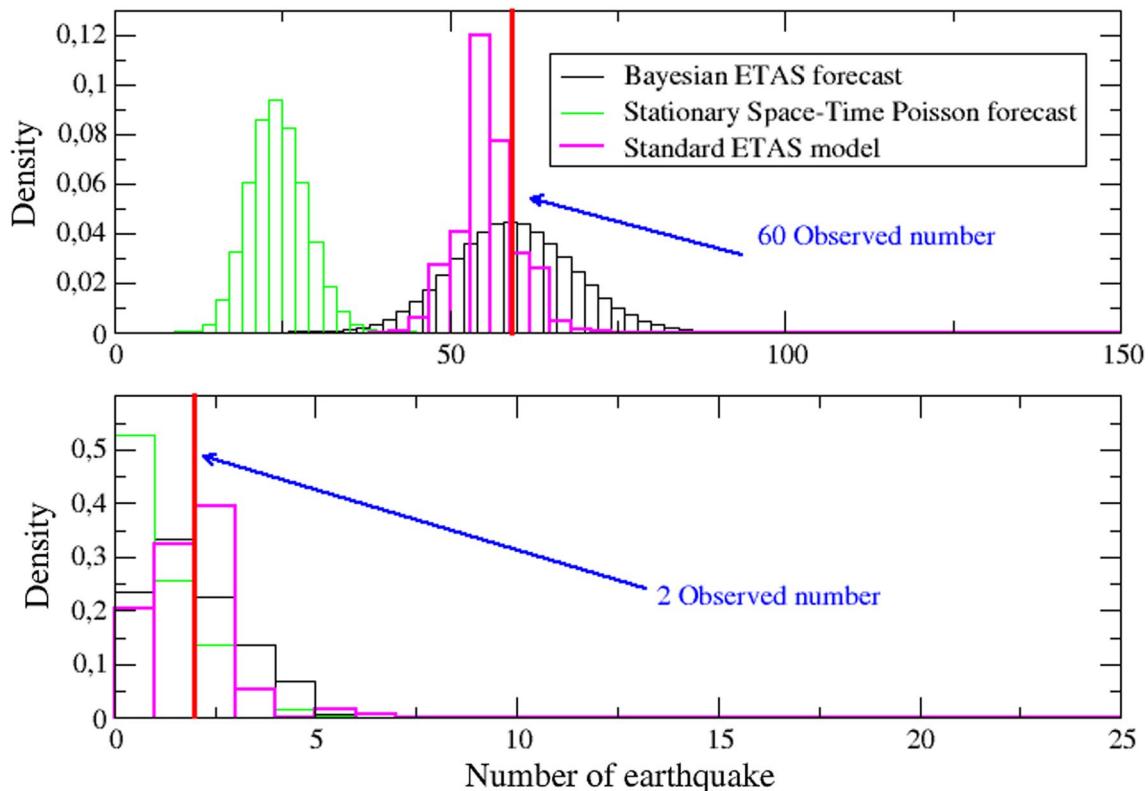


Fig. 7 (Upper panel) Forecast distribution for the number of earthquakes with magnitude $m \geq 4$ and (lower panel) for the number of earthquakes with magnitude $m \geq 5$ occurred one month after the 1999 Hector Mine earthquake. The spatial window is $[(-115, -117), (32, 36)]$. Black histogram represents the Bayesian ETAS forecast, green histogram the Stationary space-time Poissonian forecast, the magenta histogram the standard ETAS model whereas the red line represents the observed number of earthquakes (60 for $m \geq 4$ and 2 for $m \geq 5$)

possible to obtain the probability that an event is triggered by a previous one. In our case study, the seismicity of background shows prominent geographical trends in the northeast and northwest directions as confirmed by Hauksson et al. (2010).

Quantitatively, from the ETAS simulations we obtain an average value of $\bar{\mu} = 1 \times 10^{-6}$, while from the stochastic declustering on the instrumental catalog a value of 1.3×10^{-6} which confirms the correct Bayesian inversion procedure of the parameter μ .

Discussion and conclusions

In the last few decades, the now extensive literature on epidemic models applied to statistical seismology has confirmed the strong ability of the ETAS models to incorporate and capture the occurrence of earthquakes and thus become a baseline for seismic forecast (Jordan et al. 2011). The predict procedure through the use of these models, however, always involves the evaluation of log-likelihood. Unfortunately, the computation of the quantities that form the log-likelihood is computationally very demanding and also contemplates numerical

approximations due to the lack of an exact analytical solution of the integral in Eq. (6). In fact it has been shown (Lippiello et al. 2014) how the limits $t \rightarrow \infty$ and $\Sigma \rightarrow \infty$ can lead to strong bias in the evaluation of the integral.

From this point of view, with the Bayesian inference of the parameters, not only is it possible to bypass most of the problems related to the intrinsic uncertainty of the parameters, but it is also possible to avoid any bias due to the “art of simulation selection”. In fact, many efforts are being made on this line of research by estimating the parameters of the ETAS model with the Bayesian method (Ebrahimi and Jalayer 2017; Ross and Kolev 2022; Ross 2021; Shcherbakov 2014).

In this article, we carried out a retrospective forecast experiment performed after 3 major shocks of magnitude greater than 7 that occurred in Southern California, such as the 1992 Landers, the 1999 Hector Mine and the 2010 El Mayor-Cucapah earthquake (Baja California). A Bayesian inference to a spatio-temporal ETAS model with 8 free parameters by means a Metropolis–Hastings algorithm is carried

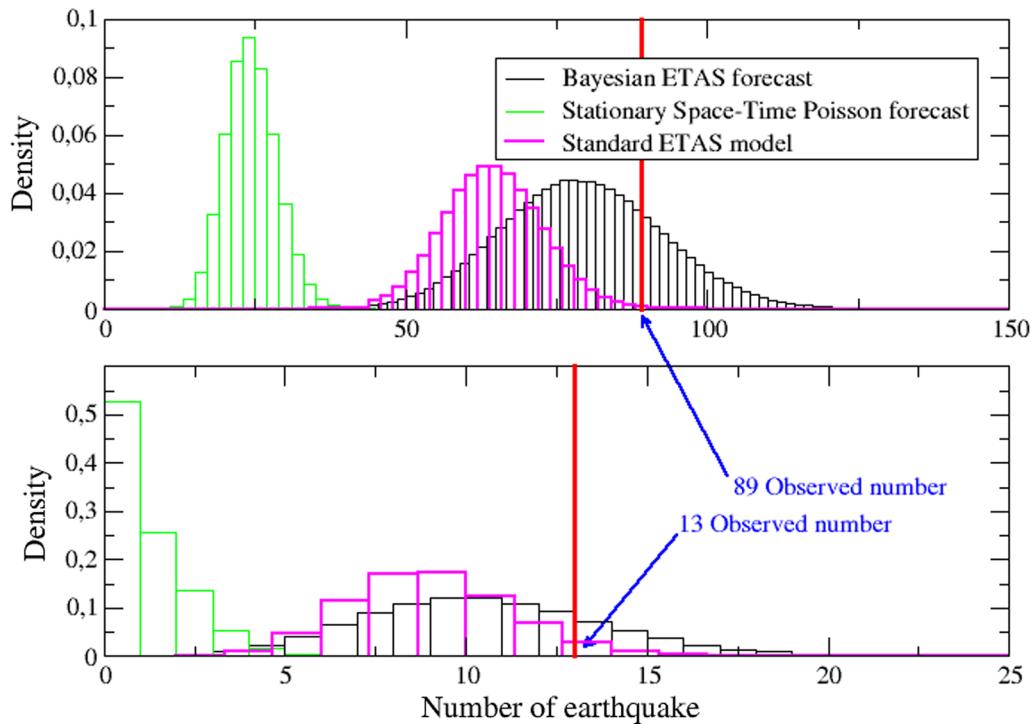


Fig. 8 (Upper panel) Forecast distribution for the number of earthquakes with magnitude $m \geq 4$ and (lower panel) for the number of earthquakes with magnitude $m \geq 5$ occurred one month after the 2010 Baja California earthquake, in particular from 2010-04-04 to 2010-05-04. The spatial window is $[(-117, -114), (31, 34)]$. Black histogram represents the Bayesian ETAS forecast, green histogram the Stationary space-time Poissonian forecast, the magenta histogram the standard ETAS model, whereas the red line represents the observed number of earthquakes (89 for $m \geq 4$ and 13 for $m \geq 5$)

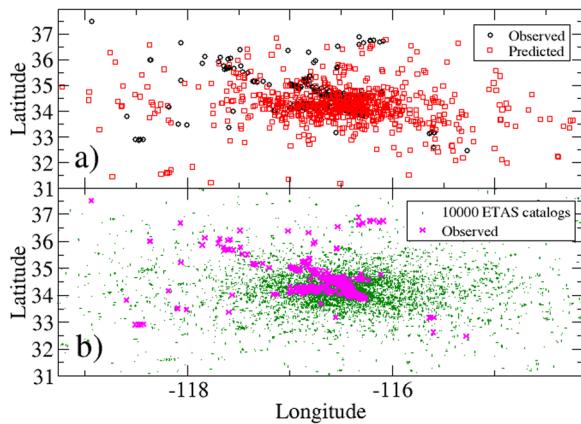


Fig. 9 **a** Earthquakes observed (995 events, black circles) and predicted (622 events, red squares) by a single realization of the ETAS model. **b** Earthquakes observed (648 events, magenta x) and 10000 realizations of the ETAS model (green dots). The total observation period is one month after the occurrence of 1992 Landers earthquake and in a spatial window $[(-119.26, -114.14), (31, 37.95)]$

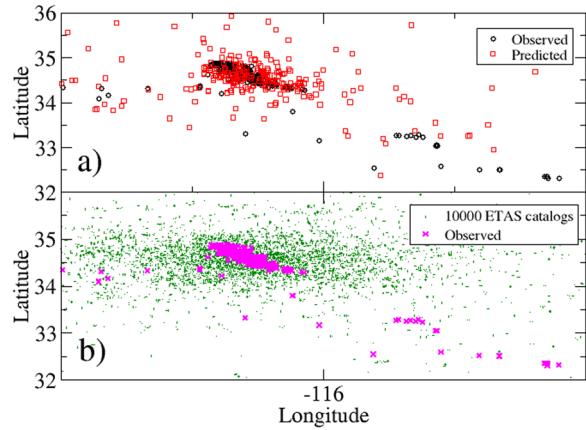


Fig. 10 **a** Earthquakes observed (680 events, black circles) and predicted (607 events, red squares) by a single realization of the ETAS model. **b** Earthquakes observed (648 events, magenta x) and 10000 realizations of the ETAS model (green dots). The total observation period is one month after the occurrence of 1999 Hector Mine earthquake and in a spatial window $[(-115, -117), (32, 36)]$

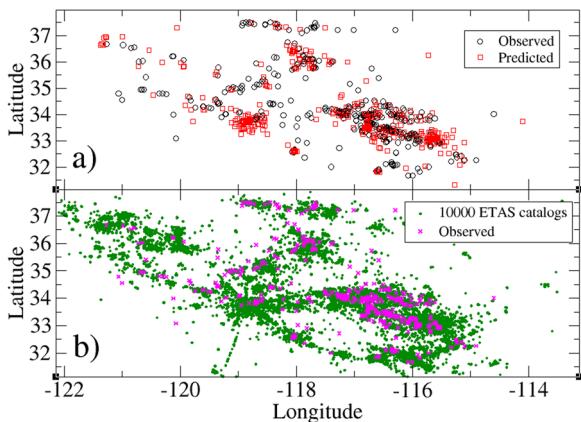


Fig. 11 **a** Earthquakes observed (648 events, black circles) and predicted (609 events, red squares) by a single realization of the ETAS model. **b** Earthquakes observed (648 events, magenta x) and 10000 realizations of the ETAS model (green dots). The total observation period is one month after the occurrence of 2010 Baja California earthquake and in a spatial window $[(-122, -113.5), (31, 38)]$

out. To assess the model optimization, we provide the Molchan diagram as a spatial predict test as well as a cloud forecast overlaying 10000 ETAS simulations. Other quantitative forecast scores, such as *S*-test and *L*-test are performed to strengthen the validity of the method and the prediction. In fact, the results show a good agreement between the predicted value and the actually observed one. The variability of the forecast distribution for the number of earthquakes allows a much less confident and more cautious forecast respect, for example, by means a direct estimate of the parameters using the MLE approach.

There are many technical points could be improved in our procedures. For example, we can introduce a latent variable for each event to identify its origin, as done in Ross (2021); Ross and Kolev (2022), or use Hamilton dynamic to fasten the MCMC sampler [e.g., Neal (2011)]. Furthermore, the background rate could be sampled as a Gaussian field (Molkenthin et al. 2022) or a Dirichlet process (Ross and Kolev 2022). Nevertheless, this direct Bayesian approach with trivial priors works well to produce stable forecasts of Southern California aftershocks.

Appendix A: Log-likelihood estimation

The log-likelihood function for in a spatio-temporal region $\Sigma \times [t_{in}, t_{fin}]$ is given by:

$$LL = LL_1 - LL_2, \quad (15)$$

with

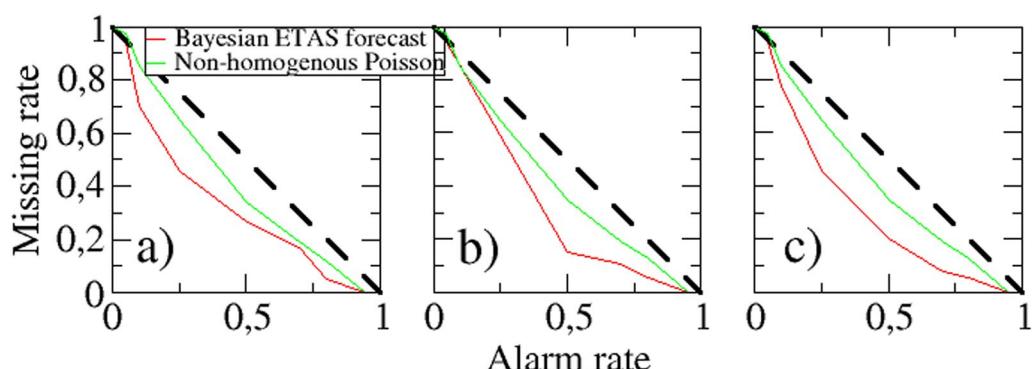


Fig. 12 The missed rate versus the alarm rate (Molchan diagram) for **a** 1992 Landers, **b** 1999 Hector Mine and **c** Baja California earthquake. The diagonal dotted line represents a completely unskilled alarm, whereas the red and green curves are the Bayesian ETAS and the non-homogeneous Poisson forecast, respectively

Table 1 L-test and S-test quantile for the three earthquakes analyzed in the paper for $m_L > 4$

	Landers	Hector Mine	Baja California
γ	0.25	0.55	0.36
ζ	0.37	0.41	0.6

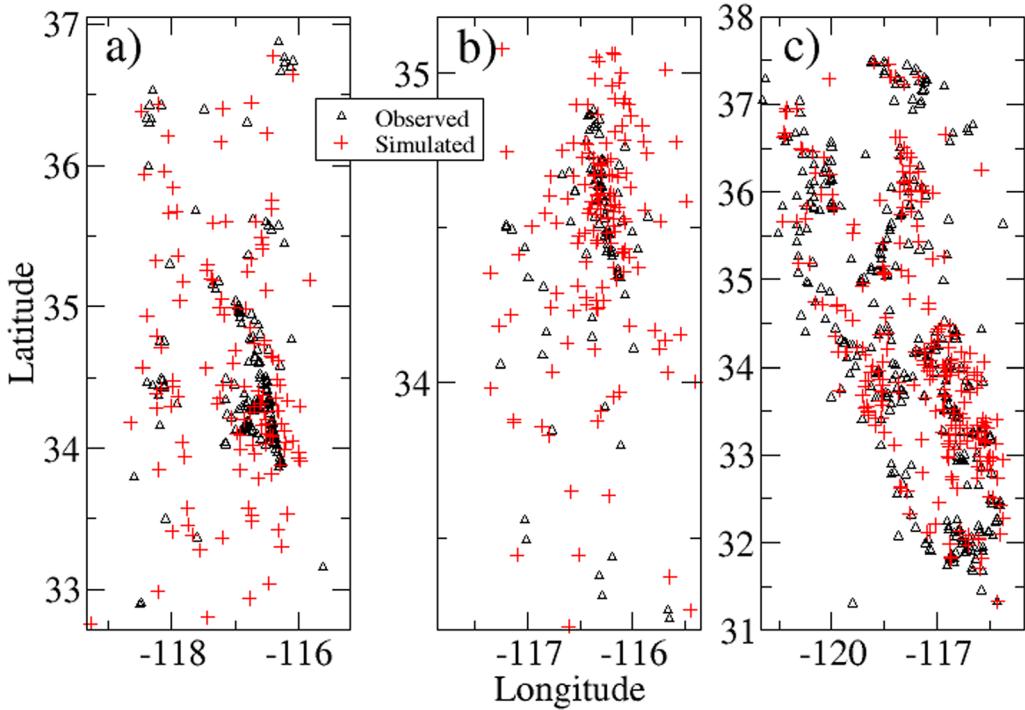


Fig. 13 Background seismicity for **a** 1992 Landers, **b** 1999 Hector Mine and **c** Baja California in the space-time window in the case study. The black triangles represent the background activity observed from the experimental catalog, while the red crosses are the background events obtained from the synthetic ETAS one

$$\begin{aligned} LL_1 &= \sum_{i \in \Sigma \times [t_{in}, t_{fin}]} \log \lambda(x_i, y_i, t_i), \\ LL_2 &= \int_{\Sigma} d\Sigma \int_{t_{in}}^{t_{fin}} \lambda(x, y, t) dt. \end{aligned} \quad (16)$$

In general, the computation of LL_1 is straightforward. Conversely, in order to evaluate LL_2 , it is necessary to perform some approximations. Schoenberg (2013) assumes that both t_{fin} and the size of the testing area tends to infinity. Under these approximations, LL_2 reduces to

$$LL_2 \simeq (t_{fin} - t_{in}) \int_{\Sigma} d\Sigma \mu(x, y) + K \sum_i 10^{\alpha(m_i - m_0)}. \quad (17)$$

However, these approximations appear to be too crude and may bias the likelihood estimate (Lippiello et al. 2014). The temporal integral can be solved

$$\int_{t_{in}}^{t_{fin}} \frac{1}{(t - t_j + c)^p} = \frac{c^{1-p}}{p-1} \left[1 - \left(1 + \frac{t_{fin} - t_j}{c} \right)^{1-p} \right], \quad \text{for } t_j \geq t_{in}. \quad (18)$$

The most challenging part is the resolution of the spatial integral. The difficulty for resolving it, in particular, is related to the fact that it is easy to solve the integral in polar coordinates, but the form of the region usually breaks the circular symmetry. Each epicenter is radially connected to a finite number of sampled points on the edge of the region. Then, the area of Σ between two successive radii is approximated with a circular sector. A good approximation can be obtained by getting many points on the edges, but this would lead to a high computational time.

To bypass this problem, we use the method proposed by Lippiello et al. (2014) which takes into account that the spatial integral is a rapidly decreasing function of the distance. In particular, for each earthquake it is possible to approximate the area of the target region using concentric annuli centered at the epicenter of the event (x_i, y_i) and with a certain width Δr . Therefore:

$$\begin{aligned} &K_r \int_{\Sigma} [(x - x_i)^2 + (y - y_i)^2 + \delta(m_i)]^{-q} \\ &\simeq \sum_k \theta(k, i) [\epsilon_i(r_{k-1}) \epsilon_i(r_k)], \end{aligned} \quad (19)$$

where $\epsilon_j(x) = (x^2/\delta(m_j) + 1)^{1-q}$ and $r_k = k\Delta r$. For the calculation of the angle $\theta(k,j)$ refer to Lippiello et al. (2014). In practice, the points on the border are not a fixed number but are a function of the distance of the earthquake from the border. This technique brings significant advantages both in terms of speed of execution and in terms of elimination of the approximation $\Sigma \rightarrow \infty$.

Appendix B: The generation tree algorithm

The generation of a forecast catalog is a standard procedure described in Zhuang et al. (2004); Zhuang and Touati (2015); de Arcangelis et al. (2016). The first step is setting the background seismicity $\mu(x,y)$. This represents the zero-th order generation in a self-exiting branching process and a certain number n_0 of events are created. Each of this element generates a certain number of offsprings, i.e., the aftershocks. The number n_1 and the occurrence space-time position of the aftershocks depends of $\lambda(x,y,t)$. In practice, the number of aftershocks will be extracted from a Poisson distribution with average dictated by the productivity law. For each offspring, the occurrence time is extracted based on the Omori–Utsu law and the location based on the spatial distribution. As a last step, the magnitude of the event is assigned obtaining the value from the GR law. This is the first-order generation of events. The previous step is repeated considering $n_j = n_{j-1}$ and it is iterated until $n_{j*} = 0$. Upon completion of the procedure, a synthetic catalog C will be obtained containing precisely the information regarding the magnitude, location and occurrence time of each event. For the production of the forecast, the procedure is simply repeated using a different set of parameters, obtained from the posterior distribution.

Acknowledgements

J.Z. would like to thank the MEXT Project for Seismology TowArd Research innovation with Data of Earthquake (STAR-E Project), Grant Number: JPJ010217. We would like to thank the two anonymous reviewers for their comments, which have significantly helped to improve the manuscript.

Author contributions

G.P and J.Z. contributed to the research, numerical results and writing of the manuscript.

Funding

This research activity has been supported by MEXT Project for Seismology TowArd Research innovation with Data of Earthquake (STAR-E Project), Grant Number: JPJ010217.

Availability of data and materials

The data considered in this study are available at https://github.com/caccioppoli/Bayesian_estimation_ETAS. The complete seismic catalog is available at <https://scedc.caltech.edu/data/alt-2011-dd-hauksson-yang-shearer.html>.

Declarations

Competing interests

The authors declare no competing interests.

Author details

¹Scuola Superiore Meridionale, Via Mezzocannone, 4, 80138 Napoli, Italy.

²The Institute of Statistical Mathematics, 10-3 Midoricho, Tachikawa, Tokyo 190-0014, Japan.

Received: 24 January 2024 Accepted: 29 April 2024

Published online: 24 May 2024

References

- Beck JL, Au SK (2002) Bayesian updating of structural models and reliability using Markov chain Monte Carlo simulation. *J Eng Mech* 128(4):380–391. [https://doi.org/10.1061/\(ASCE\)0733-9399\(2002\)128:4\(380\)](https://doi.org/10.1061/(ASCE)0733-9399(2002)128:4(380))
- Besag J, Green PJ, Higdon D, Mengerson K (1995) Bayesian computation and stochastic systems. *Stat Sci* 10(1):3–41. <https://doi.org/10.1214/ss/1177010123>
- Besag J, Green PJ (1993) Spatial statistics and Bayesian computation. *J R Stat Soc Ser B (Methodological)* 55(1):25–37. <https://doi.org/10.1111/j.2517-6161.1993.tb01467.x>
- Console R, Jackson D, Kagan Y (2010) Using the ETAS model for catalog declustering and seismic background assessment. *Pure Appl Geophys* 167:819–830. <https://doi.org/10.1007/s00024-010-0065-5>
- de Arcangelis L, Godano C, Grasso JR, Lippiello E (2016) Statistical physics approach to earthquake occurrence and forecasting. *Phys Rep* 628:1–91. <https://doi.org/10.1016/j.physrep.2016.03.002>
- de Arcangelis L, Godano C, Lippiello E (2018) The overlap of aftershock coda waves and short-term postseismic forecasting. *J Geophys Res Solid Earth* 123(7):5661–5674. <https://doi.org/10.1029/2018JB015518>
- Ebrahimi H, Jalayer F (2017) Robust seismicity forecasting based on Bayesian parameter estimation for epidemiological spatio-temporal aftershock clustering models. *Sci Rep* 7:44858. <https://doi.org/10.1038/s41598-017-09962-z>
- Fox EW, Schoenberg FP, Gordon JS (2016) Spatially inhomogeneous background rate estimators and uncertainty quantification for nonparametric Hawkes point process models of earthquake occurrences. *Ann Appl Stat* 10(3):1725–1756. <https://doi.org/10.1214/16-AOAS957>
- Grimm C, Kaser M, Hainzl S, Pagani M, Kuchenhoff H (2022) Improving earthquake doublet frequency predictions by modified spatial trigger kernels in the epidemic-type aftershock sequence (ETAS) model. *Bull Seismol Soc Am* 112(1):474–493. <https://doi.org/10.1785/0120210097>
- Hainzl S (2022) ETAS-approach accounting for short-term incompleteness of earthquake catalogs. *Bull Seismol Soc Am* 112(1):494–507. <https://doi.org/10.1785/0120210146>
- Hainzl S, Christoffersen A, Enescu B (2008) Impact of earthquake rupture extensions on parameter estimations of point-process models. *Bull Seismol Soc Am* 98(4):2066–2072. <https://doi.org/10.1785/0120070256>
- Hainzl S, Zakhrova O, Marsan D (2013) Impact of aseismic transients on the estimation of aftershock productivity parameters. *Bull Seismol Soc Am* 103(3):1723–1732. <https://doi.org/10.1785/0120120247>
- Hainzl S (2016) Rate-dependent incompleteness of earthquake catalogs. *Seismol Res Lett* 87(2A):337–344. <https://doi.org/10.1785/0220150211>
- Harte DS (2012) Bias in fitting the ETAS model: a case study based on New Zealand seismicity. *Geophys J Int* 192(1):390–412. <https://doi.org/10.1093/gji/gjs026>
- Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57(1):97–109. <https://doi.org/10.1093/biomet/57.1.97>
- Hauksson E, Stock J, Hutton K, Yang W, Vidal-Villegas JA, Kanamori H (2010) The 2010 MW 7.2 El Mayor-Cucapah earthquake sequence, Baja California, Mexico and Southernmost California, USA: active seismotectonics along the Mexican Pacific Margin. *Pure Appl Geophys* 168:1255–1277. <https://doi.org/10.1007/s00024-010-0209-7>

- Hauksson E, Yang W, Shearer PM (2012) Waveform relocated earthquake catalog for southern California (1981 to June 2011). *Bull Seismol Soc Am* 102:2239–2244. <https://doi.org/10.1785/0120120010>
- Helmstetter A, Kagan YY, Jackson DD (2006) Comparison of short-term and time-independent earthquake forecast models for Southern California. *Bull Seismol Soc Am* 96(1):90–106. <https://doi.org/10.1785/0120050067>
- Helmstetter A, Sornette D (2003) Importance of direct and indirect triggered seismicity in the ETAS model of seismicity. *Geophys Res Lett.* <https://doi.org/10.1029/2003GL017670>
- Holschneider M, Narteau C, Shebalin P, Peng Z, Schorlemmer D (2012) Bayesian analysis of the modified omori law. *J Geophys Res Solid Earth.* <https://doi.org/10.1029/2011JB009054>
- Hunter JD (2007) Matplotlib: a 2D graphics environment. *Comput Sci Eng* 9(3):90–95. <https://doi.org/10.1109/MCSE.2007.55>
- Jordan TH, Chen Y, Gasparini P, Madariaga R, Main I, Marzocchi W, Papadopoulos G, Sobolev G, Yamaoka K, Zschau J (2011) Operational earthquake forecasting: state of knowledge and guidelines for utilization. *Ann Geophys* 54:315–391. <https://doi.org/10.4401/ag-5350>
- Kagan YY (2002) Aftershock zone scaling. *Bull Seismol Soc Am* 92(2):641–655. <https://doi.org/10.1785/0120010172>
- Lippiello E, Giacco F, de Arcangelis L, Marzocchi W, Godano C (2014) Parameter estimation in the ETAS model: approximations and novel methods. *Bull Seismol Soc Am* 104(2):985–994. <https://doi.org/10.1785/0120130148>
- Lippiello E, Cirillo A, Godano C, Papadimitriou E, Karakostas V (2016) Real-time forecast of aftershocks from a single seismic station signal. *Geophys Res Lett* 43(12):6252–6258. <https://doi.org/10.1002/2016GL069748>
- Lippiello E, Petrillo G, Godano C, Tramelli A, Papadimitriou E, Karakostas V (2019) Forecasting of the first hour aftershocks by means of the perceived magnitude. *Nat Commun* 10(1):2953. <https://doi.org/10.1038/s41467-019-10763-3>
- Lombardi AM, Marzocchi W (2010) The ETAS model for daily forecasting of Italian seismicity in the CSEP experiment. *Ann Geophys* 53(3):155–164. <https://doi.org/10.4401/ag-4848>
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. *J Chem Phys* 21(6):1087–1092. <https://doi.org/10.1063/1.1699114>
- Molkenthin C, Donner RV, Reich S, Zoller G, Hainzl S, Holschneider M, Opper M (2022) GP-ETAS: semiparametric Bayesian inference for the spatio-temporal epidemic type aftershock sequence model. *Stat Comput* 32:32. <https://doi.org/10.1007/s11222-022-10085-3>
- Neal RM (2011) MCMC using hamiltonian dynamics. In: Brooks Steve, Gelman Andrew, Jones Galin, Meng Xiao-Li (eds) *Handbook of markov chain Monte Carlo*. CRC press, Boca Raton
- Ogata Y (1983) Estimation of the parameters in the modified Omori formula for aftershock frequencies by the maximum likelihood procedure. *J Phys Earth* 31(2):115–124. <https://doi.org/10.4294/jpe1952.31.115>
- Ogata Y (1978) The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Ann Inst Stat Math* 30(1):243–261. <https://doi.org/10.1007/BF02480216>
- Ogata Y (1988) Statistical models for earthquake occurrences and residual analysis for point processes. *J Am Stat Assoc* 83(401):9–27. <https://doi.org/10.1080/01621459.1988.10478560>
- Ogata Y (1998) Space-time point-process models for earthquake occurrences. *Ann Inst Stat Math* 50(2):379–402. <https://doi.org/10.1023/A:1003403601725>
- Ogata Y, Zhuang J (2006) Space-time ETAS models and an improved extension. *Tectonophysics* 413(1–2):13–23. <https://doi.org/10.1016/j.tecto.2005.10.016>
- Omi T, Ogata Y, Hirata Y, Aihara K (2015) Intermediate-term forecasting of aftershocks from an early aftershock sequence: Bayesian and ensemble forecasting approaches. *J Geophys Res Solid Earth* 120(4):2561–2578. <https://doi.org/10.1002/2014JB011456>
- Petrillo G, Lippiello E, Zhuang J (2023) Including stress relaxation in point-process model for seismic occurrence. *Geophys J Int.* <https://doi.org/10.1093/gji/ggad482>
- Petrillo G, Lippiello E (2020) Testing of the foreshock hypothesis within an epidemic like description of seismicity. *Geophys J Int* 225(2):1236–1257. <https://doi.org/10.1093/gji/gga611>
- Petrillo G, Lippiello E (2023) Incorporating foreshocks in an epidemic-like description of seismic occurrence in Italy. *Appl Sci* 13(8):4891. <https://doi.org/10.3390/app13084891>
- Petrillo G, Zhuang J (2022) The debate on the earthquake magnitude correlations: a meta-analysis. *Sci Rep* 12(1):20683. <https://doi.org/10.1038/s41598-022-25276>
- Petrillo G, Zhuang J (2023) Verifying the magnitude dependence in earthquake occurrence. *Phys Rev Lett* 131(15):154101. <https://doi.org/10.1103/PhysRevLett.131.154101>
- Ross GJ (2021) Bayesian estimation of the ETAS model for earthquake occurrences. *Bull Seismol Soc Am* 111(3):1473–1480. <https://doi.org/10.1785/0120200198>
- Ross JG, Kolev A (2022) Semiparametric Bayesian forecasting of SpatioTemporal earthquake occurrences. *Ann Appl Stat* 16(4):2083–2100. <https://doi.org/10.1214/21-AOAS1554>
- Schoenberg FP (2013) Facilitated estimation of ETAS. *Bull Seismol Soc Am* 103:601–605. <https://doi.org/10.1785/0120120146>
- Schorlemmer D, Gerstenberger MC, Wiemer S, Jackson DD, Rhoades DA (2007) Earthquake likelihood model testing. *Seismol Res Lett* 78(1):17–29. <https://doi.org/10.1785/gssrl.78.1.17>
- Seif S, Mignan A, Zechar JD, Werner MJ, Wiemer S (2017) Estimating ETAS: the effects of truncation, missing data, and model assumptions. *J Geophys Res Solid Earth* 122(1):449–469. <https://doi.org/10.1002/2016JB012809>
- Seif S, Zechar JD, Mignan A, Nandan S, Wiemer S (2018) Foreshocks and their potential deviation from general seismicity. *Bull Seismol Soc Am* 109(1):1–18. <https://doi.org/10.1785/0120170188>
- Shcherbakov R (2014) Bayesian confidence intervals for the magnitude of the largest aftershock. *Geophys Res Lett* 41(18):6380–6388. <https://doi.org/10.1002/2014GL061272>
- Shcherbakov R, Zhuang J, Zöller G, Ogata Y (2019) Forecasting the magnitude of the largest expected earthquake. *Nat Commun* 10:3956. <https://doi.org/10.1038/s41467-019-11958-4>
- Spassiani I, Petrillo G, Zhuang J (2024) Distribution related to all samples and extreme events in the ETAS cluster. Preprint. <https://doi.org/10.2254/essoar.169447347.74724727/v1>
- Wang Q, Schoenberg FP, Jackson DD (2010) Standard errors of parameter estimates in the ETAS model. *Bull Seismol Soc Am* 100(5A):1989–2001. <https://doi.org/10.1785/0120100001>
- Zechar JD, Jordan TH (2008) Testing alarm-based earthquake predictions. *Geophys J Int* 172(2):715–724. <https://doi.org/10.1111/j.1365-246X.2007.03676.x>
- Zechar JD (2010) Evaluating earthquake predictions and earthquake forecasts: a guide for students and new researchers. Community Online Resource for Statistical Seismicity Analysis. Available online: (<https://www.corssa.org>)
- Zhuang J (2011) Next-day earthquake forecasts for the Japan region generated by the ETAS model. *Earth Planets Space* 63(3):207–216. <https://doi.org/10.5047/eps.2010.12.010>
- Zhuang J (2012) Long-term earthquake forecasts based on the epidemic-type aftershock sequence (ETAS) model for short-term clustering. *Res Geophys* 2:e8. <https://doi.org/10.4081/rg.2012.e8>
- Zhuang J, Touati S (2015) Stochastic simulation of earthquake catalogs. Community Online Resource for Statistical Seismicity Analysis. Available online: [[link](https://www.corssa.org)] (<https://www.corssa.org>)
- Zhuang J, Ogata Y, Vere-Jones D (2002) Stochastic declustering of space-time earthquake occurrences. *J Am Stat Assoc* 97(458):369–380. <https://doi.org/10.1198/16214502760046925>
- Zhuang J, Ogata Y, Vere-Jones D (2004) Analyzing earthquake clustering features by using stochastic reconstruction. *J Geophys Res* 109(B5):B05301. <https://doi.org/10.1029/2003JB002879>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.