



# Inference for ETAS models with non-Poissonian mainshock arrival times

Aleksandar A. Kolev<sup>1</sup> · Gordon J. Ross<sup>2</sup>

Received: 12 April 2018 / Accepted: 26 November 2018 / Published online: 13 December 2018  
© The Author(s) 2018

## Abstract

The Hawkes process is a widely used statistical model for point processes which produce clustered event times. A specific version known as the ETAS model is used in seismology to forecast earthquake arrival times under the assumption that mainshocks follow a Poisson process, with aftershocks triggered via a parametric kernel function. However, this Poissonian assumption contradicts several aspects of seismological theory which suggest that the arrival time of mainshocks instead follows alternative renewal distributions such as the Gamma or Brownian Passage Time. We hence show how the standard ETAS/Hawkes process can be extended to allow for non-Poissonian distributions by introducing a dependence based on the underlying process' behaviour. Direct maximum likelihood estimation of the resulting models is not computationally feasible in the general case, so we also present a novel Bayesian MCMC algorithm for efficient estimation using a latent variable representation.

**Keywords** ETAS · Stress release · Renewal process · Hawkes process · Brownian passage times · RHawkes

## 1 Introduction

The Epidemic Type Aftershock Sequence (ETAS) model is commonly used for studying and forecasting the occurrence of earthquakes in a geographical region of interest (Ogata 1988). It assumes that earthquakes follow a self-exciting marked point process governed by a conditional intensity function  $\lambda(t|\mathcal{H}_t)$  which defines the instantaneous probability of an earthquake occurring at each time point  $t$  based on the historical earthquake sequence  $\mathcal{H}_t = \{(t_1, m_1), (t_2, m_2), \dots : t_i < t\}$ , where  $t_i$  and  $m_i$ , respectively, denote the time and magnitude of the  $i$ th previous earthquake.

The ETAS model can be viewed as a branching process where a sequence of so-called immigrant (also known as 'mainshock') earthquakes occurs according to a Pois-

son process with constant intensity  $\mu$ , with their magnitudes following a Gutenberg–Richter distribution. Each of these immigrant earthquakes then produces aftershocks (also known as 'offspring' or 'children') which can then produce further aftershocks, and so on. A visual example of a possible branching structure is shown in Fig. 1. Here events  $t_1$ ,  $t_6$  and  $t_{10}$  are the immigrants which initiated the other events. The events  $t_2$ ,  $t_3$  and  $t_5$  are children of  $t_1$ , while  $t_4$  is a child of  $t_3$ . Similarly  $t_7$  and  $t_9$  are offsprings of  $t_6$ , while  $t_{11}$  is caused by  $t_8$ , which is a child of  $t_7$ . There are no detected children events for  $t_{10}$ , although some might occur in the future.

Since the ETAS model assumes that the immigrant earthquakes follow a Poisson process with constant intensity  $\mu$ , this implies that they occur completely at random, i.e. that an immigrant event is equally likely to occur at each point in time, and that the time between each pair of immigrant events (known as the 'inter-arrival times') follows a time-independent Exponential( $\mu$ ) distribution. However, this conflicts with findings elsewhere in the seismology literature, where there is substantial doubt over whether the occurrence times of mainshock earthquakes is really Poissonian (Tahernia et al. 2014; Ordaz and Arroyo 2016; Marzocchi and Taroni 2014). Although ETAS immigrant events are not strictly equivalent to mainshocks as defined elsewhere in the seismology literature (since there is no requirement that an

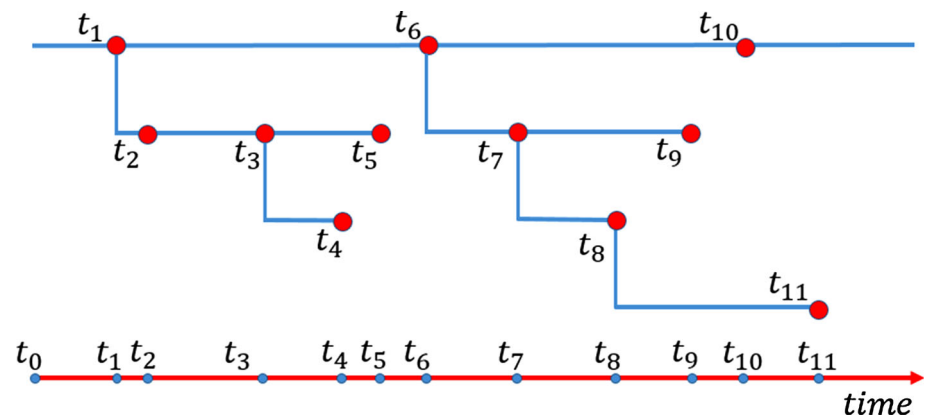
✉ Aleksandar A. Kolev  
aleksandar.kolev.15@ucl.ac.uk

Gordon J. Ross  
Gordon.Ross@ed.ac.uk

<sup>1</sup> Department of Statistical Science, University College London, London, UK

<sup>2</sup> School of Mathematics, University of Edinburgh, Edinburgh, UK

**Fig. 1** Example of a branching structure



ETAS immigrant should have larger magnitude than its offspring), this still seems to cast some doubt on the Poissonian assumption.

The concept of stress release (SR) suggests that the mainshock arrival times instead follow a renewal process that has a time-dependent hazard function, with inter-event times following a distribution such as the Gamma, Weibull, or Brownian Passage Times (BPT). Stress release models (SRM) were a representation of Reid's elastic rebound theory (Reid 1910) and were fully described by Isham and Westcott (1979) as a self-correcting point process which is updated after every event occurrence. They were introduced to seismology by Vere-Jones (1978) who developed them in order to address Reid's theory that earthquakes occur due to a release of energy which was previously accumulated strain energy along faults. SRMs were used in many locations to implement the elastic rebound theory due to their solid physical background. As outlined in Varini and Rotondi (2015), some of the examples of such implementations are present for the following countries: China (Yang et al. 2000; Liu et al. 1998; Xiaogu and Vere-Jones 1994), Greece (Rotondi and Varini 2006), Iran (Xiaogu and Vere-Jones 1994), Italy (Rotondi and Varini 2007; Varini and Rotondi 2015), Japan (Imoto 2001; Lu et al. 1999; Xiaogu and Vere-Jones 1994), New Zealand (Yang et al. 2000) and Taiwan (Zhu and Shi 2002).

SRMs are primarily applied to declustered sequences of mainshock events with large magnitudes, rather than to the full seismic sequences that are commonly used to fit ETAS models. In this paper we will develop a new class of ETAS models which we call SR-ETAS (stress release ETAS) which improve on standard ETAS models by incorporating time-dependent inter-arrival distributions. We explore two different formulations of SR-ETAS, which differ based on how they handle the inter-event time that is taken into account when calculating the immigrant event intensity. The first formulation is simpler to estimate but harder to simulate from, and addresses the Reid's elasticity rebound theory directly for all events in the catalogue. The second one is harder to estimate as it depends entirely on the branching structure as

it assumes that Reid's theory is applicable only for immigrant events, making direct maximum likelihood estimation impossible.

A model which is closely related to our SR-ETAS was proposed by Wheatley et al. (2016), who considered a Hawkes process with a renewal immigration process, which they call Renewal Hawkes (RHawkes). The authors proposed an Expectation Maximisation (EM) algorithm for parameter estimation. However, as pointed out by Wheatley (2017), Wheatley (2016) their approach crucially exploited the Markovian properties used by the Exponential offspring density  $g(\cdot)$  that they considered, which leads to instability when this is replaced by a heavy-tailed alternative such as the Omori law used in the ETAS model (Oakes 1975; Filimonov and Sornette 2015). To mitigate this, they suggest that such heavy-tailed densities should be approximated by a sum of weighted exponential kernels. Further, simulation studies were found that their EM algorithm performs poorly even for the more simplistic Renewal Immigration Hawkes process in the case where the offspring clusters are heavily overlapping, which is inevitably the case of seismic sequences. To correct this, (Chen and Stindl 2018) provided a direct maximum likelihood optimisation, as well as some conceptual corrections to the method proposed by Wheatley et al. However, both methods fail to address two fundamental issues. The first one is the potential multimodality of the ETAS mode likelihood. As discussed in Rasmussen (2013), Veen and Schoenberg (2008), Ross (2018a), such numerical instabilities can be tackled using an MCMC sampler. The second, and probably more important problem, is the lack of discussion regarding the numerical stability of the evaluation of Eq. 2. This approach is used if the intensity cannot be factorised into a single equation, i.e. it has to be evaluated as a ratio of two functions. The problem occurs since the denominator of Eq. 2 is approaching zero for large time lag.

Since the existing Expectation Maximisation (EM) and Direct Maximum Likelihood Estimation algorithms lead to either poor or limited estimation of the SR-ETAS model, we instead propose a novel Bayesian inference algorithm which

uses latent variables to allow for computationally efficient inference using a Gibbs sampler, which is an extension of that proposed for the standard Hawkes process by Ross (2018a).

The remainder of this paper proceeds as follows. In Sect. 2 we review the standard ETAS model in more detail. The SR-ETAS models are fully introduced in Sect. 3, and we discuss different choices for the immigrant process in Sect. 4. The methods for parameter estimation are present in Sect. 5. In Sect. 6, we introduce the goodness-of-fit tests which will be used to compare performance of SR-ETAS to standard ETAS models. Finally, in Sect. 7, we study the application of SR-ETAS models and compare its performance to standard ETAS using real earthquake data from the New Madrid and the North California seismic sequences.

## 2 Standard ETAS model

The standard ETAS model was firstly introduced by Ogata (Ogata 1988) and assumes earthquakes follow a marked point process with conditional intensity function:

$$\lambda(t|\mathcal{H}_t) = \mu + \sum_{t_i < t} g(t - t_i)\kappa(m_i) \tag{1}$$

where  $t_i$  and  $m_i$  denote the occurrence time and magnitude of earthquake  $i$ . All magnitudes are assumed to independently follow the Gutenberg–Richter law, which corresponds to a shifted Exponential( $\beta$ ) distribution with lower bound  $M_0$ . The  $\mu$  parameter specifies the intensity of the homogeneous point process governing the immigrant events, while  $g(\cdot)$  is a kernel function specifying how the effect of each earthquake on the intensity decays over time. It is usually taken to be the Omori law:

$$g(z) = \frac{k}{(z + c)^p}$$

where  $c$  and  $p$  are parameters controlling the decay rate, while  $k$  controls the average productivity. The magnitude kernel  $\kappa(m_i)$  determines how the magnitude of each earthquake affects the intensity and is usually defined as:

$$\kappa(m_i) = e^{\alpha(m_i - M_0)}$$

where  $\alpha$  provides similar functionality to those of  $k$ , and  $M_0$  is the catalogue’s magnitude of completeness, i.e. the minimum magnitude above which is considered that no events are missing due to physical limitations in the earthquake detection system. The unknown parameter set of the standard ETAS model is hence:  $\theta = \{\mu, \alpha, c, p, k\}$ .

Note that the form of the conditional intensity function in Eq. 1 is equivalent to a branching process, as discussed in the previous section. Suppose that at some time point

$t$  there have been  $n_t$  previous earthquakes. Then, the process intensity at  $t$  can be viewed as a linear superposition of the immigrant process with intensity  $\mu$  and the  $n_t$  processes associated with each previous event, each contributing an intensity of  $g(t - t_i)$ . It can hence be seen this formulation is equivalent to assuming that the immigrant events follow a homogeneous Poisson process with intensity  $\mu$ , and hence have exponentially distributed inter-event times.

The standard ETAS model can be generalised to include a space component, giving the spatiotemporal ETAS model (Ogata 1998). For simplicity and ease of both simulation and computation, we only consider the original temporal ETAS model in this paper rather than its spatiotemporal extension, although our model could be extended to the spatial version without difficulty.

## 3 SR-ETAS models

The theory of stress release (SR) provides a possible solution to the largely discussed concept of “crustal strain budget” by addressing the seismic elasticity as introduced by Reid in his elasticity rebound theory (Reid 1910). In it, earthquake inter-arrival times are described as a ratio of tectonic strain accumulation and strain release, without any statistical association of factors such as size, time, duration and space of the seismicity. We will use a stress release distribution for modelling immigration mainshock events rather than the typically used Exponential distribution implied by the homogeneous Poisson process assumption of the standard ETAS model. In other words, SR-ETAS models provide an alternative for modelling the structure of the immigrant arrival process. Specifically, we will assume that the intensity is time-varying and hence specify a time-dependent  $\mu(t)$ , leading to the following specification of the conditional intensity:

$$\lambda(t|\mathcal{H}_t) = \mu(t) + \sum_{t_i < t} g(t - t_i)\kappa(m_i)$$

While such time-varying specifications of  $\mu(t)$  have been considered before in the literature (Johnson et al. 2005; Imoto 2001; Varini and Rotondi 2015), they typically try to capture structural changes in the long-term earthquake rate, for example modelling  $\mu(t)$  as a piece-wise constant function. Instead, following stress release concepts, we assume that the probability of a mainshock earthquake occurring at time  $t$  depends on the time at which the last mainshock occurred. To make this clearer, we introduce the following notation. For each earthquake  $i$ , let  $B_i$  denote the index of its parent earthquake in the branching structure, with  $B_i = 0$  if it has no parent (i.e. if earthquake  $i$  is an immigrant). We hence have the branching vector  $B = (B_1, \dots, B_n)$ . For example, in Fig. 1,  $B = (0, 1, 1, 3, 1, 0, 6, 7, 6, 0, 8)$ .

Using this notation, at each time  $t$  we write the occurrence time of the last previous immigrant event prior to  $t_i$  is  $t_{I_{[i]}}$  where  $I_{[i]} = \max_j \{j | t_j < t_i \text{ and } B_j = 0\}$ . Similarly, the amount of time which has elapsed since the last previous immigrant event—known as the waiting time—is given by:

$$w_{t_i} = t_i - t_{I_{[i]}}$$

Based on the usual point process theory,  $\mu(t)$  can then be defined as the hazard function:

$$\mu(t) = \mu(t|w_t) = \frac{f_w(w_t)}{1 - F_w(w_t)} \quad (2)$$

where  $F_w(w_t)$  is the waiting time distribution and  $f_w(w_t)$  is its corresponding density. The above expression can be simplified for some choices of distribution although for more complex ones such as the BPT it has to be numerically evaluated since no explicit form is present. As the CDF goes closer to 1, the expression becomes unstable due to numerical underflow caused by the numerator being effectively 0, which cannot be avoided by transforming into the space of logarithms. Since the proposed EM and MLE algorithms depend on estimating this quantity for every time lag, they will not work for general seismological-based stress release distributions (Wheatley et al. 2016; Chen and Stindl 2018). However, we will show that our Bayesian updates do not need a full exploration of all possible inheritance structures, thus the waiting times that are taken into account are much smaller. As such, there is no numerical instability for any reasonable parametrisation of the uncaused events' distribution and we can evaluate numerically the above function as part of our MCMC sampler.

Under the definition provided by Eq. 2, the probability of an immigrant event occurring depends on the time which has elapsed since the previous immigrant event, in a manner which is consistent with SR theory since it can be interpreted with respect to Reid rebound theory where the ground state level is reached only for immigrant events and all other events are causing smaller impact on the strain accumulation/reduction. Since the branching structure is used to determine the time of the last immigrant event, we will refer to this model as the B-SR-ETAS model (Branched-SR-ETAS).

However, in practice when working with real earthquake catalogues, we do not know which events in the sequence are mainshocks since we do not have access to the true branching structure. Indeed, the branching structure is usually estimated as a by-product of the standard algorithms used to estimate the ETAS model (Ross 2018a; Rasmussen 2013; Veen and Schoenberg 2008). However, we cannot use this idea directly since we are caught in a vicious circle: our parameter estimation requires access to the branching structure in order

to define the mainshock earthquakes, but we cannot get the branching structure without first estimating the model parameters! One approach is to marginalise the branching structures out of the joint distribution by summing over all  $2^{n-1}$  unique branching structures, for a catalogue with length  $n$ . However, this is computationally intractable for even a moderate value of  $n$ . As such, we will instead introduce a Monte Carlo approach for performing this inference in a computationally tractable way.

Since defining waiting times based on the previous immigrant hence leads to computationally difficult parameter inference, we could instead define the waiting time  $w_t$  based on the occurrence time of the last earthquake prior to  $t$ , regardless of whether it was an immigrant or an offspring. At time  $t$ , the time of the last event is given by  $t_E$  where  $E = \max\{i | t_i < t\}$ . The waiting time in this case is hence:

$$w_t = t - t_E$$

with  $\mu(t)$  defined according to Eq. 2 as before. Under an SR interpretation, this implies that the strain accumulated with respect to the immigrant events causation can be assumed to reduce to a ground level after every earthquake in the sequence, which corresponds to Reid's elasticity rebound theory in which an event occurs when a specific intensity threshold is reached (Reid 1910). We denote this model by F-SR-ETAS (Full SR-ETAS).

The previously introduced concept of parameter set  $\theta$  can be adapted for both SR-ETAS models as  $\theta = \{\theta_{SR}, \alpha, c, p, k\}$ , where  $\theta_{SR}$  is taking the parameters of the waiting time distribution  $F_w(\cdot)$ .

## 4 Waiting time distributions

Regardless of which of the two approaches (B-SR-ETAS or F-SR-ETAS) we take when defining the waiting times  $w_t$ , we must specify a probability model  $F_w$  which governs their distribution. In standard ETAS, the Poisson assumption results in a memoryless Exponential distribution. In contrast, the SR approach implies other forms of distributions with nonconstant hazard rate. There is some controversy in the seismological literature over the appropriate waiting time distribution for modelling the time between mainshocks. As such, we will consider two different distributions which have been found to have strong empirical support: the Brownian Passage Time, and the Gamma.

### 4.1 Brownian passage times (BPT) immigration

The Brownian Passage Times concept was introduced to describe the inter-arrival times of earthquake events by Ellsworth et al. (1999) and Matthews et al. (2002). It is a

probabilistic physically based approach for addressing event recurrence based on long-term, load-state process assuming behaviour similar to those of the Brownian relaxation oscillator (BRO). In it, earthquakes are assumed to be an energy release of a tectonic system that accumulates strain. This approximates the event inter-arrival time probability density function as:

$$f(w_t; \lambda, \nu) = \left[ \frac{\lambda}{2\pi \nu^2 w_t^3} \right]^{\frac{1}{2}} e^{-\frac{(w_t - \lambda)^2}{2\lambda \nu^2 w_t}}.$$

The cumulative distribution function of the BPT has a close form which is the following:

$$F(w_t) = P(T \leq w_t) = \int_0^{w_t} f(u) du = \Phi[u_1(w_t)] + e^{\frac{2}{\alpha^2}} \Phi[-u_2(w_t)],$$

where

$$\Phi(w_t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{w_t} e^{-\frac{u^2}{2}} du, \tag{3}$$

for

$$u_1(w_t) = \nu^{-1} \left[ w_t^{1/2} \lambda^{-1/2} - w_t^{-1/2} \lambda^{1/2} \right],$$

$$u_2(w_t) = \nu^{-1} \left[ w_t^{1/2} \lambda^{-1/2} + w_t^{-1/2} \lambda^{1/2} \right].$$

The main attributes of BPT compared to other SR distributional alternatives are the following:

1. The mean waiting time of the (immigrant) events in the catalogue of interest,  $\lambda$ , provides a threshold until which the probability of event occurrence is continuously increasing. After reaching the mean waiting time, the conditional probability of occurrence is time independent and depends only on the aperiodicity parameter,  $\nu$ , which is associated with the scaling of the Brownian motion.
2. Earthquake occurrence corresponds to immediate stress release to ground base level. Thus, the probability of immediate events recurrence is zero.

From here after, the BPT-based SR-ETAS models will be referred as F-B-ETAS for the Full Stress Release Brownian Passage Time Epidemic After Shock Sequence model and B-B-ETAS for the branched one.

### 4.2 Gamma process immigration

The Gamma distribution is an exponential family distribution with two parameters, namely shape parameter  $a > 0$  and

scale parameter  $s > 0$ . It has been found by Kagan and Knopoff (1984), Chen et al. (2013), Wang et al. (2012) to provide a good model for main shock inter-arrival times. The Gamma distribution probability density function is:

$$f(w_t) = \frac{1}{s^a \Gamma(a)} w_t^{a-1} e^{-w_t/s}$$

where  $\Gamma(a) = \int_0^\infty u^{a-1} e^{-u} du$  is the Gamma function. The corresponding cumulative distribution function is:

$$F(w_t) = \frac{1}{s^a \Gamma(a)} \int_0^{w_t} u^{a-1} e^{-u/s} du.$$

From here after, the Gamma-based SR-ETAS models will be referred to as F-G-ETAS for the Full-SR Gamma ETAS model and B-G-ETAS for the branched one.

## 5 Estimation

We now consider parameter estimation for the SR-ETAS models. This includes estimating the ETAS model parameters  $\theta_\Phi = (\alpha, c, p, k)$ , as well as  $\theta_{SR}$ , the parameters of the waiting time distribution  $F_w$ . Let  $\theta = (\theta_{SR}, \theta_\Phi)$  denote the full set of unknown parameters. We perform Bayesian inference for the model parameters by developing a latent variable MCMC scheme that allows sampling from the full posterior.

### 5.1 Likelihood function

The general likelihood function of an ETAS-based model is (Ross 2018a):

$$p(\mathcal{H}_t | \theta) = \prod_{i=1}^n \lambda(t_i | \mathcal{H}_t) e^{-\int_0^\infty \lambda(z | \mathcal{H}_t) dz} \tag{4}$$

with corresponding log-likelihood:

$$\log(p(\mathcal{H}_t | \theta, Z)) = \sum_{i=1}^n \log(\lambda(t_i | \mathcal{H}_t)) - \int_0^{t_n} \mu(s) ds - \sum_{i=1}^{n-1} \kappa(m_i) \int_0^{t_n - t_i} g(s) ds \tag{5}$$

Plugging in the specific parametric choices for the kernel functions  $\kappa(\cdot)$  and  $g(\cdot)$  in the ETAS model gives:

$$\log(p(\mathcal{H}_t|\theta, Z)) = \sum_{i=1}^n \log \left[ \mu(t) + \sum_{j=1}^{i-1} \frac{ke^{\alpha(m_j-M_0)}}{(t_i - t_j + c)^p} \right] - \int_0^{t_n} \mu(s)ds - \sum_{i=1}^n ke^{\alpha(m_i-M_0)} \left( 1 - \frac{c^{p-1}}{(t_n - t_i + c)^{p-1}} \right) \tag{6}$$

where  $\theta$  is the set of all parameters in the model,  $Z_{1:n} \in \{0, 1\}^n$  is a vector with length  $n$  indicating whether each event is immigrant (1) or not (0). As of the branching structure introduced in Fig. 1, the immigrant information is  $Z = \{1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0\}$ .

Note that  $\mu(\cdot)$  depends on the branching vector  $Z$  for B-SR-ETAS since the intensity of the background process depends on the time at which the last immigrant event occurred. However since the true branching structure is not known in practice, it must be marginalised out by summing over all  $2^{n-1}$  possible values. Therefore, the log-likelihood of the B-SR-ETAS model is:

$$p(\mathcal{H}_t|\theta) = \sum_{j=1}^{2^{n-1}} p(\mathcal{H}_t|\theta, Z = z_j)p(Z = z_j|\theta).$$

In practice, this summation is likely to be intractable for even moderate  $n$ . As such, we will instead use a latent variable formulation where the unknown branching vector  $Z$  is treated as a parameter to be learned. In order to evaluate this quantity, we can either use a single “best” quantity or to provide a Monte Carlo approximation of it based on sampling multiple branching structures based on the true/optimised parameters  $\theta$ .

While the proposed by Wheatley et al. (2016) log-likelihood function is conceptually the same as the one shown above, (Chen and Stindl 2018) Sect. 3, Remark 1, claims that the log-likelihood form is wrong with respect to the examined by them RHawkes process. The full algorithm that is proposed for the calculation of the (log-)likelihood of RHawkes by Chen and Stindl (2018) is provided in “Appendix A”. This method requires the calculation of probabilities associated with all possible inheritance structures. In other words, the immigrant intensity Eq. 2 has to be evaluated for all possible temporal lags when in the calculation of Eqs. 6–8, Sect. 3 of Chen and Stindl (2018). As discussed before, such expression cannot be evaluated for immigrant distributions that do not have explicit intensity function (Eq. 2). However, from a Bayesian perspective, the branching structure is a feature that we learn. Rather than being an unknown quantity, it is a data characteristics that we evaluate based on our inheritance believes. Thus, the provided log-likelihood function in Eq. 5 is feasible for the scope of a Bayesian algorithm.

### 5.2 Bayesian analysis

We will use Monte Carlo Markov Chain techniques for obtaining samples from the posterior distribution of the ETAS parameters  $\theta$ . The main aim of Bayesian analysis is to fully explore the parameter’s distribution rather than trying to obtain a single best value. A prior distribution  $\pi(\theta)$  is set based on our prior knowledge of the parameters’ distribution. Then, using the Bayes theorem, the posterior distribution of the parameter  $\theta$  can be represented as follows:

$$\pi(\theta|\mathcal{H}_t) = \frac{p(\mathcal{H}_t|\theta)\pi(\theta)}{\int_{\Theta} p(\mathcal{H}_t|\theta)\pi(\theta)d\theta} \tag{7}$$

The multi-dimensional integral in Eq. 7 is extremely hard to be rearranged in order to obtain a closed form. For this reason, we use the well-known Metropolis–Hastings (MH) algorithm for sampling the Markov Chain of interest (Chib and Greenberg 1995; Hamra et al. 2013; Rotondi and Varini 2007). According to this method, we firstly initialise the parameter set with some reasonable values  $\theta^{(0)}$ . At step  $i$  we would like to propose a new sampled value of  $\theta^{(i)}$  based on  $\theta^{(i-1)}$ . For example we might consider a White Noise transformation such as  $\theta^{(i)} = \theta^{(i-1)} + \epsilon$  were  $\epsilon \sim N(0, \sigma^2)$ . The acceptance probability of the proposed value  $\theta^{(i)}$  is  $\pi(\theta^{(i)}|\mathcal{H}_t)/\pi(\theta^{(i-1)}|\mathcal{H}_t)$ . If the value is rejected, we fail to obtain a new sample at this step and assign the  $(i - 1)$ st sample to the  $i$ th (i.e.  $\theta^{(i)} = \theta^{(i-1)}$ ) and repeat the procedure for the next step.

As shown by Ross (2018b), the efficiency of MCMC algorithms for ETAS type models can be drastically improved by including the branching structure as a latent variable. What is more, the full conditional MCMC MH techniques are only applicable for the ETAS and F-SR-ETAS, while the B-SR-ETAS is only feasible to be implemented based on the latent variable approach since the branching structure is needed to obtain the last immigrant event time used in the evaluation of  $\mu(\cdot)$ . To introduce this latent variable approach, we first make some small reparameterisations of the(SR-)ETAS intensity function:

$$\begin{aligned} \Phi(t|\mathcal{H}_t) &= \sum_{t>t_i} g(t - t_i)\kappa(m_i) \\ &= \sum_{t>t_i} \frac{k}{(t - t_i + c)^p} e^{\alpha(m_i-M_0)} \\ &= \sum_{t>t_i} \frac{(p - 1)c^{p-1}}{(t - t_i + c)^p} K e^{\alpha(m_i-M_0)} \\ &= \sum_{t>t_i} h(t - t_i)t(m_i) \end{aligned}$$

where  $\iota(m_i) = Ke^{\alpha(m_i - M_0)}$ ,  $K = \frac{k}{(p-1)c^{p-1}}$ , and  $h(z) = (p-1)c^{p-1} \frac{1}{(z+c)^p}$  is a reparameterisation of  $g(\cdot)$  that now integrates to 1. The log-likelihood function as of Eq. 6 is then:

$$\begin{aligned} &\log(p(\mathcal{H}_t|\theta, Z)) \\ &= \sum_{i=1}^n \log \left[ \mu(t) + \sum_{j=1}^{i-1} \frac{K(p-1)c^{p-1}e^{\alpha(m_j - M_0)}}{(t_i - t_j + c)^p} \right] \\ &\quad - \int_0^{t_n} \mu(s)ds - \sum_{i=1}^n Ke^{\alpha(m_i - M_0)} \\ &\quad \left( 1 - \frac{c^{p-1}}{(t_n - t_i + c)^{p-1}} \right) \end{aligned} \tag{8}$$

Performing MCMC directly is difficult due to the high correlation of all the parameters in this likelihood. The goal of including the latent branching variables is to break this dependence and decouple the parameters. For this to be done, we need to be able to sample from the posterior distribution of the full branching structure, which requires extending the approach used in Ross (2018b).

### 5.2.1 Branching procedure

Let  $B$  denote the branching structure vector where  $B_i = j$  indicates that the  $i$ -th event in the sequence is caused by the  $j$ -th event ( $j < i$ ). Immigrant events are notated as uncaused, i.e. caused by an event with index 0. If we refer again to the branching structure, that was introduced on Fig. 1, we can visually assign corresponding values for our branching inheritance measure  $B_i$  as follows  $B = \{0, 1, 1, 3, 1, 0, 6, 7, 6, 0, 8\}$ . The immigrant events are coming from a in-homogeneous Poisson process with intensity function  $\mu(\cdot)$  while the offspring events of the  $j$ -th event are generated from in-homogeneous Poisson process with intensity  $h(t_i - t_j)\iota(m_j)$ . Assuming that each event in the sequence is generated by a single process, we can assign probabilities distribution to each event with respect to its branching pedigree and therefore sample a branching structure from its conditional posterior as follows:

1. Initiate the branching by setting  $B_1 = 0$  as we assume that always the first term is immigrant.
2. Sample each  $B_i$  in turn from  $P(B_i|\mathcal{H}_t, \theta, B_{1:(i-1)})$
3. Return the sequence of generated  $B_i$ s

As pointed out by an anonymous reviewer, the form of  $P(B_i|\mathcal{H}_t, \theta, B_{1:(i-1)})$  in the general SR-ETAS model is substantially more complex than for the standard ETAS, since using a general renewal process for the mainshocks introduces substantial dependence in the process. It was pre-

viously shown in Ross (2018a) that for a standard ETAS model, the conditional posterior for each  $B_i$  is independent of all other  $B_j$  for  $i \neq j$  and can be written as:

$$\begin{aligned} -P(B_i = 0|\mathcal{H}_t, \theta, B_{1:(i-1)}) &= \frac{\mu(t_i - t_{I_{[i]}})}{\mu(t_i - t_{I_{[i]}}) + \Phi(t_i|\mathcal{H}_t)} \\ -P(B_i = j|\mathcal{H}_t, \theta, B_{1:(i-1)}) &= \frac{h(t_i - t_j)\iota(m_j)}{\mu(t_i - t_{I_{[i]}}) + \Phi(t_i|\mathcal{H}_t)} \end{aligned}$$

for  $j$  in 1 to  $i - 1$

where  $I_{[i]}$  takes the last immigrant event index before the  $i$ -th event in order to obtain branching for B-SR-ETAS, and  $I_{[i]} = i - 1$  for F-SR-ETAS and ETAS models.

However, for our more general SR-ETAS models with renewal process immigration, this independence no longer holds. Instead, as we show in ‘‘Appendix B’’, the conditional posteriors are:

$$\begin{aligned} -P(B_i = 0|\mathcal{H}_t, \theta, B) &\propto \mu(t_i - t_{I_{[i]}}) \\ &\times e^{-\int_{t_i}^{t_{I_{[i]}^*}} \mu(t|B_i=0)dt} \mu(t_{I_{[i]}^*} - t_i) \end{aligned} \tag{9}$$

$$\begin{aligned} -P(B_i = j|\mathcal{H}_t, \theta, B) &\propto \Phi(t_i|\mathcal{H}_t) \\ &\times e^{-\int_{t_i}^{t_{I_{[i]}^*}} \mu(t|B_i \neq 0)dt} \mu(t_{I_{[i]}^*} - t_{I_{[i]}}) \end{aligned}$$

for  $j$  in 1 to  $i - 1$  (10)

where at each time  $t_i$  we write the occurrence time of the first immigrant event after  $t_i$  as  $t_{I_{[i]}^*}$  where  $I_{[i]}^* = \min_j \{j|t_j > t_i \text{ and } B_j = 0\}$ .

### 5.2.2 Log-likelihood latent variable transformations

The (log-)likelihood function of the process can be rewritten conditional on the branching structure. From Eq. 1 the process intensity at time  $t$  is a sum of the contribution of  $\mu(t)$  from the background process, and a contribution of  $h(t - t_i)$  for each of the previous event  $t_i$ . Let us define  $S_0$  to be the set of all immigrant events (conditional on the branching structure), and  $S_l$  to be the set of all events triggered by each event  $t_i$ . We write  $|S_l|$  to denote the number of events in each set. For a given branching structure, the likelihood function can then be rewritten as:

$$\begin{aligned} p(\mathcal{H}_t|\theta, B) &= e^{-\sum_{t_i \in S_0} \int_0^{t_i - t_i - 1} \mu(u)du} \prod_{s \in S_0} \mu(s) \\ &\times \prod_{j=1}^n \left( e^{-\iota(m_j) \int_0^{t_n - t_i - 1} h(u)du} \iota(m_j)^{|S_j|} \right. \\ &\quad \left. \prod_{t_i \in S_j} h(t_i - t_j) \right). \end{aligned} \tag{11}$$

In this notation  $B$  is a full branching structure realisation, and the integrals are summed over all immigrant events except the first one since there is no waiting time for the first event. The permutation over  $\mu(s)$  is a permutation of the spot values of  $\mu(\cdot)$  at the triggering times of all immigrant events in the catalogue and  $\theta_{SR}$  represents the parameter set of the chosen SR distribution. Note that  $\mu(t)$  in this case is actually  $\mu(t|w_t) = \frac{f_w(w_t)}{1-F_w(w_t)}$ , where  $w_t$  is the waiting time from the last immigration for the B-SR-ETAS model and the waiting time between every event for the F-SR-ETAS. The  $f_w(\cdot)$  and  $F_w(\cdot)$  are the corresponding PDF and CDF of the candidate immigration distribution (SR). Additional approximation can be obtained based on the previously mentioned infinite time assumption, namely that the end time of the catalogue is very large and as such the integral over the Modified Omori law ( $h(\cdot)$ ) for the range of values in the catalogue converges to 1, or in other words

$$\lim_{t_n \rightarrow \infty} \int_0^{t_n - t_{i-1}} h(u) du = 1. \tag{12}$$

Based on the branching structure, the likelihood function (and hence the posterior, given independent priors) in Eq. 11 is separable into three functions that can update separately the following parameters' sets  $\theta_{SR}$ ,  $\{K, \alpha\}$  and  $\{c, p\}$ . They have the following posterior probabilities that will be used for the Metropolis–Hastings accept–reject ratios:

$$\log(\pi(\theta_{SR} | \mathcal{H}_t, \theta, B)) \propto \log(\pi(\theta_{SR})) + \sum_{t_i \in S_0} \left( \log(\mu(t_i)) - \int_0^{t_i - t_{i-1}} \mu(u) du \right) \tag{13}$$

$$\log(\pi(K, \alpha | \mathcal{H}_t, \theta, B)) \propto \log(\pi(K, \alpha)) - \sum_{j=1}^n \left( \iota(m_j) \left( 1 - \frac{c^{p-1}}{(t_n - t_j + c)^{p-1}} \right) - |S_j| \log(\iota(m_j)) \right) \tag{14}$$

$$\log(\pi(c, p | \mathcal{H}_t, \theta, B)) \propto \log(\pi(c, p)) - \sum_{j=1}^n \left( \iota(m_j) \left( 1 - \frac{c^{p-1}}{(t_n - t_j + c)^{p-1}} \right) - \sum_{t_i \in S_j} \log(h(t_i - t_j)) \right) \tag{15}$$

Note that based on the infinite time assumption, as defined in Eq. 12, the term  $\frac{c^{p-1}}{(t_n - t_j + c)^{p-1}}$  is effectively zero and as such the above expressions can be simplified so that every posterior probability to be independent from the other parameters' behaviour. In other words, based on the infinite time assumption are achieved three independently updated chains that have interaction only when a new full branching structure is sampled. The conducted analysis was carried out without the infinite time assumption for all datasets because it appeared

that there are large differences for the North California catalogue.

### 5.2.3 Choice of prior and proposal distributions

The (SR-ETAS) parameter estimates in this paper were obtained by running a latent variable MCMC for each of the 5 proposed models—ETAS, B-B-ETAS, F-B-ETAS, B-G-ETAS and F-G-ETAS. We use noninformative priors for all (SR-) ETAS parameters. For the standard ETAS model there exists a conjugate Gamma prior for the fixed ground intensity  $\mu$  (Ross 2018a). We used a flat Uniform prior for  $\theta_{SR}$ ,  $\alpha$ ,  $\log(c)$ ,  $\log(p)$  and  $\log(K)$  with bounds  $\alpha \in [0, 10]$ ,  $c \in [0, 10]$ ,  $p \in [1, 30]$ ,  $K \in [0, \infty]$ , although more informative priors could be used if desired. For a infinite time catalogue, the overall productivity of the offspring decay, i.e. the mean number of offsprings by every event is  $K$  thus we might want to reduce it to be smaller than 1 for simulation purposes. Since in reality the time is not infinite, the overall productivity is not  $K$  anymore. It is catalogue dependent and as such we decided to use a higher upper bound for  $K$ . The support for the other parameters is greatly influenced by the potential multimodality and were taken to be identical to those used in the Bayesian ETAS R package (Ross 2018a).

We use as a proposal distribution a Normal with standard deviation of 0.1 for all parameters that require Metropolis–Hastings updates. The New Madrid catalogue parameters' sequences are with overall length of 15,000 after burn-in of 5, 100 and 100 for the  $\theta_{SR}$ ,  $\{K, \alpha\}$  and  $\{c, p\}$ , respectively. The branching structure was sampled from its conditional posterior at every iteration. The North California catalogue is much larger, thus we updated the branching structure less frequently at every 20 iterations of the Gibbs sampler, overall 12,000 parameter sets were obtained after burn-in of 4, 100, 20 for the  $\theta_{SR}$ ,  $\{K, \alpha\}$  and  $\{c, p\}$ , respectively.

## 6 Model comparison: diagnostic tests

In order to compare the performance of the two SR-ETAS models to the standard ETAS model for the purpose of modelling real earthquake data, we require model comparison metrics. In this section, we discuss the various tests which we will use for the comparison.

### 6.1 Bayesian information criterion (BIC)

The Bayesian Information Criterion (BIC) is a popular penalised likelihood technique which incorporates a penalty based on the number of parameters in order to reduce the risk of overfitting (Schwarz 1978). In a Bayesian framework, it functions as an asymptotic approximation to the model marginal likelihood in situations where the marginal likeli-



hood is intractable to compute. Given a model with parameter vector  $\theta$ , the model’s BIC is defined as:

$$\text{BIC}(\theta) = -l(\theta) + \frac{d}{2} \cdot \log(n),$$

where  $d$  is the number of free model parameters, i.e.  $d = |\theta|$ ,  $l(\theta)$  is the log-likelihood value evaluated at the MLE  $\hat{\theta}$  and  $n$  is the number of observations. The best model is associated with the lowest value of BIC coefficient.

### 6.1.1 Deviance information criterion (DIC)

The DIC is a fully Bayesian alternative to the AIC. It replaces the maximum likelihood parameter estimates of  $\theta$  with their posterior mean  $\bar{\theta}$ . The number of parameters correction is replaced with a measure of parameter adequacy based on the goodness of sample of  $\theta$  in terms of log-likelihood (Gelman et al. 2014). For every set of model parameters  $\theta$  the model’s DIC value is:

$$\text{DIC}(\theta) = -2l(\bar{\theta}) + 2 \cdot p_{\text{DIC}},$$

where  $l(\cdot)$  is the log-likelihood function and  $p_{\text{DIC}}$  is the effective number of parameters. It is defined as:

$$p_{\text{DIC}} = 2l(\bar{\theta}) - 2\mathbb{E}(l(\text{data}|\theta)) \cong 2l(\bar{\theta}) - 2\frac{1}{S} \sum_{s=1}^S l(\theta_s),$$

where  $\theta_s$  indicates the specific obtained value of the parameter(s) in the chain. Alternatively, we can compute the effective sample size as the variance of the obtained log-likelihood values for all sampled parameters as follows:

$$p_{\text{DICalt}} = 2\text{var}(l(\text{data}|\theta))$$

This method is not as numerically stable as the other one but it is easier to compute and it is guaranteed to provide positive values. There are many different alternatives of DIC. There are a number of limitations of all alternatives of DIC that the reader can consider while interpreting the results (Spiegelhalter et al. 2014).

## 6.2 Time rescaling residuals

The time rescaling concept aims to rescale the observations from a point process based on its conditional intensity function, in order to produce residuals which follow a homogeneous Poisson process (Brown et al. 2002; Lallouache and Challet 2016). For a given temporal point process sequence  $0 \leq t_1 < \dots < t_n \leq T$  with corresponding conditional

intensity  $\lambda(t|\mathcal{H}_t^c) > 0$  for  $t \in (0, T]$ , the residuals are defined as:

$$\Lambda(t_k) = \int_0^{t_k} \lambda(u|\theta, \mathcal{H}_u^c) du.$$

Assuming that the cumulative intensity is finite, i.e.  $\lambda(\cdot) < \infty$ , then all  $\Lambda(\cdot)$ s are a realisation from a Poisson process with rate 1. The inter-arrival times  $\delta_k = \Lambda(t_k) - \Lambda(t_{k-1})$  are hence independent Exponentially distributed with mean and standard deviation of 1. As such, testing these residuals to check they follow an Exponential distribution is equivalent to testing whether the conditional intensity function has the form specified (SR-ETAS vs ETAS, in our formulation). This test can be carried out using a goodness-of-fit test such as The Cramér-Von Mises (CVM) or Engle Russell Excess Dispersion test.

### 6.2.1 Cramér-Von Mises test

The Cramér-Von Mises (CVM) test (Stephens 1970) compares a set of observations to a hypothesised distribution function by computing the average distance between the empirical and hypothesised distributions. In other words, for an ordered data  $X = (x_1, \dots, x_n)$  we are interested to examine, whether the sample cumulative distribution function (CDF)  $F(\cdot)$  is close to a specific CDF  $F_0(\cdot)$ . The CVM test statistics is:

$$\text{CVM} = n \int_{-\infty}^{\infty} (F(x) - F_0(x))^2 w(x) dF_0(x)$$

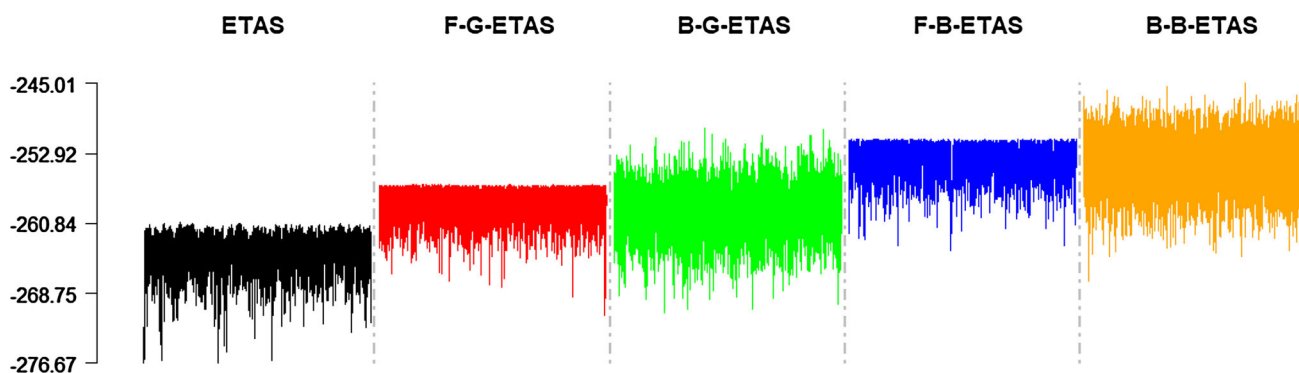
where  $w(x)$  is a weight function which is assumed to be equal to 1 in the standard CVM test.

### 6.2.2 Ljung–Box test

The Ljung–Box (LB) test was introduced by Ljung and Box (1978) in order to check whether a sequence is an independent realisation from a target distribution of interest. This is done by examining  $m$  autocorrelations of the residuals. The test statistic is:

$$Q = n(n+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{n-k}$$

where  $n$  is the length of the data,  $\hat{\rho}_k$  is the estimated autocorrelation at the lag of interest  $k$  with respect to the number of lags ( $m$ ) that are taken into account. The choice of appropriate number of lags  $m$  is critical for obtaining adequate test results (Hyndman 2014). Given that the data is expected to be nonseasonal and always have more than 200 observations, we always use 10 lags, i.e.  $m = 10$ .



**Fig. 2** Log-likelihood of the MCMC sequences based on the used full branching structures for the New Madrid catalogue with respect to ETAS/F-G-ETAS/B-G-ETAS/F-B-ETAS/B-B-ETAS

### 6.2.3 Engle Russell excess dispersion test

Excessive dispersion of the exponentially distributed residuals can be examined using the Engle Russell Excess Dispersion (ER) test (Lallouache and Challet 2016; Engle and Russell 1998). It takes into account only the sample variance  $\hat{\sigma}^2$  and has the following test statistics:

$$ER = \sqrt{n} \frac{\hat{\sigma}^2 - 1}{\sqrt{8}}.$$

Under the null hypothesis of lack of excess dispersion,  $\hat{\sigma}$  is distributed as a standard Normal random variable.

## 7 Applications

In this section, we discuss and compare the model fit across ETAS-based models on two seismic catalogue of interest. The first one is the New Madrid catalogue which is much smaller but of great interest for underwriting community while the second one, the North California, is more dense and should behave similarly to a typical single fault catalogue.

### 7.1 New Madrid seismic sequence

We first compare the performance of the ETAS and SR-ETAS models on the catalogue of New Madrid earthquakes obtained from The University of Memphis website <http://www.memphis.edu/cei/seismic/catalog.php>. This catalogue starts on 29/06/1974 and ends on 23/02/2017. Only earthquakes of magnitude greater than 3 are considered since smaller ones are typically considered harmless. The resulting catalogue contains 308 events. We fit the ETAS model and BPT and Gamma-based SR-ETAS models to this catalogue.

Figure 2 shows how the sequence of log-likelihoods for each model evolves over each iteration of the Gibbs sampler (after convergence). It is clearly observable that there is a difference between the overall fitting capabilities between the 5 models. What is more, the overall mixing for branched-SR models is greater and relatively more symmetric. Figure 3 plots the posterior distribution of the model parameters for the B-B-ETAS, which is the most difficult model to estimate due to the need to estimate the unknown branching structure. The posterior distributions for the parameters in the other models are similar. As expected, the obtained parameters' distributions are smooth, symmetric and not very different from a bell-shaped-base form.

The Goodness-of-fit and model comparison results are shown on Table 1. Amongst all ETAS-based models it appears that SR-ETAS models are superior to the standard ETAS model according to both BIC and DIC. BPT-based models are supreme to their corresponding Gamma alternatives and B-SR-ETAS models are supreme to the F-SR-ETAS. The best model within all examined models is evidently the B-B-ETAS.

Figure 4 presents the informal diagnostic plots of the time residuals. On the left are the raw time residuals for all 5 models versus a diagonal line. Ideally, these should overlap. The overall pattern is very similar for all models. They all experience a bias towards the middle of the catalogue which might indicate a potential minor nonstationarity in the data (Kumazawa and Ogata 2013). The right part of Fig. 4 shows a Q–Q plot for the residuals of all 5 models versus Exponential(1) distribution. All 5 models behave similarly, with minor spread from the expected results for large quantiles.

The time rescaling diagnostic tests conclude that the CVM and ER tests were passed by all 5 models at the 5% significance level. The LB test is passed only by F-B-ETAS at 5% significance level, while all other models pass it at 1% significance level. Thus, there might be minor dependence in the residuals which we believe is negligible.

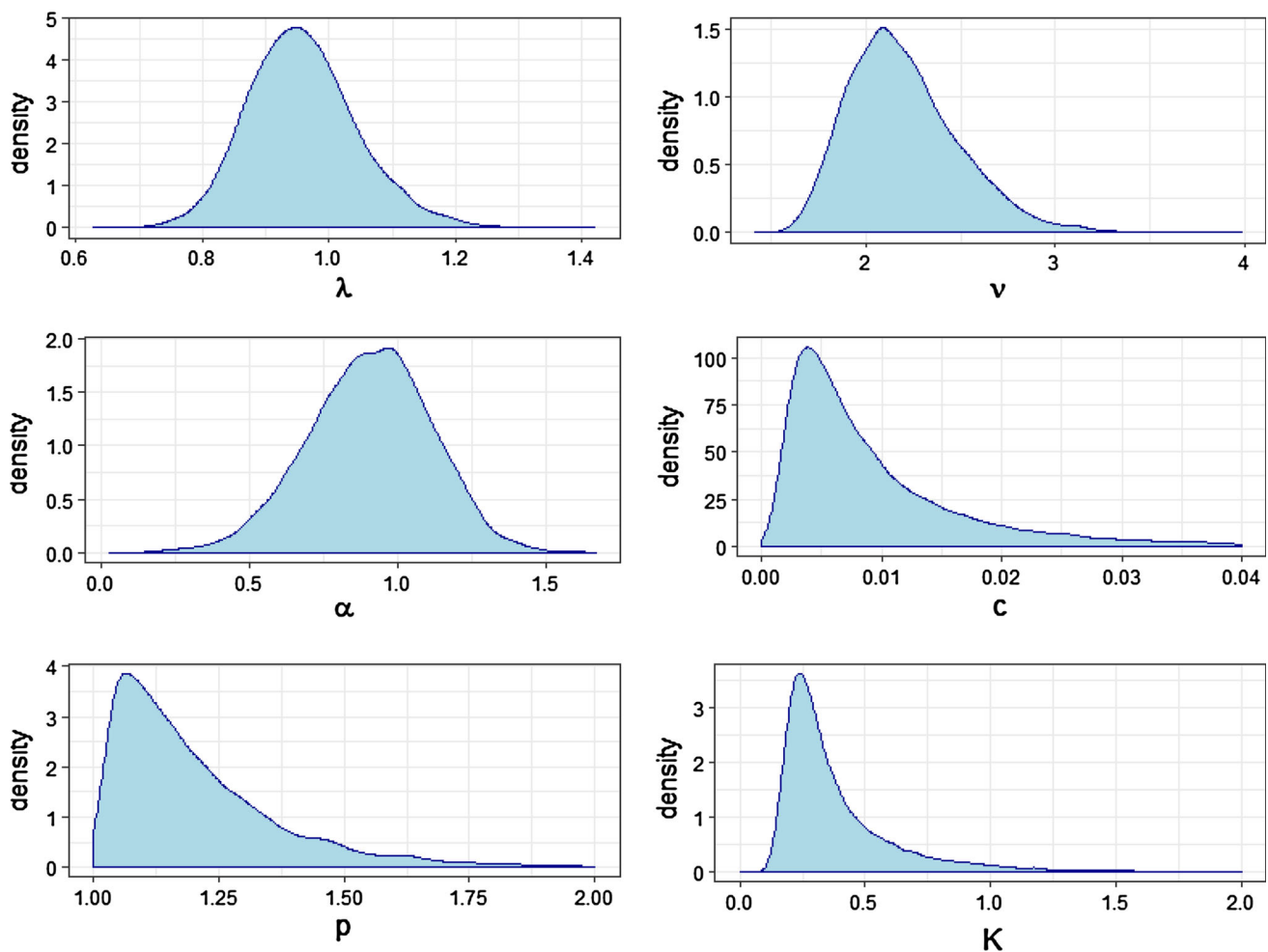


Fig. 3 B-B-ETAS MCMC parameters’ density for the New Madrid catalogue

**Table 1** Goodness-of-fit summary—New Madrid; ETAS, BPT and Gamma-based SR-ETAS

	ETAS	F-G-ETAS	B-G-ETAS	F-B-ETAS	B-B-ETAS
Log-likelihood	− 260.75	− 256.44	− 250.13 <sup>a</sup>	− 251.28	− 245.01 <sup>a</sup>
Number of parameters	5	6	6	6	6
BIC	275.05	273.23	− 267.32 <sup>a</sup>	268.47	− 262.20 <sup>a</sup>
DIC	519.82	515.47	484.28 <sup>a</sup>	507.60	483.80 <sup>a</sup>

Lower values of the BIC/DIC indicate superior fit

<sup>a</sup>The value is approximate

### 7.2 North California seismic sequence

The previous analysis was repeated using a North California seismic sequence. The historical catalogue of earthquake events can be obtained from <http://www.ncedc.org/ncedc/catalog-search.html>. We took into account all events from 01/01/1987 until 31/12/2015, with magnitude of completeness of 3.5. This created a catalogue consisting of 3442 events.

The full sequences of the log-likelihood calculated using the Gibbs sampler are shown in Fig. 5. As before, all the SR-

ETAS models appear to give substantial improvements over the basic ETAS model. Again we decided to report the posterior density only for B-B-ETAS which are shown in Fig. 6. The heavy tails that appeared for the New Madrid catalogue are not present. Overall the shapes of all 6 parameters appear to be roughly symmetric. The goodness-of-fit results of the un-simplified (finite time) runs are shown on Table 2.

According to the BIC, the F-B-ETAS is the worst model while all other SR models are slightly better than the standard ETAS. Due to the larger number of observations in this catalogue, we decided to examine the DICalt that depends on the

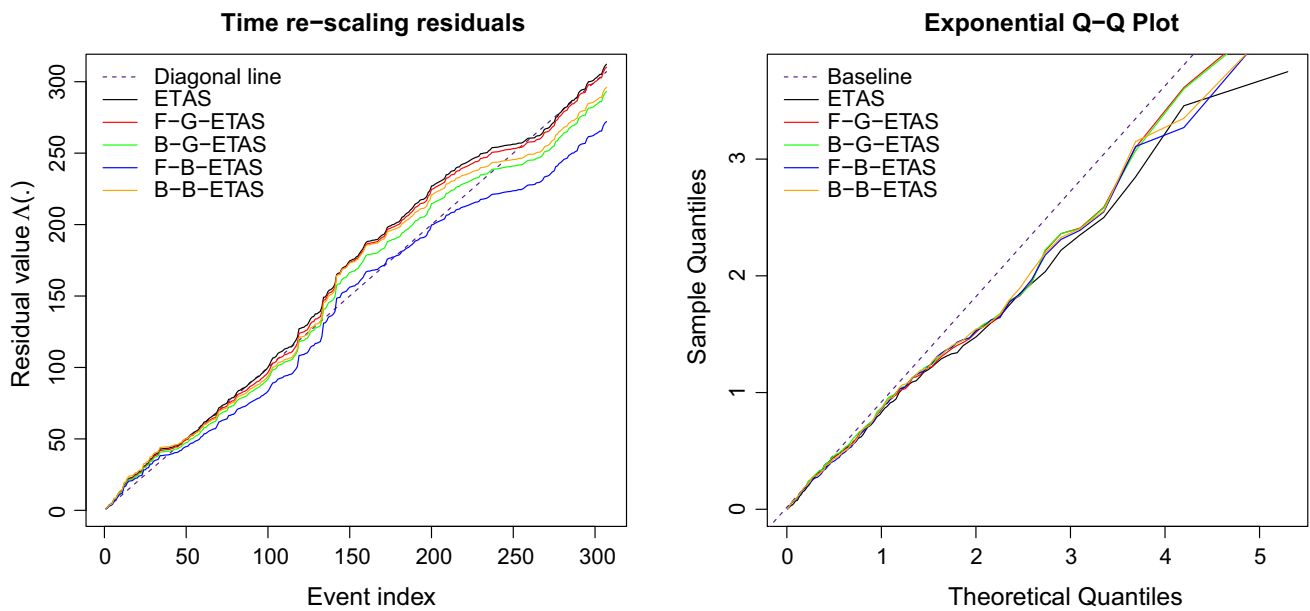


Fig. 4 Time rescaling diagnostic plots for the New Madrid catalogue

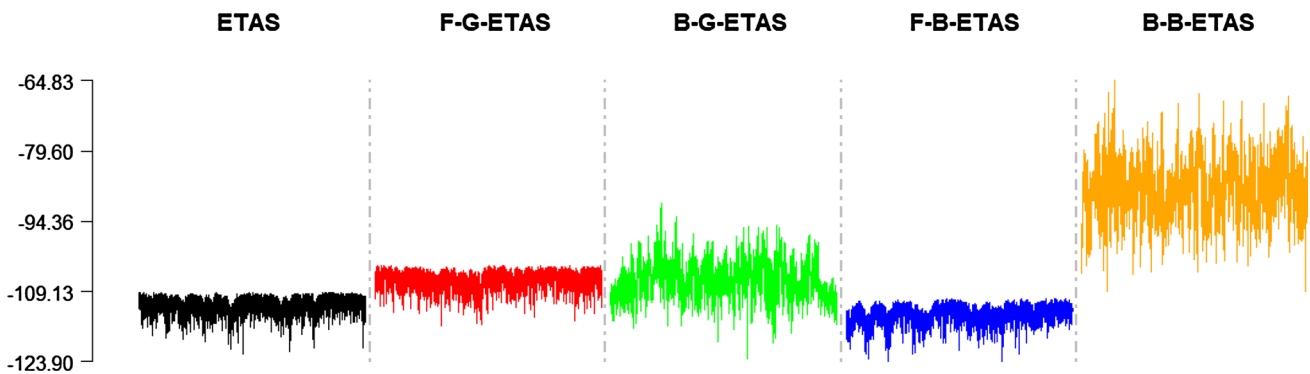


Fig. 5 Log-likelihood of the MCMC sequences based on the used full branching structures for the North California catalogue with respect to ETAS/F-G-ETAS/B-G-ETAS/F-B-ETAS/B-B-ETAS

previously defined  $p_{DIC_{alt}}$ . According to it, the branched-SR models are providing a considerable performance improvement compared to the full models while the standard ETAS performs the worst. For this catalogue BPT-based models are no longer superior to their corresponding Gamma alternatives. Interestingly, F-B-ETAS is currently the worst model. This is probably attributed to the fact that Gamma-SR-ETAS models are guaranteed to be at least as good as the ETAS model since they can reduce to it, since the Exponential inter-arrival time distribution used in the standard ETAS is nested inside the Gamma distribution. It is clear that B-B-ETAS has a great advantage amongst all other models.

Figure 7 presents the time residuals informal diagnostic plots. On the left are shown the residuals and on the right is shown the Q–Q plot for all 5 models. The residuals plot indicates that all SR-ETAS models behave very similarly, while standard ETAS indicates a flow from the desired pattern. Sim-

ilarly to the results for the New Madrid catalogue, all models experience a bias towards the middle of the catalogue which might indicate a potential minor nonstationarity in the data (Kumazawa and Ogata 2013). The Q–Q plot again shows that SR-ETAS models provide a better realisation to Exponential(1) distribution with minor overestimation for large quantiles, while standard ETAS underestimates it for a large proportion of the support.

The time rescaling diagnostic tests conclude that the CVM test was passed by all models except standard ETAS at the 5% significance level. All five models failed the LB test. We believe this was caused by the large sample size since overall there are no indications for dependence in the residuals. The ER test was passed by all non-Gamma-based models at the 5% significance level indicating that Gamma distribution might induce an excessive dispersion in the residuals. All things considered, we conclude that ETAS model

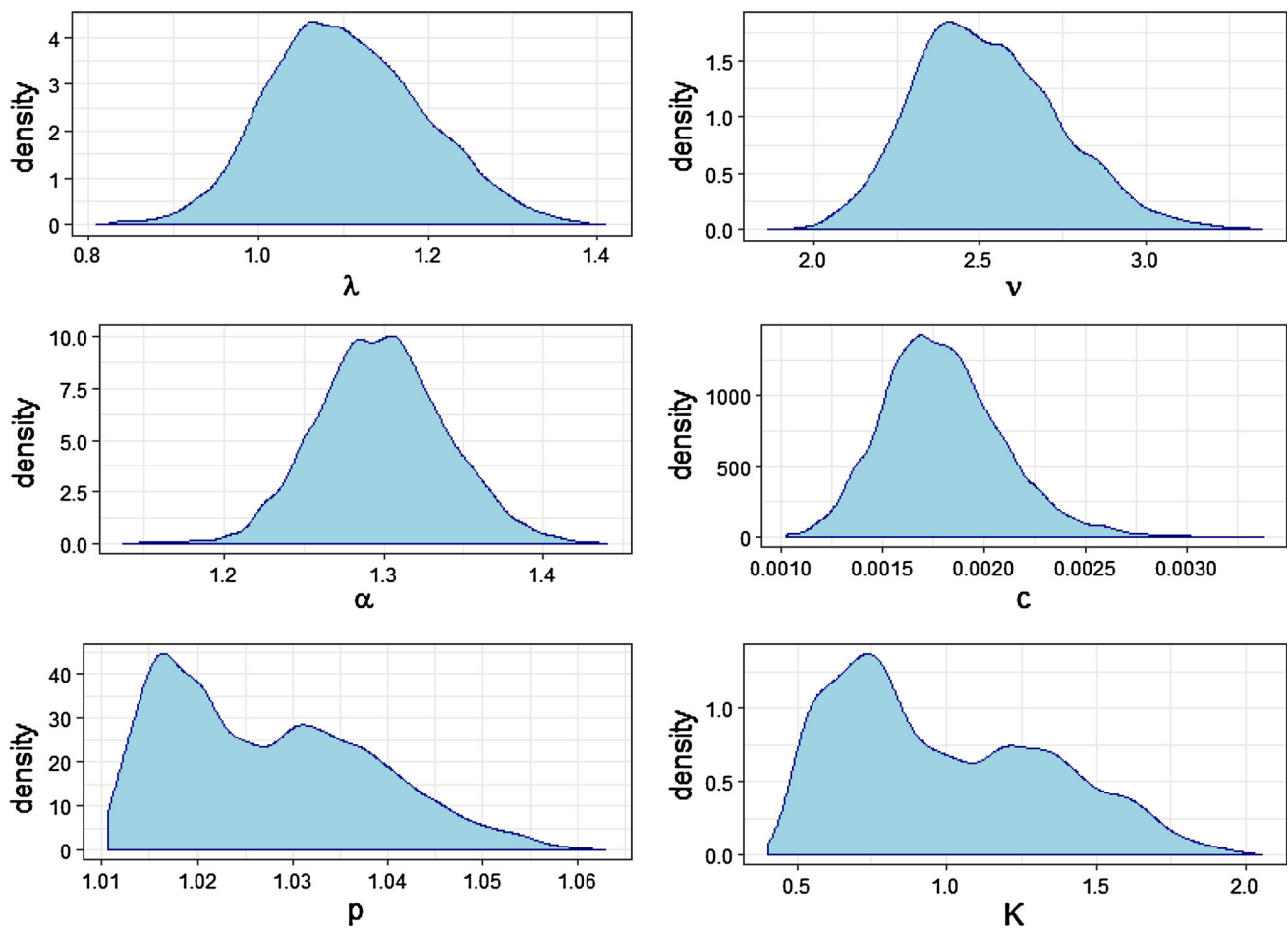


Fig. 6 B-B-ETAS MCMC parameters’ density for the North California catalogue

**Table 2** Goodness-of-fit summary—North California; ETAS, BPT and Gamma-based SR-ETAS

	ETAS	F-G-ETAS	B-G-ETAS	F-B-ETAS	B-B-ETAS
Log-likelihood	− 109.32	− 103.68	− 90.61 <sup>a</sup>	− 110.58	− 64.83 <sup>a</sup>
Number of parameters	5	6	6	6	6
BIC	129.68	128.11	115.04 <sup>a</sup>	135.01	89.26 <sup>a</sup>
DICalt	1108.51	799.57	725.33 <sup>a</sup>	553.48	338.07 <sup>a</sup>

Lower values of the BIC/DIC indicate superior fit

<sup>a</sup>The value is approximate

is not providing adequate fit to the North California catalogue, Gamma-SR-ETAS models are supreme to it but some researchers might disregard them due to the present excessive dispersion in the residuals. The B-SR-ETAS is providing the most stable results, with Branched-BPT ETAS being the supreme model that should be chosen for this dataset.

### 8 Conclusion

The ETAS model has proved to be one of the most widely used tools for modelling seismic activity in terms of both capturing specific features of interest and forecasting future

events. Its estimation can be considered challenging due to identifiability issues. In this work, we introduced the concept of temporally variable ground intensity based on stress release modelling. In this, we specified two families of SR-ETAS model that depend on either the occurrence time of the previous event in the sequence (Full-SR-ETAS), or the elapsed time from the last immigrant (main) event (Branched-SR-ETAS). Our experimental results suggest that these models capture observed features of real earthquake catalogues that the standard ETAS model does not.

Our experimental results suggest that these models capture features of real earthquake catalogues related to crustal

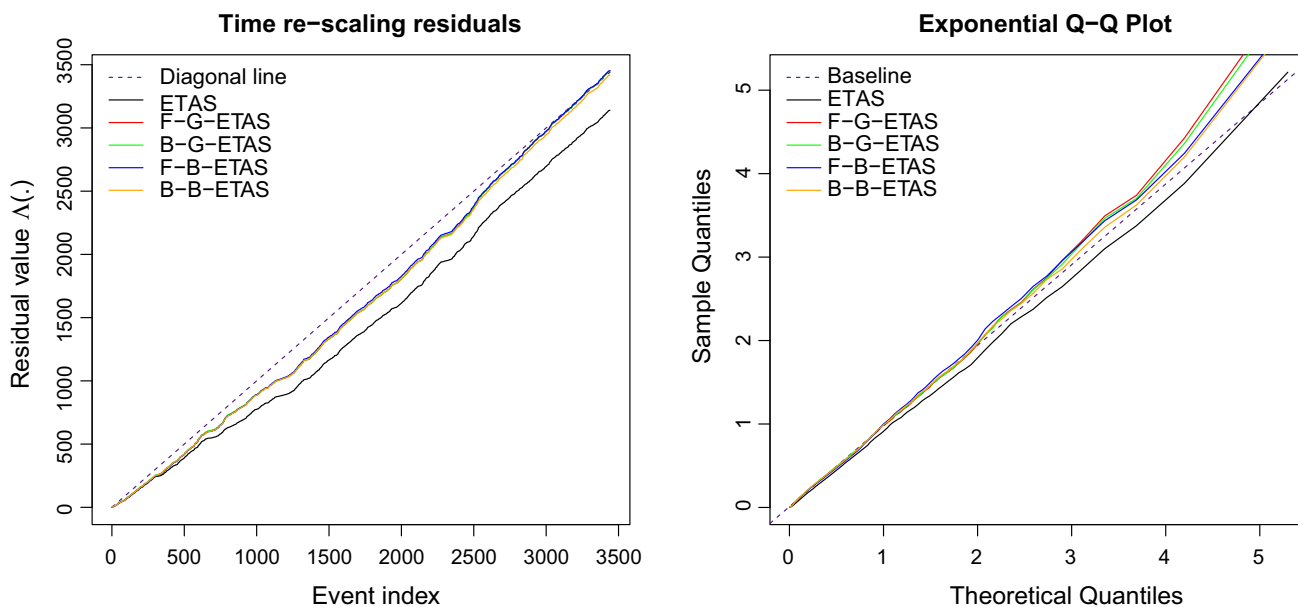


Fig. 7 Time rescaling diagnostic plots for the North California catalogue

strain budget that the standard ETAS model does not. Currently, we examined a single fault, nonspatial occurrence that is typically used by general seismologist for the analysis of a seismic fault activity. All concepts are directly applicable to the Spatial extension of ETAS. There are many alternatives of the spatial component(s) of the standard ETAS that provide a great differentiation amongst them which makes direct comparison of the introduced family of model impractical. Overall, the nonspatial alternative introduced in this paper will provide excellent results as long as there are no strong nonlinear or nonuniform patterns in the spatial distribution of the earthquakes along the fault of interest.

All methods are introduced for a general distribution, as such the SR-ETAS family can grow very quickly to accommodate the modelling needs of any sort of data. Direct application to stock daily changes, insurance claims, fraud and terrorist threats is feasible.

**Acknowledgements** We would like to express sincere appreciation for the detailed comments of the two reviewers. Their immense contribution helped us to address very critical issues that resulted in a substantial improvements to the paper.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

### Appendix A: Maximum likelihood estimation of RHawkes

The overall scope of the defined by Wheatley et al. (2016), Chen and Stindl (2018) RHawkes process is very similar to ETAS and relies on the following intensity function

$$\lambda^{Rh}(t) = \mu(t - t_{[i]}) + \sum_{t_i < t} \eta r(t - t_i)$$

where  $\int_0^\infty r(u)du = 1$ ,  $I_{[i]} = \max_j \{j|t_j < t_i \text{ and } B_j = 0\}$  and  $B = \{B_1, \dots, B_n\}$  is a branching realisation as described in Sect. 3. The proposed form for the likelihood is the following:

$$p(\mathcal{H}_i|\theta) = \begin{cases} \exp(-U(T)), & n = 0. \\ \mu(t_1 - t_0) \exp(-U(t_1) \times -U(T - t_1) - \eta R(T - t_1)), & n = 1. \\ \mu(t_1 - t_0) \exp(-U(t_1) \times \prod_{i=2}^n \sum_{j=1}^{i-1} p_{ij} d_{ij} \times \sum_{j=1}^n s_{n+1} p_{n+1,j}) & n \geq 2. \end{cases} \tag{16}$$

where

$$d_{ij} = (\mu(t_i - t_j) + \sum_{t < t_i} \eta r(t_i - t)) \times \exp(-U(t_i - t_j) + U(t_{i-1} - t_j) - \eta \sum_{t < t_i} R(t_i - t) + \eta \sum_{t < t_{i-1}} R(t_{i-1} - t)) \quad (17)$$

$$s_{n+1,j} = \exp \left\{ -[U(T - t_j) - U(t_n - t_j)] - \eta \left[ \sum_{t < T} R(T - t) - \sum_{t < t_n} R(t_n - t) \right] \right\} \quad (18)$$

the  $p_{ij}, i = \{2, \dots, n + 1\}, j = \{1, \dots, i - 1\}$  are obtained by initiating  $p_{21} = 1$  and updated recursively as follows

$$p_{ij} = \begin{cases} \frac{\sum_{t < t_{i-1}} \eta r(t_{i-1} - t)}{\mu(t_{i-1} - t_j) + \sum_{t < t_{i-1}} \eta r(t_{i-1} - t)} \times \frac{d_{i-1,j} p_{i-1,j}}{\sum_{j=1}^{i-2} p_{i-1,j} d_{i-1,j}}, & j = 1, \dots, i - 2 \\ 1 - \sum_{k=1}^{j-2} p_{ik} & j = i - 1 \end{cases} \quad (19)$$

for  $i = 3, \dots, n + 1$ ;

$$U(t) = \int_0^t \mu(s) ds \quad \text{and} \quad R(t) = \int_0^t r(s) ds.$$

The evaluation of  $d_{ij}$  is only feasible based on Stress Release/ Renewal density that has explicit intensity function. Otherwise, the quantity in Eq. 2 will be numerically unstable as previously discussed. Thus, the method is not directly applicable to a general family of renewal process distributions (e.g. B-B-ETAS).

### Appendix B: Branching update probabilities

In order to evaluate the branching probabilities we have to address the conditional probability of the branching ( $B$ ) given the data  $Y$ , model parameters  $\theta = \{\theta_{SR}, \theta_\Phi\}$  and potentially the previous obtained branching structure(s). The general likelihood function as introduced in Eq. 4 can be rearranged to indicate the immigration and offspring process as follows:

$$p(\mathcal{H}_t | \theta, B) = e^{-\int_0^T \mu(z | \mathcal{H}_t, B) dz} \prod_{i: B_i=0} \mu(t_i | \mathcal{H}_t, B) \times \prod_{j=1}^n \left[ e^{-\int_0^{T-t_j} \Phi(z | \mathcal{H}_t) dz} \prod_{i: B_i=j} \Phi(t_i - t_j | \mathcal{H}_t, B) \right] \quad (20)$$

For ETAS and Full-ETAS models, the branching structure ( $B$ ) is not needed for the evaluation of Eq. 2, thus the term  $\int_0^T \mu(z | \mathcal{H}_t) dz$  is independent of  $B$ . The conditional distribution of the branching structure, given a flat prior is then

$$p(B | Y, \theta_{SR}, \theta_\Phi) \propto \prod_{i: B_i=0} \mu(t_i) \prod_{j=1}^n \prod_{i: B_i=j} \Phi(t_i - t_j)$$

which lead us, as required, to the following branching updates:

$$-P(B_i = 0 | \mathcal{H}_t, \theta, B_{1:(i-1)}) = \frac{\mu(t_i - t_{I_{[i]}})}{\mu(t_i - t_{I_{[i]}}) + \Phi(t_i | \mathcal{H}_t)} \\ -P(B_i = j | \mathcal{H}_t, \theta, B_{1:(i-1)}) = \frac{h(t_i - t_j) \iota(m_j)}{\mu(t_i - t_{I_{[i]}}) + \Phi(t_i | \mathcal{H}_t)} \\ \text{for } j \text{ in } 1 \text{ to } i - 1$$

where  $I_{[i]} = \max_j \{j | t_j < t_i \text{ and } B_j = 0\}$ . In the case of Branched-ETAS or Rhawkes, as defined by Wheatley et al. (2016), Chen and Stindl (2018),  $\int_0^T \mu(z | \mathcal{H}_t, B)$  requires a branching realisation  $B$  for the calculation of the ground intensity as of Eq. 2. Let us define  $z_0, z_1, \dots, z_{m+1}$  to be,  $z_0 = 0, z_1, \dots, z_m$ —all  $m$  immigrant events’ times that were estimated to occur in the catalogue and  $z_{m+1} = T$ . Thus, we can rewrite the full data likelihood components that depend on  $\mu(\cdot)$  as follows:

$$e^{-\int_0^T \mu(z | \mathcal{H}_t, B) dz} \prod_{i: B_i=0} \mu(t_i | \mathcal{H}_t, B) = e^{-\int_0^{z_1} \mu(z) dz - \int_{z_1}^{z_2} \mu(z) dz - \dots - \int_{z_m}^{z_{m+1}} \mu(z) dz} \prod_{i=1}^m \mu(z_i - z_{i-1}) \propto e^{-\int_{z_{i-1}}^{z_i} \mu(z) dz - \int_{z_i}^{z_{i+1}} \mu(z) dz} \mu(z_i - z_{i-1}) \mu(z_{i+1} - z_i) \quad (21)$$

Based on the above approximation, the full conditional posterior distribution is:

$$p(B_i | Y, \theta, B) \propto \left[ e^{-\int_{t_{[i]}}^{t_i} \mu(z) dz - \int_{t_i}^{I_{[i]}^*} \mu(z) dz} \mu(t_i - t_{[i]}) \mu(t_{I_{[i]}^*} - t_i) \right]^{\mathbb{1}_{B_i=0}} \times \left[ e^{-\int_{t_{[i]}}^{I_{[i]}^*} \mu(z) dz} \mu(t_{I_{[i]}^*} - t_{[i]}) \right]^{\mathbb{1}_{B_i \neq 0}} \times \prod_{j=1}^n \Phi(t_i - t_j)^{\mathbb{1}_{B_i=j}} \quad (22)$$

where  $I_{[i]}^* = \min_j \{j | t_j > t_i \text{ and } B_j = 0\}$  and  $\mathbb{1}_x$  is 1 if  $x$  is True and 0 otherwise.

Thus, the required branching probability updates are

$$\begin{aligned}
 & -P(B_i = 0 | \mathcal{H}_i, \theta, B) \propto \mu(t_i - t_{I_{[i]}}) \\
 & \quad \times e^{-\int_{t_{I_{[i]}}}^{t_i} \mu(z) dz - \int_{t_i}^{t_{I_{[i]}^*}^*} \mu(t|B_i=0) dt} \mu(t_{I_{[i]}^*}^* - t_i) \\
 & -P(B_i = j | \mathcal{H}_i, \theta, B) \propto \Phi(t_i | \mathcal{H}_i) \\
 & \quad \times e^{-\int_{t_{I_{[i]}}}^{t_{I_{[i]}^*}^*} \mu(t|B_i \neq 0) dt} \mu(t_{I_{[i]}^*}^* - t_{I_{[i]}}) \\
 & \text{for } j \text{ in } 1 \text{ to } i - 1
 \end{aligned}$$

which can be further simplified to those stated in Eqs. 9 and 10 since  $e^{-\int_{t_{I_{[i]}}}^{t_i} \mu(z) dz}$  has the same value regardless of the state of the branching assignment of  $t_i$ .

## References

- Brown, E.N., Barbieri, R., Ventura, V., Kass, R.E., Frank, L.M.: The time-rescaling theorem and its application to neural spike train data analysis. *Neural Comput.* **14**(2), 325–346 (2002)
- Chen, C.-H., Wang, J.-P., Wu, Y.-M., Chan, C.-H., Chang, C.-H.: A study of earthquake inter-occurrence times distribution models in Taiwan. *Nat. Hazards* **69**(3), 1335–1350 (2013)
- Chen, F., Stindl, T.: Direct likelihood evaluation for the renewal Hawkes process. *J. Comput. Graph. Stat.* **27**(1), 119–131 (2018)
- Chib, S., Greenberg, E.: Understanding the Metropolis–Hastings algorithm. *Am. Stat.* **49**(4), 327–335 (1995)
- Ellsworth, W.L., Matthews, M.V., Nadeau, R.M., Nishenko, S.P., Reasenberg, P.A., Simpson, R.W.: A physically based earthquake recurrence model for estimation of long-term earthquake probabilities. *US Geol. Surv.* **522**, 23 (1999)
- Engle, R.F., Russell, J.R.: Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica* **66**, 1127–1162 (1998)
- Filimonov, V., Sornette, D.: Apparent criticality and calibration issues in the Hawkes self-excited point process model: application to high-frequency financial data. *Quant. Finance* **15**(8), 1293–1314 (2015)
- Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B.: *Bayesian Data Analysis*, vol. 2. CRC Press, Boca Raton (2014)
- Hamra, G., MacLehose, R., Richardson, D.: Markov chain Monte Carlo: an introduction for epidemiologists. *Int. J. Epidemiol.* **42**(2), 627–634 (2013)
- Hyndman, R.J.: Thoughts on the Ljung–Box test. (2014) <https://robjhyndman.com/hyndsight/ljung-box-test/>. Accessed 4 Aug 2017
- Imoto, M.: Application of the stress release model to the Nankai earthquake sequence, southwest Japan. *Tectonophysics* **338**(3), 287–295 (2001)
- Isham, V., Westcott, M.: A self-correcting point process. *Stoch. Process. Appl.* **8**(3), 335–347 (1979)
- Johnson, N.L., Kemp, A.W., Kotz, S.: *Univariate Discrete Distributions*, vol. 444. Wiley, Hoboken (2005)
- Kagan, Y., Knopoff, L.: A stochastic model of earthquake occurrence. In: *Proceedings of the Eighth International Conference on Earthquake Engineering*, vol. 1, pp. 295–302 (1984)
- Kumazawa, T., Ogata, Y.: Quantitative description of induced seismic activity before and after the 2011 Tohoku–Oki earthquake by non-stationary ETAS models. *J. Geophys. Res. Solid Earth* **118**(12), 6165–6182 (2013)
- Lallouache, M., Challet, D.: The limits of statistical significance of Hawkes processes fitted to financial data. *Quant. Finance* **16**(1), 1–11 (2016)
- Liu, J., Vere-Jones, D., Ma, L., Shi, Y.-L., Zhuang, J.-C.: The principle of coupled stress release model and its application. *Acta Seismologica Sinica* **11**(3), 273–281 (1998)
- Ljung, G.M., Box, G.E.: On a measure of lack of fit in time series models. *Biometrika* **65**(2), 297–303 (1978)
- Lu, C., Harte, D., Bebbington, M.: A linked stress release model for historical Japanese earthquakes: coupling among major seismic regions. *Earth Planets Space* **51**(9), 907–916 (1999)
- Marzocchi, W., Taroni, M.: Some thoughts on declustering in probabilistic seismic-hazard analysis. *Bull. Seismol. Soc. Am.* **104**, 1838–1845 (2014)
- Matthews, M.V., Ellsworth, W.L., Reasenberg, P.A.: A Brownian model for recurrent earthquakes. *Bull. Seismol. Soc. Am.* **92**(6), 2233–2250 (2002)
- Oakes, D.: The Markovian self-exciting process. *J. Appl. Probab.* **12**(1), 69–77 (1975)
- Ogata, Y.: Statistical models for earthquake occurrences and residual analysis for point processes. *J. Am. Stat. Assoc.* **83**(401), 9–27 (1988)
- Ogata, Y.: Space-time point-process models for earthquake occurrences. *Ann. Inst. Stat. Math.* **50**(2), 379–402 (1998)
- Ordaz, M., Arroyo, D.: On uncertainties in probabilistic seismic hazard analysis. *Earthq. Spectra* **32**(3), 1405–1418 (2016)
- Rasmussen, J.G.: Bayesian inference for Hawkes processes. *Methodol. Comput. Appl. Probab.* **15**(3), 623–642 (2013)
- Reid, H.F.: *The Mechanics of the Earthquake*, vol. 2. Carnegie Institution of Washington, Washington (1910)
- Ross, G.: Bayesian estimation of the ETAS model for earthquake occurrences. Preprint (2018a)
- Ross, G.: Nonparametric Bayesian inference for the Hawkes process with seasonal event data. Preprint (2018b)
- Rotondi, R., Varini, E.: Bayesian analysis of marked stress release models for time-dependent hazard assessment in the western Gulf of Corinth. *Tectonophysics* **423**(1), 107–113 (2006)
- Rotondi, R., Varini, E.: Bayesian inference of stress release models applied to some Italian seismogenic zones. *Geophys. J. Int.* **169**(1), 301–314 (2007)
- Schwarz, G., et al.: Estimating the dimension of a model. *Ann. Stat.* **6**(2), 461–464 (1978)
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Linde, A.: The deviance information criterion: 12 years on. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **76**(3), 485–493 (2014)
- Stephens, M.A.: Use of the Kolmogorov–Smirnov, Cramér–Von Mises and related statistics without extensive tables. *J. R. Stat. Soc. Ser. B (Methodol.)* **32**, 115–122 (1970)
- Tahernia, N., Khodabin, M., Mirzaei, N.: Non-Poisson probabilistic seismic hazard assessment. *Arab. J. Geosci.* **7**(8), 3259–3269 (2014)
- Varini, E., Rotondi, R.: Probability distribution of the waiting time in the stress release model: the Gompertz distribution. *Environ. Ecol. Stat.* **22**(3), 493–511 (2015)
- Veen, A., Schoenberg, F.P.: Estimation of space-time branching process models in seismology using an em-type algorithm. *J. Am. Stat. Assoc.* **103**(482), 614–624 (2008)
- Vere-Jones, D.: Earthquake prediction—a statistician’s view. *J. Phys. Earth* **26**(2), 129–146 (1978)
- Wang, J.-H., Chen, K.-C., Lee, S.-J., Huang, W.-G., Wu, Y.-H., Leu, P.-L.: The frequency distribution of inter-event times of  $m \geq 3$  earthquakes in the Taipei metropolitan area: 1973–2010. *Terr. Atmos. Ocean. Sci.* **23**(3), 269–281 (2012)



- Wheatley, S.: Extending the Hawkes Process, A General Outlier Test, Case Studies in Extreme Risk. Ph.D. Thesis, ETH Zurich, Zurich (2016)
- Wheatley, S.: Personal communication (2017)
- Wheatley, S., Filimonov, V., Sornette, D.: The Hawkes process with renewal immigration and its estimation with an EM algorithm. *Comput. Stat. Data Anal.* **94**, 120–135 (2016)
- Xiaogu, Z., Vere-Jones, D.: Further applications of the stochastic stress release model to historical earthquake data. *Tectonophysics* **229**(1–2), 101–121 (1994)
- Yang, W.-Z., Vere-Jones, D., Ma, L., Liu, J.: A method for locating the critical region of a future earthquake using the critical earthquake concept. *Earthquake* **20**(4), 28–38 (2000)
- Zhu, S.-B., Shi, Y.-L.: Improved stress release model: application to the study of earthquake prediction in Taiwan area. *Acta Seismologica Sinica* **15**(2), 171–178 (2002)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.