



# A Bayesian multiple changepoint model for marked poisson processes with applications to deep earthquakes

Lu Shaochuan<sup>1</sup>

Published online: 30 November 2018

© Springer-Verlag GmbH Germany, part of Springer Nature 2018

## Abstract

A multiple changepoint model for marked Poisson process is formulated as a continuous time hidden Markov model, which is an extension of Chib's multiple changepoint models (J Econ 86:221–241, 1998). The inference on the locations of changepoints and other model parameters is based on a two-block Gibbs sampling scheme. We suggest a continuous time version of forward-filtering backward-sampling algorithm for simulating the full trajectories of the latent Markov chain without utilizing the uniformization method. To retrieve the optimal posterior path of the latent Markov chain, i.e. the maximum a posteriori estimation of changepoint locations, a continuous-time version of Viterbi algorithm (CT-Viterbi) is proposed. The set of changepoint locations is obtainable either from the CT-Viterbi algorithm or the posterior samples of the latent Markov chain. The number of changepoints is determined according to a modified BIC criterion tailored particularly for the multiple changepoint problems of a marked Poisson process. We then perform a simulation study to demonstrate the methods. The methods are applied to investigate the temporal variabilities of seismicity rates and the magnitude-frequency distributions of medium size deep earthquakes in New Zealand.

**Keywords** Multiple changepoint models · Marked temporal point process · Continuous-time forward filtering backward sampling · Continuous-time Viterbi algorithms · Deep earthquakes · b-Values

## 1 Introduction

Deep Earthquakes, with the focal depth more than 50 km, form a significant portion in earthquake catalogue around the world, including New Zealand. Deep earthquakes are important in that they give indications of the structure of the earth, the dynamics of the crust and mantle, see Frohlich (2006). Although deep earthquakes are usually less destructive as shallow earthquakes, they do cause damage in some occasions. A study on the occurrence patterns of deep earthquakes may prove valuable for understanding the subduction process at convergent plate boundaries and the evaluation of geological hazards near surface such as shallow earthquakes and volcano activities. A common feature typically observed in the deep earthquake catalogue is the nonstationarity of two important statistics, i.e. the

seismicity rate and the magnitude-frequency distribution (MFG). We consider one type of nonstationarity: abrupt changes in the deep seismicity rate and the associated magnitude-frequency distribution. We characterize the time-varying pattern of the two statistics by use of a Bayesian multiple changepoint model for the marked temporal Poisson processes.

Changepoint models are widely applied for modelling heterogeneities appearing in a set of observations collected sequentially. These models split the data into disjoint segments with a (random) number of changepoints, so that observations in the same segment come from the same pattern and observations in different segments show heterogeneity. Changepoint models have been applied extensively in engineering, signal processing, bioinformatics, earthquake modelling (Yip et al. 2017), hydrology (Kehagias 2004) and finance, among many others. There are vast literatures on the methods and applications of changepoint models. Among them, wild binary segmentation (Fryzlewicz 2014) and simultaneous multiscale change point estimator (Frick et al. 2014) are very popular. We

✉ Lu Shaochuan  
lvshaochuan@bnu.edu.cn

<sup>1</sup> School of Statistics, Beijing Normal University, Xin Jie Kou Wai Da Jie 19, Beijing, People's Republic of China

consider Bayesian off-line inference for the changepoint models. Often, inference for the number, the locations of changepoints and other model parameters is based on Markov chain Monte Carlo methods, see Stephens (1994), Green (1995), Chib (1998), Lavielle and Lebarbier (2001) and Fearnhead (2006), among many others. It is noted that most of these models are applicable only for discrete time observations. Although some continuous time random processes may be approximated ideally by their discrete-time counterparts, it is necessary to develop multiple changepoint models in continuous time on its own right. The advantage of this approach is to avoid approximation errors by their discrete-time counterparts, which is often difficult to quantify. Marked point process is widely utilized in statistical modelling for hurricane occurrences, insurance claims (Elliott et al. 2007), earthquake occurrences (Ogata 1988; Yip et al. 2018) etc. For Poisson processes, Galeano (2007) proposed a binary segmentation algorithm along with a centralized and normalized cumulative sum statistic to detect changepoints for the intensity rate of Poisson events. The approach is based on asymptotic arguments for providing consistent estimates for the locations of changepoints. Yang and Kuo (2001) suggested a binary segmentation procedure for locating the changepoints and the associated heights of the intensity function of a Poisson process by using Bayes factor or its BIC approximation.

We consider a formulation of Bayesian multiple changepoint models to simultaneously monitor the structural breaks in Poisson intensity rate and the associated mark distribution, which is a continuous time extension of Chib's multiple changepoint models (1998). For this continuous time hidden Markov model, we suggest an approach to directly simulate the full trajectory of the latent Markov chain in a block Gibbs sampling scheme, which is often implemented through the uniformization method (Fearnhead and Sherlock 2006; Rao and Teh 2013). The number of changepoints is determined via a modified Bayes information criterion, which is tailored particularly for this multiple changepoint models for the marked Poisson processes.

The outline of the paper is as follows. In Sects. 2 and 3, we formulate a multiple changepoint model for the marked Poisson processes via a continuous-time hidden Markov models. We then introduce a continuous-time forward filtering backward sampling algorithm for sampling the full trajectory of the latent Markov process without resorting to the uniformization method. The maximum a posteriori (MAP) estimate of the trajectory of the latent Markov chain  $x(t)$ , i.e. the locations of changepoints, can be obtained from a continuous time version of Viterbi algorithm as indicated in Sect. 4. The number of changepoints in a marked Poisson process is chosen by a modified BIC

criterion. We then carry out simulation studies to demonstrate the method in Sect. 5. In the last section, we perform a case study for the New Zealand deep earthquakes. The temporal variabilities of the *deep seismicity rate* and the *magnitude-frequency distribution* are analysed via this multiple changepoint model and its implications for seismic hazards are illustrated.

## 2 Model formulation and the likelihood

### 2.1 Model formulation

Let  $Y(t)$  be a Poisson process attached with marks. Suppose it is subject to abrupt changes at  $m$  unknown time points  $0 \triangleq \tau_0 < \tau_1 < \dots < \tau_m < \tau_{m+1} \triangleq T$ . The process  $Y(t)$  is partitioned into  $m + 1$  segments by  $m$  changepoints, with the  $i$ -th segment consisting observations within  $[\tau_{i-1}, \tau_i)$ . We consider the multiple changepoint models with changepoint locations associated with the state transition times of an unobservable continuous time finite Markov chain  $x(t)$ . The intensity rate function of  $Y(t)$  is specified by  $\lambda_{x(t)} f_{x(t)}(y)$ , where  $\lambda_{x(t)}$  is the stochastic intensity rate of the ground process and  $f_{x(t)}(y)$  is the probability density function of the mark distribution. A sequence of Poisson events at  $\{t_1, \dots, t_n\}$  and associated marks  $\{y_1, \dots, y_n\}$  are observed over a time interval  $[0, T]$ . The transition rate matrix of  $x(t)$  is constrained to be consistent with a changepoint model, such that  $x(t)$  either sojourns in the previous state or jumps to the next level. Correspondingly, the transition rate matrix of  $x(t)$  is parameterized by

$$\mathbf{Q} = \begin{pmatrix} -q_1 & q_1 & 0 & \dots & 0 & 0 \\ 0 & -q_2 & q_2 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & & \\ 0 & 0 & 0 & \dots & -q_m & q_m \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}.$$

The chain  $x(t)$  starts from the state 1 and ends in the state  $m + 1$ . When  $x(t)$  is in the state  $i$ , the stochastic intensity rate of the ground process and the mark distribution of  $Y(t)$  is given by  $\lambda_i f_i(y)$ . The attached marks can be any univariate or multivariate variables. We shall specify the mark distributions in later sections. This type of model is exactly a Markov modulated Poisson process (MMPP) attached by state-dependent marks, which is a doubly stochastic point process with the stochastic intensity of the ground process and the mark distributions determined by an underlying irreducible finite Markov chain, see Lu (2012). This type of multiple changepoint models have the *conditional independence property*, such that conditional on the position of a changepoint, observations after the changepoint contain no information about segments and observations prior to

the changepoint. It is also a type of product partition models (Barry and Hartigan 1992) and a continuous-time generalization of Chib’s multiple changepoint model (1998).

**Remark 1** The marks can be any type of variables, which may be dependent or independent of the ground process. Particularly, when the attached mark is an indicator of the class to which the point belongs, it forms a multiple changepoint model for multivariate Poisson processes (Ramesh et al. 2013). In this case, changepoints in multiple Poisson sequences can be simultaneously monitored.

**Remark 2** This model formulation allows jointly monitoring the changepoints of Poisson rates and associated marks in two cases: changepoints occurring both in the ground process and the associated marks simultaneously, or changepoints occurring either in the ground process or the attached marks alone. This type of model formulation is more flexible than modelling changepoints of the ground process and the attached marks separately or individually, which is ideal for modelling “common” structural breaks.

**Remark 3** When all  $q_i$ s in  $\mathbf{Q}$  are equal, it suggests that the changepoints follow a constant rate Poisson process. As  $q_i$ s in  $\mathbf{Q}$  may be different, varying scale of segment lengths can be represented, which is desirable for modelling highly variable segment lengths between changepoints and avoids potential model bias.

This model formulation is a special case of MMPP attached by state-dependent marks (Lu 2012). However, current model parameterization is consistent with a multiple changepoint model with  $m + 1$  segment constraints and no “state reciprocal” is assumed by specifying a full transition rate matrix of the latent Markov chain as indicated in Lu (2012). See also the last paragraph in Sect. 6 for further discussions.

We define some notations used in the later sections. Denote the inter-event time  $t_i - t_{i-1}$  by  $\Delta t_i$  and the observation  $(t_i, y_i)$  by  $Y_i$ .  $x(t_k)$  is denoted by  $x_k$ . Generically, we define the brackets  $[j, i]$  after a matrix  $A$  as the  $(j, i)$ -th entry of  $A$ . The sample path of a random process  $Z(t)$  over a time interval  $[a, b]$  or  $[a, b)$  is denoted by  $Z[a, b]$  or  $Z[a, b)$  respectively.

### 2.2 The likelihood

The sequence  $\{(x_i, \Delta t_i, y_i)\}_{i=1}^n$  forms a Markov sequence with transition density matrix  $e^{(\mathbf{Q}-\Lambda)(t_i-t_{i-1})}\Lambda\Upsilon(y_i)$ , where  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_{m+1})$  and  $\Upsilon(y) = \text{diag}(f_1(y), \dots, f_{m+1}(y))$ . The likelihood is given by

$$L(\mathbf{Q}, \Lambda) = \mathbf{e}'_1 e^{(\mathbf{Q}-\Lambda)\Delta t_1} \Lambda \Upsilon(y_1) \dots e^{(\mathbf{Q}-\Lambda)\Delta t_n} \Lambda \Upsilon(y_n) \mathbf{1}, \tag{1}$$

where  $\mathbf{e}_i$  is a unit column vector with all entries being zero except the  $i$ th entry and  $\mathbf{1}$  is a column vector with all entries being unity, see Lu (2012) for the derivation of the likelihood.

The evaluation of the likelihood is facilitated by use of the forward and backward recursions. Denote  $e^{(\mathbf{Q}-\Lambda)(t_k-t_{k-1})}\Lambda\Upsilon(y_k)$  by  $L_k$ . The forward and backward probabilities are written by  $\alpha_{t_k}(i) = \mathbf{e}'_1 L_1 \dots L_k \mathbf{e}_i$  and  $\beta_{t_k}(j) = \mathbf{e}'_j L_{k+1} L_{k+2} \dots L_n \mathbf{1}$ . Therefore, the forward and backward probabilities are recursively given by

$$\begin{cases} \alpha_{t_{k+1}}(i) = \sum_j \alpha_{t_k}(j) L_{k+1}[j, i] \\ \beta_{t_k}(i) = \sum_j L_{k+1}[i, j] \beta_{t_{k+1}}(j), \end{cases} \tag{2}$$

where  $L_{k+1}[j, i]$  is the  $(j, i)$ -th entry of  $L_{k+1}$ . The likelihood in terms of this device is obviously written by  $L(\mathbf{Q}, \Lambda) = \sum_i \alpha_{t_i}(i) \beta_{t_i}(i)$  for each  $t$  or equivalently  $L(\mathbf{Q}, \Lambda) = \alpha_T(m + 1)$ . The forward and backward densities tend to zero or infinity exponentially fast as the number of observations accrue, leading to “under flow” or “over flow” problem in practical computations. This “numerical instability” problem is often treated by use of floating-point software or incorporation of scaling procedures as outlined in our R implementation.

## 3 Bayesian inference of multiple changepoint models

Let all the model parameters be denoted by  $\Theta = (\mathbf{Q}, \theta)$ . Suppose a prior  $\pi(\Theta)$  is specified for the model parameters  $(\mathbf{Q}, \theta)$ . In the Bayesian context, the posterior distribution  $\pi(\Theta|Y[0, T]) \propto \pi(\Theta)p(Y[0, T]|\Theta)$  is of interest. We discuss a block Gibbs sampling scheme to sample approximately from the posterior distribution of  $\Theta$ . The Gibbs sampling scheme involves two full conditionals: the full trajectories of the latent Markov chain  $x(t)$  given the model parameters and the model parameters conditioned on the trajectories of the latent Markov chain.

### 3.1 A continuous time forward filtering and backward sampling algorithm

Sampling the full trajectories of the latent Markov chain given model parameters is typically implemented by continuous time versions of forward filtering backward sampling algorithms, see Fearnhead and Sherlock (2006) and Rao and Teh (2013). In Fearnhead and Sherlock (2006) and Rao and Teh (2013), the continuous time forward filtering

backward sampling algorithms are based on the uniformization method. Alternatively, we suggest a direct approach to simulate trajectories of the latent Markov chain  $x(t)$  without resorting to uniformization method. We start from the discrete time version of forward filtering backward sampling algorithm, sampling from  $x_k \triangleq x(t_k)$ ,  $k = 1, \dots, n$  at Poisson arrival times conditioned on  $\{Y_t, 0 \leq t \leq T\}$ , see Chib (1998), Scott (2002) and Fearnhead and Sherlock (2006). The forward filtering recursion is given by

$$p(x_i = k | Y[0, t_i], \Theta) = \frac{\sum_{l=k-1}^k p(x_{i-1} = l | Y[0, t_{i-1}], \Theta) p(x_i = k, Y(t_{i-1}, t_i) | x_{i-1} = l, \Theta)}{\sum_k \sum_{l=k-1}^k p(x_{i-1} = l | Y[0, t_{i-1}], \Theta) p(x_i = k, Y(t_{i-1}, t_i) | x_{i-1} = l, \Theta)} \quad (3)$$

The state filtering is calculated recursively from the initial condition:  $p(x_0 = 1 | \Theta) = 1$ .

The state sequence  $X_n = (x_1, \dots, x_n)$  is sampled from the joint distribution

$$p(X_n | Y[0, T], \Theta) = p(x_{n-1} | x_n, Y[0, T], \Theta) \cdot p(x_{n-2} | x_{n-1}, Y[0, T], \Theta) \cdots p(x_1 | x_2, Y[0, T], \Theta).$$

After the filtering probabilities are stored, the backward sampling is implemented according to

$$p\left(x_i \middle| x_{i+1}, Y[0, T], \Theta\right) \propto p(x_i | Y[0, t_i], \Theta) p(x_{i+1}, Y(t_i, t_{i+1}) | x_i, \Theta). \quad (4)$$

In (3) and (4), it is required to evaluate  $p(x_{i+1}, Y(t_i, t_{i+1}) | x_i, \Theta)$  exactly. In this case,

$$p(x_{i+1}, Y(t_i, t_{i+1}) | x_i, \Theta) = \mathbf{e}_{x_i}' e^{(\mathbf{Q}-\Lambda)(t_{i+1}-t_i)} \Lambda Y(y_{i+1}) \mathbf{e}_{x_{i+1}}. \quad (5)$$

See the previous section for the likelihood of a Markov modulated Poisson process with state-dependent marks.

It is noted that the state sequence  $\{x_1, x_2, \dots, x_n\}$  should be in consecutive order such that  $x_{k+1} = x_k$  or  $x_{k+1} = x_k + 1$ . Otherwise, there exists at least one segment within  $(x_k, x_{k+1})$ , which is totally devoid of any Poisson observations. It is reasonable to assume no such segment appears in many practical scenes. To simulate the full trajectory of the latent Markov chain, it is necessary to simulate the exact state transition times of  $x(t)$ . When a state transition happens between two consecutive  $x_k$  and  $x_{k+1}$ , the exact transition time  $\tau$  may be simulated by the uniformization method, see Fearnhead and Sherlock (2006) and Rao and Teh (2013). Alternatively, we consider a direct approach to simulate it. The exact jump timing  $\tau$  is simulated according to the probability density

$$p\left(\tau \middle| x_k = i, x_{k+1} = i + 1, Y(0, T)\right) = \frac{q_i e^{-(q_i + \lambda_i)(\tau - t_k)} e^{-(q_{i+1} + \lambda_{i+1})(t_{k+1} - \tau)}}{e^{(Q-\Lambda)(t_{k+1} - t_k)} [i, i + 1]}, t_k \leq \tau \leq t_{k+1}, \quad (6)$$

where  $e^{(Q-\Lambda)(t_{k+1} - t_k)} [i, i + 1]$  is the  $(i, i + 1)$ -th entry of the matrix exponential  $e^{(Q-\Lambda)(t_{k+1} - t_k)}$ . The numerator of the above equation is the probability density of  $x(t)$  sojourn in the state  $i$  from  $t_i$  until  $\tau$ , then jumping to the state  $i + 1$  in  $[\tau, t_{k+1}]$ , while no Poisson event happens in the entire  $(t_k, t_{k+1})$ . The denominator of the above equation is the probability density of  $x(t)$  sojourn in the state  $i$  at  $t_k$  and stay in the state  $i + 1$  at  $t_{k+1}$  without Poisson arrivals in  $(t_k, t_{k+1})$ . The cumulative distribution function  $F(u) = \int_{t_k}^u p(v | x_k = i, x_{k+1} = i + 1, Y(0, T)) dv$  is given in closed form. Hence,  $\tau$  can be sampled directly by the inverse transformation  $\tau = F^{-1}(U)$  as follows:

$$\tau = t_k + \log\left(\frac{V e^{(Q-\Lambda)(t_{k+1} - t_k)} [i, i + 1]}{q_i e^{-(q_{i+1} + \lambda_{i+1})(t_{k+1} - t_k)}} U + 1\right) / V, \quad (7)$$

where  $U \sim U[0, 1]$  and  $V = (q_{i+1} + \lambda_{i+1} - q_i - \lambda_i)$ .

### 3.2 Simulation of $\mathbf{Q}$

The block Gibbs sampling scheme is completed by sampling the model parameters conditioned on the full path of the underlying Markov chain. Given  $x(t)$ , the distribution of  $\mathbf{Q}$  is independent of  $Y(t)$  and  $\theta$ . It is straightforward to simulate from  $\mathbf{Q}$  when the priors of the model parameters are selected from the conjugate ones. Suppose the prior distribution of  $q_i$  is  $\Gamma(a, b)$  with probability density  $\frac{b^a}{\Gamma(a)} q_i^{a-1} e^{-bq_i}$ . Then the joint prior of  $\mathbf{Q}$  is given by  $\prod_{i=1}^m \frac{b^a}{\Gamma(a)} q_i^{a-1} e^{-bq_i}$ . The hyper-parameters  $a$  and  $b$  are often selected empirically. The likelihood of  $x(t)$ ,  $0 \leq t \leq T$  is written by

$$\prod_{i=1}^m q_i e^{-q_i(\tau_i - \tau_{i-1})}.$$

Therefore, the posterior distribution of  $\mathbf{Q}$  is given by

$$p(\mathbf{Q} | x[0, T]) \propto \prod_{i=1}^m \frac{b^a}{\Gamma(a)} q_i^{(a+1)-1} e^{-[b+(\tau_i - \tau_{i-1})]q_i}. \quad (8)$$

According to (8),  $q_i$  is simulated from  $\Gamma(a + 1, b + (\tau_i - \tau_{i-1}))$ . The final step of each Gibbs iteration involves sampling from the full conditionals  $\theta | x(t), Y(t), t \in [0, T], \mathbf{Q}$ . We shall discuss it in the following sections once the mark distributions are specified exactly.

**Algorithm 3.1** Block Gibbs sampler for multiple changepoint models of the marked Poisson processes:

1. Sample the state sequence  $\{x_1, \dots, x_n\}$  of the latent Markov chain  $x(t)$  conditioned on the model parameters and observations by a discrete-time forward filtering backward sampling algorithm;
2. If there exists one jump between two consecutive  $x_k$ s, the exact jumping time, i.e. the location of a changepoint  $\tau$  is simulated according to (7);
3. Sample  $\mathbf{Q}$  and other model parameters  $\theta$  given the trajectories of the latent Markov chain  $x(t)$ .
4. Repeat the above steps until the last iteration.

### 4 The continuous-time viterbi algorithm and the number of changepoints

In discrete time setting, it is well-known that the optimal set of changepoints can be given by the dynamic programming algorithm- Viterbi algorithm. However, for the continuous time HMMs, there exists an infinite number of potential paths for the latent Markov chain. In Bebbington (2007), a continuous-time Viterbi algorithm is suggested to retrieve the optimal path of the latent Markov chain of a Markov modulated Poisson process. The method can be tailored to fit the current models. Define

$$M_i(t_k) = \max_{x[0,t_k]} \log p(x[0, t_k], x_k = i, Y[0, t_k]). \tag{9}$$

Viterbi recursion shows that

$$M_j(t) = \max_i \{M_i(t_k) + \log p(x(t) = j, Y(t_k, t] | x_k = i)\},$$

$$t_k \leq t \leq t_{k+1}. \tag{10}$$

In this case,  $\log p(x(t) = j, Y(t_k, t] | x_k = i)$  needs to be maximized over all possible paths. It is noted that  $x(t)$  has at most one jump over  $[t_k, t_{k+1})$ . When  $j = i$ ,  $\log p(x(t) = j, Y(t_k, t] | x_k = i)$  is a constant. Only  $\log p(x(t) = i + 1, Y(t_k, t] | x_k = i)$  needs to be maximized over the path space. Assume that the sample path of  $x(t)$  over  $(t_k, t]$  is given by  $x(t_k, u) = i, x[u, t] = i + 1$ . The probability density is given by  $q_i e^{-(q_i + \lambda_i)(u - t_k)} e^{-(q_{i+1} + \lambda_{i+1})(t - u)}$ , which is maximized by setting either  $u - t_k$  or  $t - u$  to zero. So the maximum of it is given by

$$q_i \max \{e^{-(q_i + \lambda_i)(t - t_k)}, e^{-(q_{i+1} + \lambda_{i+1})(t - t_k)}\}$$

$$= q_i e^{-\min\{q_i + \lambda_i, q_{i+1} + \lambda_{i+1}\}(t - t_k)}.$$

Hence, the latent Markov chain has state transitions only at event times of  $Y(t)$ . From the above argument, we have the following proposition.

**Proposition 1** *The optimal posterior path of the latent Markov chain  $x(t)$  of a Markov modulated Poisson process with marks given  $Y(t), 0 \leq t \leq T$  has state transitions only at the Poisson event times.*

From Proposition 1, it is suggested that the optimal set of changepoint locations are given exactly from the Poisson event times. Hence, the search for the set of most probable changepoint locations can be equivalently given by a discrete time version of Viterbi algorithm.

Let  $f_k(j)$  be the probability density of the most probable sample path of  $x(t)$  up to  $t_k$  when reaching to the state  $j$ , and  $\phi_k(j)$  be the optimal state at time  $t_{k-1}$  for the sample path reaching to the state  $j$  at time  $t_k$ .

**Algorithm 2** (Continuous-time Viterbi Algorithm)

1. Initialize  $f_1(j) = e^{-\Lambda \Delta t_1} \Lambda Y(y_1)[i, j]$  and set  $\phi_1(j) = 0$  for all  $j$ .
2. For  $k = 2, \dots, n$  and all  $j$ , recursively compute
 
$$f_k(j) = \max_i \{f_{k-1}(i) e^{(\mathbf{Q} - \Lambda) \Delta t_k} \Lambda Y(y_k)[i, j]\}$$
 and
 
$$\phi_k(j) = \operatorname{argmax}_i \{f_{k-1}(i) e^{(\mathbf{Q} - \Lambda) \Delta t_k} \Lambda Y(y_k)[i, j]\}.$$
3. Let  $x_n = \operatorname{argmax}_j f_n(j)$  and backtrack the state sequence as follows:
 
$$\text{For } k = n - 1, \dots, 1, x_k = \phi_{k+1}(x_{k+1}).$$

In the above algorithm,  $f_k(j)$  propagates to zero or infinity exponentially fast, which again cause underflow or overflow problem. Proper scaling procedure by taking logarithm of it or by other approaches are required in numerical computations, see our R implementation.

There exist some popular model selection methods specifically tailored for the challenging changepoint problems, see Green (1995), Zhang and Siegmund (2007) and Harchaoui and Lévy-Leduc (2010), among many others. One popular criteria is Bayes information criterion (BIC). However, due to irregularities of the likelihood in the multiple changepoint models, there is no justification for using BIC in this scenario. Recently, a large sample approximation to the Bayes factor bypassing the Taylor expansion is derived for Gaussian changepoint models (Zhang and Siegmund 2007) and Poisson changepoint models (Shen and Zhang 2012) under the assumption of uniform prior for the locations of changepoints. Similar to Zhang and Siegmund (2007), we estimate the Bayes factor of the changepoint model versus the homogeneous Poisson model with stationary exponential marks for large T and



$\lim_{T \rightarrow \infty} \tau_i/T = r_i, i = 1, \dots, m$ . Denote the multiple changepoint models with  $m$  changepoints by  $\mathcal{M}_m$ . We have the following theorem:

**Proposition 2** *Assuming a uniform prior for the changepoints  $\tau$  and other model parameters, then*

$$\log \frac{P(\mathcal{M}_m|Y[0, T])}{P(\mathcal{M}_0|Y[0, T])} = \log \left( \frac{\sup \hat{L}_m}{\sup \hat{L}_0} \right) - \sum_{i=0}^m \log(\hat{\tau}_{i+1} - \hat{\tau}_i) + (1 - m) \log T + O_p(1).$$

In (11), the first term is the generalized log-likelihood ratio statistics of the model with  $m$  changepoints relative to the null model with no changepoint. The rest of it is interpreted as a (negative) penalty term posed to the model complexity. Generally, the penalty term favors evenly distributed changepoints. See the proof sketch in “Appendix” section. Thus, the modified BIC is defined as a penalized likelihood criterion:

$$\log L_m(\hat{\Theta}) - \sum_{i=0}^m \log(\hat{\tau}_{i+1} - \hat{\tau}_i) + (1 - m) \log T, \quad (11)$$

where  $\hat{\Theta}$  is the maximum likelihood parameter estimation of the model  $\mathcal{M}_m$ .

## 5 Simulation studies

We perform a simulation study to demonstrate the methods. The simulation will provide some insights for an application of the methods in deep earthquakes modelling in the next section. We choose the exponential marks as the attached marks in this simulation for the reason illuminated in the application in the next section. We also choose conjugate priors  $\Gamma(\alpha, \beta)$  and  $\Gamma(\zeta, \eta)$  for the Poisson intensity rates  $\Lambda$  and the rate parameters  $\rho$  in the exponential type of marks respectively. Therefore, the full conditionals of  $\Lambda$  are written by

$$\Lambda|x[0, T], Y[0, T], \mathbf{Q} \propto \prod_{i=1}^{m+1} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda_i^{\alpha+N_i-1} e^{-[\beta+(\tau_i-\tau_{i-1})]\lambda_i}, \quad (12)$$

where  $N_i$  is the number of Poisson events arrived in the  $i$ -th segment. Similarly, the full conditionals of  $\rho$  are given by

$$\rho|x[0, T], Y[0, T], \mathbf{Q}, \Lambda \propto \prod_{i=1}^{m+1} \frac{\zeta^\eta}{\Gamma(\eta)} \rho_i^{\eta+N_i-1} e^{-[\zeta+S_i]\rho_i}, \quad (13)$$

where  $S_i$  is the cumulative summation of the marks in the  $i$ -th segment.

We perform three simulations. In the first case, a multiple changepoint model with simultaneous changepoints occurring in both the Poisson rate and associated marks is considered. A three-changepoint model is assumed the true model with specified parameters listed in (a) rows of Table 1. In each segment of the model, we simulate only 50 Poisson events attached by exponential marks. The total number of simulated observations is 200. So, the exact locations of changepoints are 51, 101, 151. We fit the simulated data by multiple changepoint models with 2~4 changepoints. For each model, Gibbs sampling iterates 100,000 times, starting from two different initial values and hyperparameters and the last 10,000 samples are treated as posterior samples. The estimates of parameters, including  $\Lambda$  and  $\rho$ , are given by the posterior means of the samples. Generally, the results are not very sensitive to the initial values and hyperparameters in the priors. We present only one of the simulation results.

It is noted that the four segments are moderately separated, as indicated by the Poisson rates  $\Lambda$  and  $\rho$  specified in the true model, see (a) rows of Table 1. However, with only 50 Poisson arrivals attached with marks in each segment, it is still hard for the algorithm to accurately locate the changepoints and estimate the model parameters. For a three-changepoint model, the locations of changepoints, the Poisson rates and  $\rho$  in each segment are properly estimated in comparison to the true values, see (c) rows of Table 1. However, according to (b) rows of Table 1, it is noted that the log-likelihood of a two-changepoint model, rather than that of a three-changepoint model, is the highest among all the models. With greater penalties for the 3~4 changepoint models, the modified BIC of 3~4 changepoint models is smaller. So, it is unnecessary to list all of them in Table (1). Obviously, a parsimonious model, i.e. a two-changepoint model, is preferred for this short sequence with only 200 events in total in terms of the modified BIC, which actually deviates from the true data generating mechanism.

The second simulation is a piecewise marked Poisson process parameterized nearly the same as the previous numerical example, with the Poisson rate  $\Lambda = (2, 5, 2, 5)$  and the exponential rate  $\rho = (3, 2, 3, 2)$ , see (a) rows in Table (2). However, in each segment, we simulate 150 observations, instead of only 50 events as in the previous numerical example. The total number of simulated observations is 600. The exact locations of changepoints are 151, 301, 451. We fit the simulated data by multiple changepoint models with 2~4 changepoints. Again, the Gibbs sampler iterates 100,000 times in each case, with the last 10,000 samples treated as the posterior samples. With more observations available, it is reasonably well for the algorithm to accurately locate the changepoints and estimate the model parameters. For the three-changepoint model,

**Table 1** The table lists the estimated  $\mathbf{Q}, \Lambda, \rho$ , the locations of changepoints  $\tau_s$ , along with  $\log L$  for multiple changepoint models

Model	$\mathbf{Q}$	$\Lambda$	$\rho$	$\tau$	
(a) True model		$\lambda_1 = 2$	$\rho_1 = 2$	$\tau_1 = 51$	
		$\lambda_2 = 5$	$\rho_2 = 3$	$\tau_2 = 101$	
		$\lambda_3 = 2$	$\rho_3 = 2$	$\tau_3 = 151$	
		$\lambda_4 = 5$	$\rho_4 = 3$		
(b) Model for 3 segments	$q_{12} = 0.057$	$\lambda_1 = 2$	$\rho_1 = 1.95$	$\tau_1 = 49$	$\log L = -164.1$
	$q_{23} = 0.12$	$\lambda_2 = 5.4$	$\rho_2 = 2.86$	$\tau_2 = 94$	
	$q_{33} = 0$	$\lambda_3 = 2.7$	$\rho_3 = 2.88$		
(c) Model for 4 segments	$q_{12} = 0.066$	$\lambda_1 = 2.1$	$\rho_1 = 2.0$	$\tau_1 = 45$	$\log L = -168$
	$q_{23} = 0.11$	$\lambda_2 = 4.9$	$\rho_2 = 2.7$	$\tau_2 = 89$	
	$q_{34} = 0.076$	$\lambda_3 = 2.8$	$\rho_3 = 2.6$	$\tau_3 = 134$	
	$q_{44} = 0$	$\lambda_4 = 3.5$	$\rho_4 = 3.0$		
(d) Model for 5 segments	$q_{12} = 0.1$	$\lambda_1 = 2.3$	$\rho_1 = 2$	$\tau_1 = 35$	$\log L = -181.7$
	$q_{23} = 0.11$	$\lambda_2 = 3.7$	$\rho_2 = 2.4$	$\tau_2 = 68$	
	$q_{34} = 0.11$	$\lambda_3 = 3.7$	$\rho_3 = 2.7$	$\tau_3 = 104$	
	$q_{45} = 0.10$	$\lambda_4 = 2.8$	$\rho_4 = 2.4$	$\tau_4 = 140$	
	$q_{55} = 0$	$\lambda_5 = 3.6$	$\rho_5 = 3$		

**Table 2** The table lists the estimated  $\mathbf{Q}, \Lambda, \rho$ , the locations of changepoints  $\tau_s$ , along with **MBIC** and  $\log L$  for multiple changepoint models

Model	$\mathbf{Q}$	$\Lambda$	$\rho$	$\tau$	
(a) True model		$\lambda_1 = 2$	$\rho_1 = 3$	$\tau_1 = 151$	
		$\lambda_2 = 5$	$\rho_2 = 2$	$\tau_2 = 301$	
		$\lambda_3 = 2$	$\rho_3 = 3$	$\tau_3 = 451$	
		$\lambda_4 = 5$	$\rho_4 = 2$		
(b) Model for 3 segments	$q_{12} = 0.02$	$\lambda_1 = 3$	$\rho_1 = 2.3$	$\tau_1 = 309$	$\log L = -541.3$ $\text{MBIC} = 29.2$
	$q_{23} = 0.03$	$\lambda_2 = 2.2$	$\rho_2 = 3.1$	$\tau_2 = 452$	
	$q_{33} = 0$	$\lambda_3 = 5.8$	$\rho_3 = 2.0$		
(c) Model for 4 segments	$q_{12} = 0.028$	$\lambda_1 = 2.2$	$\rho_1 = 3.2$	$\tau_1 = 150$	$\log L = -533.4$ $\text{MBIC} = 66.5$
	$q_{23} = 0.063$	$\lambda_2 = 5.1$	$\rho_2 = 1.8$	$\tau_2 = 300$	
	$q_{34} = 0.027$	$\lambda_3 = 2.1$	$\rho_3 = 3.1$	$\tau_3 = 452$	
	$q_{44} = 0$	$\lambda_4 = 5.8$	$\rho_4 = 1.9$		
(d) Model for 5 segments	$q_{12} = 0.05$	$\lambda_1 = 2.2$	$\rho_1 = 3.1$	$\tau_1 = 132$	$\log L = -542.6$ $\text{MBIC} = 35.7$
	$q_{23} = 0.15$	$\lambda_2 = 3.9$	$\rho_2 = 2.1$	$\tau_2 = 218$	
	$q_{34} = 0.11$	$\lambda_3 = 4.2$	$\rho_3 = 2.1$	$\tau_3 = 328$	
	$q_{45} = 0.07$	$\lambda_4 = 2.4$	$\rho_4 = 3.0$	$\tau_4 = 454$	
	$q_{55} = 0$	$\lambda_5 = 5.8$	$\rho_5 = 2.0$		

the lag-5 auto-correlations of  $\mathbf{Q}$  are all below 0.01, suggesting good mixing of the Gibbs sampler. The locations of the changepoints are accurately located and the estimated parameters are close to the true values, see (c) rows of Table 2. For multiple changepoint models with two or four changepoints, the estimated  $\mathbf{Q}, \Lambda, \rho$  and the locations of changepoints are listed in (b) and (d) rows of Table 2. In this case, MBIC is able to identify the number of changepoints of the true model.

The previous two numerical examples were designed for the marked Poisson process with simultaneous changepoints appearing in both the Poisson rates and associated marks. The following third numerical simulation is

designed for a multiple changepoint model with miscellaneous type of changepoints, in which some of the changepoints appear either in the Poisson rates or in the associated marks alone, others appear in both components. For this marked Poisson process, the Poisson rates are given by  $\Lambda = (2, 5, 2, 2)$  and the rate parameters in the exponential marks are given by  $\rho = (2, 3, 3, 2)$ , see (a) rows in Table 3. So, both the two types of changepoints appear in this numerical example. The true parameters of this simulation is comparable to the previous simulations. Similar to the second simulation, we generate 150 observations in each segment, with 600 observations in total. The exact locations of changepoints are 151, 301, 451.

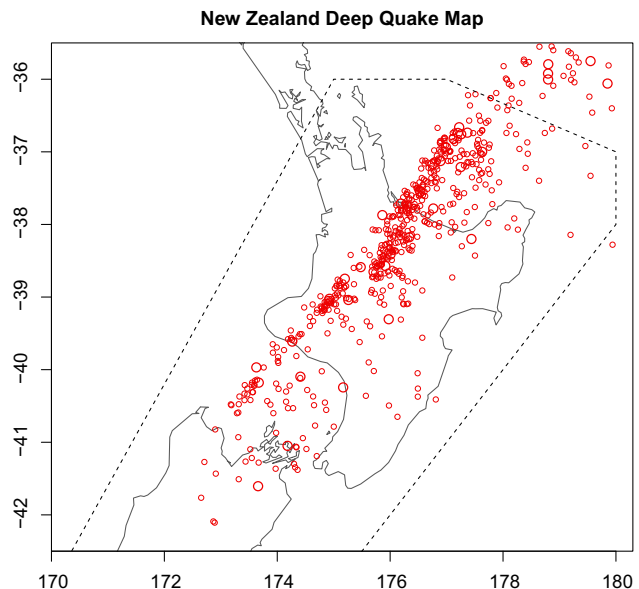
**Table 3** The table lists the estimated  $Q, \Lambda, \rho$ , the locations of changepoints  $\tau_s$ , along with **MBIC** and  $\log L$  for multiple changepoint models

Model	$Q$	$\Lambda$	$\rho$	$\tau$	
(a) True model		$\lambda_1 = 2$	$\rho_1 = 2$	$\tau_1 = 151$	
		$\lambda_2 = 5$	$\rho_2 = 3$	$\tau_2 = 301$	
		$\lambda_3 = 2$	$\rho_3 = 3$	$\tau_3 = 451$	
		$\lambda_4 = 2$	$\rho_4 = 2$		
(b) Model for 3 segments	$q_{12} = 0.023$	$\lambda_1 = 2$	$\rho_1 = 2.3$	$\tau_1 = 163$	$\log L = -549.8$ <b>MBIC</b> = 37.1
	$q_{23} = 0.07$	$\lambda_2 = 5.4$	$\rho_2 = 3.2$	$\tau_2 = 303$	
	$q_{33} = 0$	$\lambda_3 = 1.9$	$\rho_3 = 2.4$		
(c) Model for 4 segments	$q_{12} = 0.024$	$\lambda_1 = 2$	$\rho_1 = 2.3$	$\tau_1 = 162$	$\log L = -545.4$ <b>MBIC</b> = 44.4
	$q_{23} = 0.069$	$\lambda_2 = 5.3$	$\rho_2 = 3.2$	$\tau_2 = 303$	
	$q_{34} = 0.024$	$\lambda_3 = 2$	$\rho_3 = 3.4$	$\tau_3 = 457$	
	$q_{44} = 0$	$\lambda_4 = 1.8$	$\rho_4 = 1.8$		
(d) Model for 5 segments	$q_{12} = 0.07$	$\lambda_1 = 2.1$	$\rho_1 = 2.1$	$\tau_1 = 122$	$\log L = -564.3$ <b>MBIC</b> = 23.9
	$q_{23} = 0.17$	$\lambda_2 = 3.3$	$\rho_2 = 2.3$	$\tau_2 = 187$	
	$q_{34} = 0.1$	$\lambda_3 = 5$	$\rho_3 = 3.1$	$\tau_3 = 308$	
	$q_{45} = 0.034$	$\lambda_4 = 2$	$\rho_4 = 3.3$	$\tau_4 = 457$	
	$q_{55} = 0$	$\lambda_5 = 1.8$	$\rho_5 = 1.8$		

With the same number of observations as the previous simulation, it is more difficult for the algorithm to locate the changepoints and estimate other model parameters in this case. We fit the simulated data by multiple changepoint models with 2~4 changepoints. In each case, the Gibbs sampler iterates 100,000 times, with the last 10,000 samples treated as the posterior samples. For three-changepoint and two-changepoint models, the lag-5 auto-correlation of  $Q$  are all below 0.01. However, the Gibbs sampler converges rather slow for misspecified models. The estimated  $Q, \Lambda, \rho$  and the locations of changepoints are listed in Table 3. For the three-changepoint model, the locations of changepoints still can be properly located and the estimated parameters are close to the true values of the simulated model, see (c) rows of Table 3. In this case, the modified BIC properly identify the number of changepoints of the simulated time series.

### 6 An application to the deep earthquakes

The data set studied in this analysis is from New Zealand catalogue, which is freely obtainable from GNS Science of New Zealand via Geonet ([www.geonet.org.nz](http://www.geonet.org.nz)). We choose those events from New Zealand catalogue within confine as defined in Fig. 1 at depth greater than 50km with magnitude above 5 in New Zealand version of local magnitude scale. All the chosen events are either beneath the land or close to the shore. So, these events are under good coverage of monitoring networks. To avoid the analysis is biased by missing data, it is still necessary to assess the completeness threshold of the selected events. Generally, it is believed that the magnitude of completeness of the selected



**Fig. 1** Epicenter distribution of deep earthquakes with magnitude above 5 between 1965 and 2013. Events encircled by dash lines and map boundaries within a polygon with vertexes  $(170^\circ E, 43^\circ S)$ ,  $(175^\circ E, 36^\circ S)$ ,  $(177^\circ E, 36^\circ S)$ ,  $(180^\circ E, 37^\circ S)$ ,  $(180^\circ E, 38^\circ S)$ ,  $(173^\circ E, 45^\circ S)$  are considered

events is below 5 in this period, see, e.g., the analysis of the catalogue completeness for New Zealand deep earthquakes by various techniques in Lu and Vere-Jones (2011) and Lu (2012). Some descriptive properties of New Zealand deep earthquakes, such as the epicentral and depth distributions etc., are given in Lu and Vere-Jones (2011).

We focus on deep earthquakes modelling in this analysis. The occurrence rate of earthquakes is a direct indicator of the level of seismicity, which is often characterized



by the intensity function of a finite point process (Daley and Vere-Jones 2003). The magnitude-frequency distribution, also called Gutenberg-Richter (G-R) law, indicates that the cumulative number of earthquakes  $N(M)$  above  $M_c$  with magnitude  $M$  follows the log-linear relation:

$$\log_{10} N(M) = a - b(M - M_c), \tag{14}$$

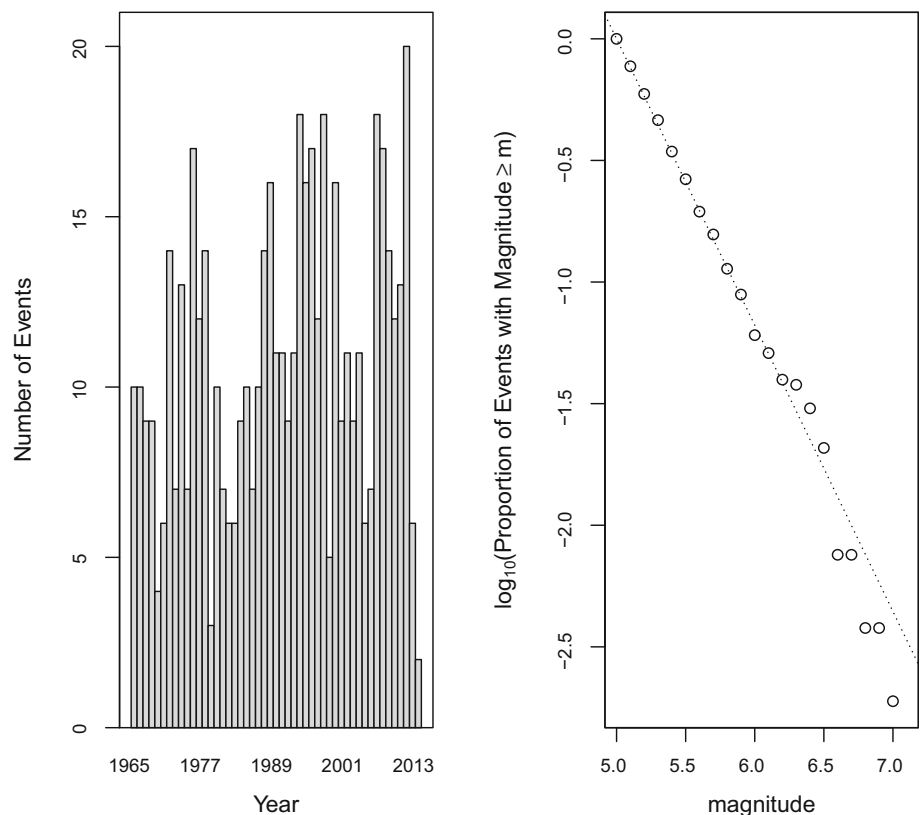
where  $a$  and  $b$  are constants and  $M_c$  is the magnitude threshold. The slope  $b$ , also called “ $b$ -value” in geophysical communities, indicates the relative proportion of the number of small and large earthquakes, corresponding to  $\beta$  in the exponential distribution:  $F(x) = 1 - e^{-\beta x}$ ,  $\beta = b \log 10$ . See the right part of Fig. 2 for the magnitude-frequency distribution of selected deep earthquakes. It seems that the log-linear relation of the magnitude-frequency distribution holds well. So, the figure suggests the completeness of the magnitude for the selected earthquakes. Otherwise, if the left end of the magnitude-frequency distribution is below the predicted log-linear relation of the G-R law, it is often believed that the catalogue is incomplete for smaller events due to limited detectability.

Generally, unlike shallow earthquakes, deep earthquakes rarely have following sequence of small aftershocks which decay according to Omori’s law. Instead, the typical occurrence pattern of deep earthquakes is that it varies from time to time, active in one period and relatively

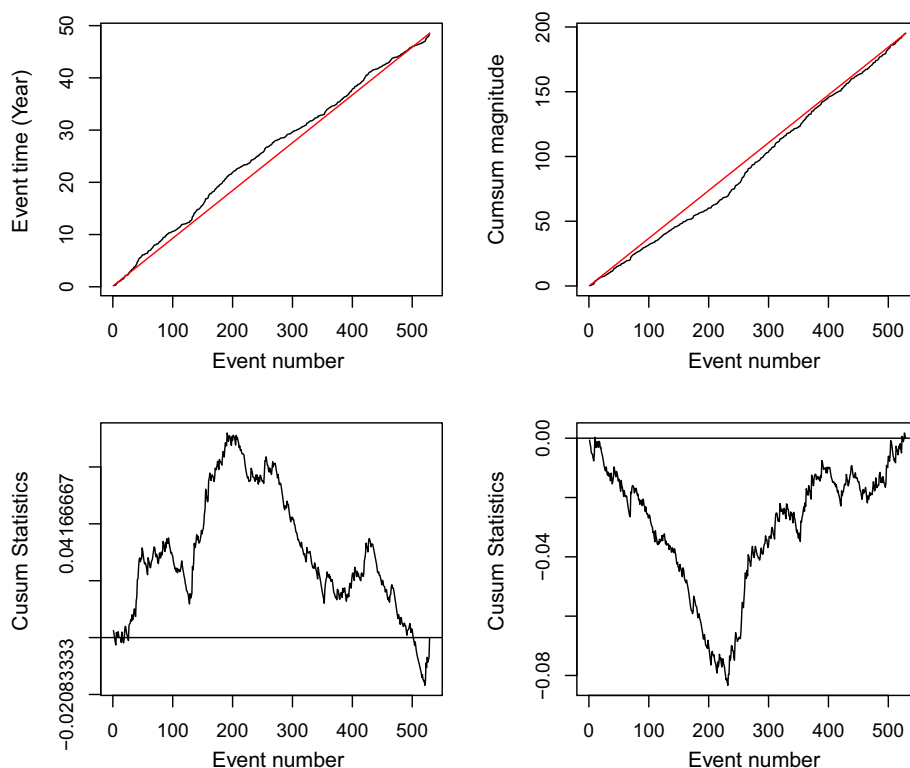
quiescent in another, see the left part of Fig. 2 for the yearly counts of deep events. We also demonstrate the centralized and normalized cumulative sums of inter-event times  $\Delta t_i$  in the left bottom of Fig. 3, which is given by  $\frac{\sum_{i=1}^j \Delta t_i}{\sum_{i=1}^n \Delta t_i} - \frac{j}{n}$ . For a homogeneous Poisson process, the statistic should be close to the line segment:  $y = 0, 0 \leq x \leq 1$ , which behaves as a standard Brownian bridge on  $[0, 1]$ , see Galeano (2007). From the left panel of Fig. 3 and the left part of Fig. 2, it is noted that the deep earthquakes show some non-stationarity. A constant rate Poisson process will not be adequate for fitting the occurrence rate of the deep earthquakes. Instead, Markov modulated Poisson process is viable for characterizing the time-varying behavior appearing in the occurrence rate of deep earthquakes.

Another statistic of interest is the magnitude-frequency distributions (MFD). Interpretation of the  $b$ -value of earthquake MFD has led to considerable attention in geophysical community. Wide range of observation studies suggest the  $b$ -value varies both spatially and temporally, see Wiemer et al. (1998). The  $b$ -value variability has been considered to be due to the ambient stress state, material heterogeneity, focal depth and geothermal gradient, which are directly or indirectly associated with the effective stress state. The relationship has been utilized for seismic hazard evaluation and risk forecasting, see Nanjo et al. (2012),

**Fig. 2** The left part of the plot is the yearly counts of deep earthquakes. The right part of the plot is the magnitude-frequency distribution of deep earthquakes. The logarithm of the frequency of magnitude is nearly linear with respect to the magnitude for a complete catalogue



**Fig. 3** The two figures in the left panel are the cumulative sum, the centralized and normalized cumulative sum of the inter-event times for deep earthquakes respectively. The two figures in the right panel are the cumulative sum, the centralized and normalized cumulative sum of earthquake magnitudes respectively



Schorlemmer and Wiemer (2005), Nuannin et al. (2005) and Lu (2017), among many others. Similarly, we demonstrate the  $b$ -value variability by the centralized and normalized cumulative sums of (trimmed) magnitudes  $\frac{\sum_{i=1}^j (m_i - M_c)}{\sum_{i=1}^n (m_i - M_c)} - \frac{j}{n}$  in the right bottom of Fig. 3. Again, when there is no changepoint occurring in the  $b$ -value, the statistic should behave like a standard Brownian bridge on  $[0, 1]$ . From the right top and right bottom of Fig. 3, it is noted that  $b$ -value variation appears for the deep earthquakes. Also, it is observed that there exists some sort of “coupling” between the deep seismicity rates and the  $b$ -value, see the bottom two graphs of Fig. 3. Both the seismicity rate and the  $b$ -value show a change at about 200.

One approach to jointly modelling the temporal variabilities of deep seismicity rate and the  $b$ -value is by use of the Markov modulated Poisson process attached by state-dependent marks, formulated as in the previous sections. The occurrence time distribution and the magnitude-frequency distribution of deep earthquakes are fitted by the multiple changepoint models with  $1 \sim 6$  changepoints. In each case, the Gibbs sampling scheme iterates 500,000 times, with the last 10,000 samples treated as the posterior samples. For this short sequence, it seems that the models with  $2 \sim 6$  changepoints overfit the data, as indicated by the MBIC listed in Table 4. From Table 4, it is also observed obvious similarities appearing among the Poisson intensity rates and the rate parameters in the mark distributions,

particularly for multiple changepoint models with  $3 \sim 6$  changepoints, which again suggests the multiple changepoint models with  $2 \sim 6$  changepoints overfit the data. The single changepoint model is sufficient to characterize the heterogeneity of the data. For this single changepoint model, after 500,000 Gibbs iterations, the 5-lag autocorrelations for all the parameters are less than or around 0.01, which suggests the Markov chain is mixing well. The posterior summary for the single changepoint model is displayed in Fig. 4. The figure displays the kernel density estimations and the 95% highest posterior density (HPD) intervals from 10,000 samples for some of the model parameters such as the location of the changepoint, the transition rate, the Poisson intensity rate and the rate parameter in the mark distribution. Both the upper and lower limits of the HPD intervals are indicated in the figure. The location of the changepoint appears a bit diffusive as indicated in the left top of the figure, which may result from a progressive rather than an abrupt change in the seismicity rate and (or) the  $b$ -value. The posterior inference is relatively robust over the selection of a range of proper hyperparameters.

The posterior mean of the location of changepoint  $\tau$  is around 1987. Before this point, the deep seismicity is relatively quiescent, with 9 deep events per year and a relatively high  $b$  value above  $3.2 / \log(10) \approx 1.3$ . Since 1987, it appears that the deep seismicity rate increased and the  $b$  value dropped to  $2.4 / \log(10) \approx 1$ . The deep seismicity

**Table 4** The table lists the estimated  $Q, \Lambda, \rho$ , the locations of changepoints  $\tau_s$ , along with MBIC and logL for multiple changepoint models

Model	$Q$	$\Lambda$	$\rho$	$\tau$	
1 Segments		$\lambda = 10.8$	$\rho = 2.7$		$\text{LogL}_0 = 717.3$
2 Segments	$q_{12} = 0.08$	$\lambda_1 = 9$	$\rho_1 = 3.2$	$\tau_1 = 203$	
	$q_{22} = 0$	$\lambda_2 = 11.9$	$\rho_2 = 2.4$		MBIC = 722.1
Model for 3 segments	$q_{12} = 0.13$	$\lambda_1 = 9$	$\rho_1 = 3$	$\tau_1 = 151$	
	$q_{23} = 0.18$	$\lambda_2 = 9.3$	$\rho_2 = 2.6$	$\tau_2 = 254$	MBIC = 713
	$q_{33} = 0$	$\lambda_3 = 11.3$	$\rho_3 = 2.37$		
Model for 4 segments	$q_{12} = 0.18$	$\lambda_1 = 8.9$	$\rho_1 = 2.9$	$\tau_1 = 110$	
	$q_{23} = 0.25$	$\lambda_2 = 9.0$	$\rho_2 = 2.5$	$\tau_2 = 174$	MBIC = 709
	$q_{34} = 0.23$	$\lambda_3 = 9.1$	$\rho_3 = 2.6$	$\tau_3 = 256$	
	$q_{44} = 0$	$\lambda_4 = 11.4$	$\rho_4 = 2.4$		
Model for 5 segments	$q_{12} = 0.2$	$\lambda_1 = 8.9$	$\rho_1 = 2.9$	$\tau_1 = 100$	
	$q_{23} = 0.26$	$\lambda_2 = 8.9$	$\rho_2 = 2.5$	$\tau_2 = 155$	MBIC = 707.5
	$q_{34} = 0.26$	$\lambda_3 = 9$	$\rho_3 = 2.46$	$\tau_3 = 213$	
	$q_{45} = 0.24$	$\lambda_4 = 9.3$	$\rho_4 = 2.45$	$\tau_4 = 288$	
	$q_{55} = 0$	$\lambda_5 = 11$	$\rho_5 = 2.36$		
Model for 6 segments	$q_{12} = 0.22$	$\lambda_1 = 8.9$	$\rho_1 = 2.8$	$\tau_1 = 82$	
	$q_{23} = 0.28$	$\lambda_2 = 8.8$	$\rho_2 = 2.49$	$\tau_2 = 127$	MBIC = 709.6
	$q_{34} = 0.26$	$\lambda_3 = 8.93$	$\rho_3 = 2.52$	$\tau_3 = 178$	
	$q_{45} = 0.26$	$\lambda_4 = 8.95$	$\rho_4 = 2.47$	$\tau_4 = 232$	
	$q_{56} = 0.24$	$\lambda_5 = 9.6$	$\rho_5 = 2.42$	$\tau_5 = 313$	
	$q_{66} = 0$	$\lambda_6 = 10.7$	$\rho_6 = 2.34$		
Model for 7 segments	$q_{12} = 0.24$	$\lambda_1 = 8.8$	$\rho_1 = 2.77$	$\tau_1 = 69$	
	$q_{23} = 0.28$	$\lambda_2 = 8.6$	$\rho_2 = 2.51$	$\tau_2 = 111$	MBIC = 702.4
	$q_{34} = 0.27$	$\lambda_3 = 8.9$	$\rho_3 = 2.52$	$\tau_3 = 157$	
	$q_{45} = 0.28$	$\lambda_4 = 8.94$	$\rho_4 = 2.46$	$\tau_4 = 201$	
	$q_{56} = 0.27$	$\lambda_5 = 9$	$\rho_5 = 2.41$	$\tau_5 = 249$	
	$q_{67} = 0.24$	$\lambda_6 = 9.6$	$\rho_6 = 2.38$	$\tau_6 = 324$	
	$q_{77} = 0$	$\lambda_7 = 10.6$	$\rho_7 = 2.33$		

changed from a relatively quiescent period to an active period. The accelerated energy release by the deep seismicity after 1987 suggests that a relatively high risk for the occurrences of large deep earthquakes. Although most of the deep earthquakes including large ones don't pose direct threat to people, occasionally, some great deep earthquakes do cause severe damages. It is also worthwhile to evaluate the risks of the occurrence of large deep earthquakes. For instance, the expected number of deep earthquakes with magnitude greater than 6 per year after 1987 is

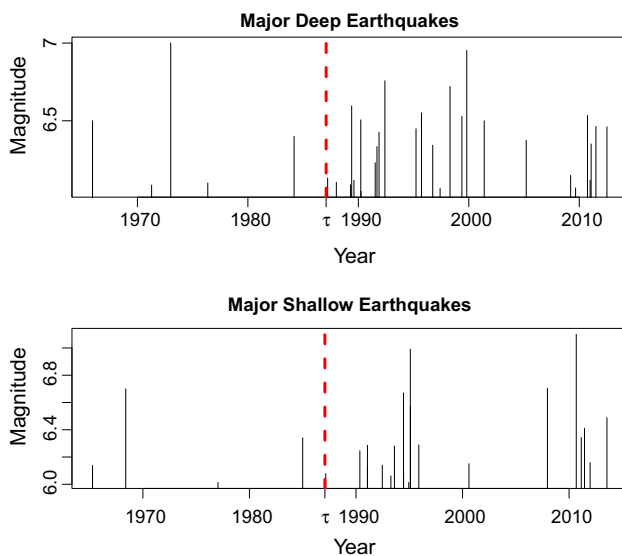
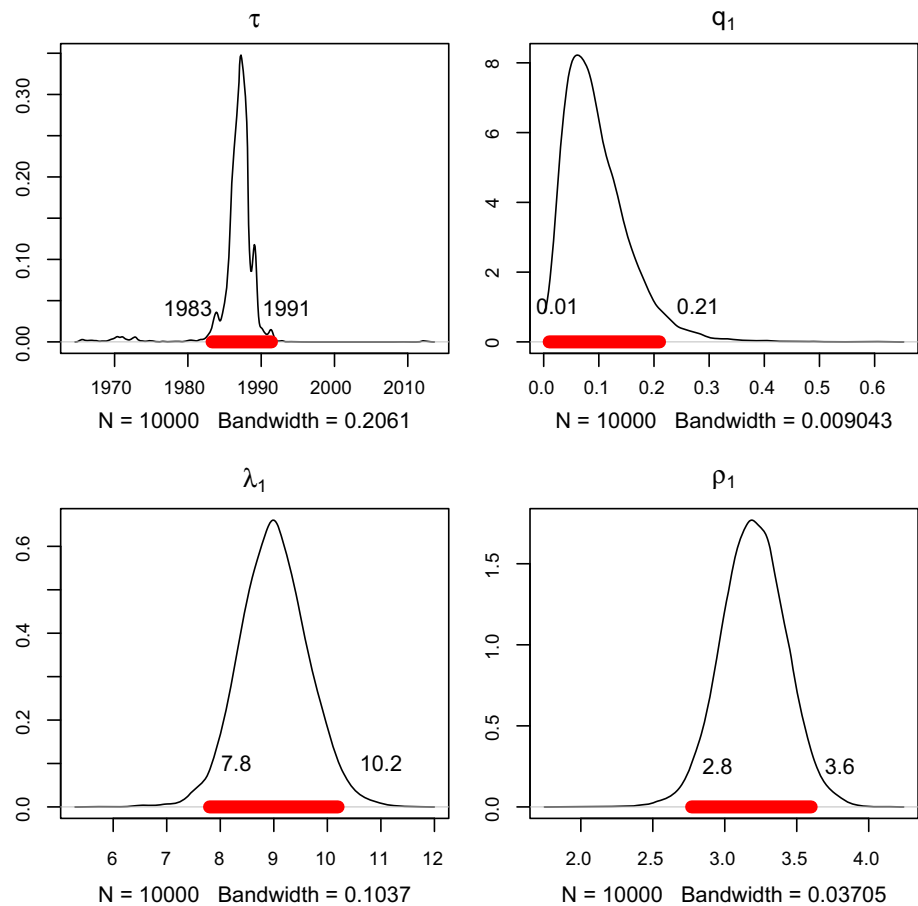
$$E\left(\sum_{i=1}^{N(1)} I_{M_i \geq 6}\right) = E(N(1))P(M \geq 6) = 1.09, \tag{15}$$

which is nearly three times of that before 1987. In the above equation,  $N(1)$  is the number of earthquakes occurred per year and  $I_{M_i \geq 6}$  is an indicator function for the  $i$ -th earthquake with magnitude greater than 6. The top of Fig. 5 demonstrates the Magnitude-Time plot for major deep earthquakes with magnitude above 6. The vertical red dash line shows the location of changepoint  $\tau$ , indicating a

strong contrast for the risks of large deep earthquakes before and after that point. In addition, it seems that some sort of “weak coupling” exists between the deep and shallow seismicity, see the bottom of Fig. 5. From the bottom of Fig. 5, it is observed that the seismically active episode of deep earthquakes roughly coincides with that of shallow earthquakes. The episode of high seismicity rate and high mean energy release per quake is also the episode of frequent occurrences of large shallow earthquakes. The increase of the moment release rate by deep earthquakes may be attributed to the increase in convergence rate of the tectonic plates, and hence increasing risks of geological hazards near surface such as shallow earthquakes and volcano activities in the most recent decade.

Nearly the same data set was analyzed in Lu (2012). In Lu (2012), a MMPP attached with state-dependent marks is applied to characterize the variabilities of seismicity rates and b-values, in which a full transition intensity rate matrix of the latent Markov chain is specified and the estimation of model parameters is implemented by the EM algorithm. This type of model formulation assumes there exists “state

**Fig. 4** Kernel density estimates for part of the model parameters. Line segments in bold beneath each kernel density estimate indicating the 95% highest posterior density for the model parameters. Both the lower and upper limits of the 95% HPD are indicated. The left top, right top, left bottom and right bottom of the figure show the posterior summary of the changepoint location  $\tau$ , the transition rate  $q_1$ , the Poisson rate  $\lambda_1$  and the rate parameter  $\rho_1$  in the exponential mark respectively. The number of posterior samples and the Bandwidth used in the kernel density estimation are given



**Fig. 5** The magnitude vs. time plot for major deep earthquakes and shallow earthquakes, with the location of the changepoint indicated by  $\tau$

reciprocals” in deep seismicity, which may cause model bias when there is actually no such “state reciprocals”. Furthermore, for some short sequence with only a few

“state transitions”, the estimation for the transition intensity rate matrix is unreliable, causing obvious instabilities in the state filtering and smoothing by methods in Lu (2012). Tiny perturbations may cause wholly different estimates for the latent state trajectories, leading to different decisions and scientific insights in practice. With a left-to-right restriction in the transition rate matrix, our model is suitable for modelling both a long sequence and a short sequence when the “state transitions” are rare. The current model formulation also avoids potential bias from the use of MMPPs with a full transition rate matrix when there is no underlying “state reciprocal” appearing in the MFD or (and) the deep seismicity rate. In addition, potential gains by Bayesian approach may include quantifying uncertainties of the number and positions of changepoints, which is difficult to deal with by other approaches.

## 7 Concluding remarks and discussion

In this study, we propose a Bayesian multiple changepoint model to jointly detect changepoints appearing in both the deep seismicity rate and the magnitude-frequency

distribution, which is an extension of Chib’s multiple changepoint model in continuous time. We suggest an approach to directly simulate the full trajectory of the latent Markov chain in a block Gibbs sampling scheme. The locations and numbers of changepoints can be given by a continuous time Viterbi algorithm and a modified BIC, tailored particularly for changepoint problems of marked Poisson processes. The model is applied to analyse the time-varying pattern of deep seismicity in New Zealand. It has been seen an increase in both the deep seismicity rate and the mean energy release per quake since 1987, in which most major deep earthquakes and shallow earthquakes occurred. The deep seismicity change may be attributed to an increase in the convergence rate of the tectonic plates, suggesting relatively high risk for the hazards near surface such as large shallow earthquakes in recent decades. The high seismicity and high mean energy release per quake show no signs of turning down until 2014 and continues. The method is potentially applicable for modelling insurance claims (Elliott et al. 2007).

The current model considers only one type of nonstationarity, in which the model parameters are piecewise constant, subject to abrupt changes at a fixed number of locations. In practice, it might be sensible to consider other forms of nonstationarity, such as progressive change or trend appearing in the model parameters, which is beyond the current model to characterize. In addition, in contrast to other approaches, current model formulation assumes only a fixed number of changepoints. It might be desirable to allow the number of changepoints to accrue unboundedly upon the arrival of new data. Finally, it seems that there exists some "weak coupling" between the deep seismicity and shallow seismicity. However, whether the large shallow earthquakes are preceded by an increase in deep seismicity or vice versa has not been thoroughly investigated so far. A study of it is potentially valuable for seismic hazards evaluation and earthquake risk forecasting (Kagan 2017) in subduction zones.

**Acknowledgements** Two referees’ suggestions are acknowledged.

### Appendix

Poisson process attached by exponential marks belongs to a two-parameter exponential family with canonical form:

$$\frac{d^2 F_j}{d\mu d\nu} = e^{\theta_j \mu - \Psi(\theta_j) + \eta_j \nu - \Phi(\eta)}$$

After fixing changepoints  $\tau$ , the log-likelihood can be written in a second order Taylor series around the

maximum likelihood estimate  $\hat{\Theta}(\tau) = \arg \max_{\Theta} l(\Theta, \tau)$ , such that:

$$l(\Theta, \tau) \approx l(\hat{\Theta}(\tau), \tau) + (\Theta - \hat{\Theta}(\tau))' H(\hat{\Theta}(\tau), \tau) (\Theta - \hat{\Theta}(\tau)) / 2, \tag{16}$$

where  $H(\cdot)$  is the Hessian matrix of the log-likelihood and  $\Theta = (\theta, \eta) \in \mathcal{R}^2$ . Therefore, after exponentiating  $l(\Theta, \tau)$  and treating it as a Gaussian kernel, under the Uniform priors for the changepoint locations, the marginal likelihood of the model  $\mathcal{M}_m$  is given by

$$P(\mathcal{M}_m | Y[0, T]) = \int_{\mathcal{D}_m} \int_{\mathcal{R}^{2m}} e^{l(\Theta, \tau)} \frac{m!}{T^m} d\Theta d\tau \tag{17}$$

$$= C \int_{\mathcal{D}_m} e^{l(\hat{\Theta}(\tau), \tau)} |H(\hat{\Theta}(\tau), \tau)|^{-\frac{1}{2}} \frac{m!}{T^m} d\tau,$$

where  $\mathcal{D}_m = \{(t_0, t_1, \dots, t_{m+1}) : 0 = t_0 < t_1 < \dots < t_{m+1} = T\}$  and  $C$  is a normalizing constant. The data points in the sample are independent,

$$l(\Theta, \tau) = \sum_{i=1}^{m+1} \left\{ \theta_i (S_{\tau_i}^u - S_{\tau_{i-1}}^u) - (\tau_i - \tau_{i-1}) \Psi(\theta_i) + \eta_i (S_{\tau_i}^v - S_{\tau_{i-1}}^v) - (\tau_i - \tau_{i-1}) \Phi(\eta_i) \right\},$$

where  $S_{\tau_i}^u$  denotes the sum of  $u$  from time 0 to  $\tau_i$ .

Obviously, the Hessian matrix is a diagonal matrix

$$H(\hat{\Theta}(\tau), \tau) = \text{diag}(\tau_1 \ddot{\Psi}(\hat{\theta}_1(\tau)), (\tau_2 - \tau_1) \ddot{\Psi}(\hat{\theta}_2(\tau)), \dots, (T - \tau_m) \ddot{\Psi}(\hat{\theta}_m(\tau)), \tau_1 \ddot{\Phi}(\hat{\eta}_1(\tau)), (\tau_2 - \tau_1) \ddot{\Phi}(\hat{\eta}_2(\tau)), \dots, (T - \tau_m) \ddot{\Phi}(\hat{\eta}_m(\tau)))$$

and

$$|H(\hat{\Theta}(\tau), \tau)| = \prod_{i=1}^{m+1} (\tau_i - \tau_{i-1})^2 \prod_{i=1}^{m+1} \ddot{\Psi}(\hat{\theta}_i(\tau)) \ddot{\Phi}(\hat{\eta}_i(\tau)). \tag{18}$$

In the denominator of Bayes factor, the marginal likelihood of the model  $\mathcal{M}_0$  is simply given by

$$P(\mathcal{M}_0 | Y[0, T]) = \int e^{l_0(\Theta)} d\Theta = \frac{1}{2\pi} e^{l_0(\hat{\Theta}_0)} |\ddot{l}_0(\hat{\Theta}_0)|^{-\frac{1}{2}}, \tag{19}$$

where  $\hat{\Theta}_0 = \arg \max_{\Theta} l_0(\Theta)$  and  $|\ddot{l}_0(\hat{\Theta}_0)| = T^2 \ddot{\Psi}(\hat{\theta}_0) \ddot{\Phi}(\hat{\eta}_0)$ .

Ignoring constant factors, the Bayes factor has approximation:

$$\int_{\mathcal{D}_m} e^{(l(\hat{\Theta}(\tau), \tau) - l_0(\hat{\Theta}_0))} \frac{1}{T^{m-1}} \prod_{i=1}^m (\tau_i - \tau_{i-1})^{-1} d\tau. \tag{20}$$

In above equation, the log of the integrant at the maximum likelihood value  $\hat{\tau}$  is the Modified BIC. To prove the



Modified BIC given in (11), it is necessary to show that the remainder term

$$\int_{\mathcal{D}_m} e^{l(\hat{\Theta}(\tau), \tau) - l(\hat{\Theta}(\hat{\tau}), \hat{\tau})} \prod_{i=1}^m (\hat{\tau}_i - \hat{\tau}_{i-1}) / \prod_{i=1}^m (\tau_i - \tau_{i-1}) d\tau. \quad (21)$$

is uniformly bounded in  $T$ , see the proof in Zhang and Siegmund (2007).

## References

- Barry D, Hartigan JA (1992) Product partition models for change point problems. *Ann Stat* 20:260–279
- Bebbington MS (2007) Identifying volcanic regimes using Hidden Markov Models. *Geophys J Int* 171:921–942
- Chib S (1998) Estimation and comparison of multiple change-point models. *J Econ* 86:221–241
- Daley DJ, Vere-Jones D (2003) An introduction to the theory of point processes. *Elementary theory and Methods*, vol 1. Springer, New York
- Elliott RJ, Siu TK, Yang H (2007) Insurance claims modulated by a hidden marked point process. *Am Control Confer* 2007:390–395
- Fearnhead P (2006) Exact and efficient bayesian inference for multiple changepoint problems. *Stat Comput* 16:203–213
- Fearnhead P, Sherlock C (2006) An exact Gibbs sampler for the Markov-modulated Poisson process. *J R Stat Soc B* 68(5):767–784
- Frick K, Munk A, Sieling H (2014) Multiscale change point inference. *J R Stat Soc B* 76(3):495–580
- Frohlich C (2006) *Deep earthquakes*. Cambridge University Press, Cambridge
- Fryzlewicz P (2014) Wild binary segmentation for multiple change-point detection. *Ann Stat* 42(6):2243–2281
- Galeano P (2007) The use of cumulative sums for detection of changepoints in the rate parameter of a Poisson process. *Comput Stat Data Anal* 51:6151–6165
- Green PJ (1995) Reversible Jump Markov Chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82:711–732
- Harchaoui Z, Lévy-Leduc C (2010) Multiple change-point estimation with a total variation penalty. *J Am Stat Assoc* 105:1480–1493
- Kagan YY (2017) Worldwide earthquake forecasts. *Stoch Environ Res Risk Assess* 31:1273. <https://doi.org/10.1007/s00477-016-1268-9>
- Kehagias A (2004) A hidden markov model segmentation procedure for hydrological and environmental time series. *Stoch Environ Res Risk Assess* 18(2):117–130
- Lavielle M, Lebarbier E (2001) An application of MCMC methods for the multiple change-points problem. *Signal Process* 81:39–53
- Lu S (2012) Markov modulated Poisson process associated with state-dependent marks and its applications to the deep earthquakes. *Ann Inst Stat Math* 64(1):87–106
- Lu S (2017) Long-term b value variations of shallow earthquakes in New Zealand: a HMM-based analysis. *Pure Appl Geophys* 174:1629–1641
- Lu S, Vere-Jones D (2011) Large occurrence patterns of New Zealand deep earthquakes: characterization by use of a switching poisson process. *Pure Appl Geophys* 168:1567–1585
- Nanjo KZ, Hirata N, Obara K, Kasahara K (2012) Decade-scale decrease in b value prior to the M9-class 2011 Tohoku and 2004 Sumatra quakes. *Geophys Res Lett* 39:L20304. <https://doi.org/10.1029/2012GL052997>
- Nuannin P, Kulhanek O, Persson L (2005) Spatial and temporal b value anomalies preceding the devastating off coast of NW Sumatra earthquake of December 26, 2004. *Geophys Res Lett* 32(11):L11307
- Ogata Y (1988) Statistical models for earthquake occurrences and residual analysis for point processes. *J Am Stat Assoc* 83(401):9–27
- Ramesh NI, Thayakaran R, Onof C (2013) Multi-site doubly stochastic poisson process models for fine-scale rainfall. *Stoch Environ Res Risk Assess* 27(6):1383–1396
- Rao V, Teh YW (2013) Fast MCMC sampling for Markov jump processes and extensions. *J Mach Learn Res* 14:3295–3320
- Schorlemmer D, Wiemer S (2005) Microseismicity data forecast rupture area. *Nature* 434:1086. <https://doi.org/10.1038/4341086a>
- Scott SL (2002) Bayesian methods for hidden Markov models: recursive computing in the 21st century. *J Am Stat Assoc* 97:337–351
- Shen JJ, Zhang NR (2012) Change-point model on nonhomogeneous Poisson processes with application in copy number profiling by next-generation DNA sequencing. *Ann Appl Stat* 6(2):476–496
- Stephens DA (1994) Bayesian retrospective multiple-changepoint identification. *Appl Stat* 43:159–178
- Wiemer S, McNutt S, Wyss M (1998) Temporal and three-dimensional spatial analyses of the frequency magnitude distribution near Long Valley Caldera, California. *Geophys J Int* 134(2):409–421
- Yang TY, Kuo L (2001) Bayesian binary segmentation procedure for a Poisson Process with multiple changepoints. *J Comput Graph Stat* 10:772–785
- Yip CF, Ng WL, Yau CY (2018) A hidden Markov model for earthquake prediction. *Stoch Environ Res Risk Assess* 32:1415. <https://doi.org/10.1007/s00477-017-1457-1>
- Zhang NR, Siegmund DO (2007) A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics* 63(1):22–32

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