BAYESIAN AND NON PARAMETRIC ESTIMATION OF ETAS MODELS APPLIED TO SEISMIC RECURRENCE IN ECUADOR 2016

Crespo Fernandez. Fausto Fabian[1], Jimenez Mosquera. Carlos[2],

[1]Department of Mathematics, University San Francisco de Quito, Quito, Ecuador

[2]Department of Mathematics, University San Francisco de Quito, Quito, Ecuador

In this paper, we analyzed, from the Bayesian point of view, the occurrence rate of earthquakes in Ecuador since March 2016 to July 16th, 2016. We use the Stan modeling language to implement the epidemic-type aftershock sequence (ETAS) model, starting with the temporal model, then considering the magnitudes of the earthquakes, and later the spatio-temporal models (both isotropic and anisotropic), and finally the hypo-central model. We use the Welzl algorithm to evaluate the log-likelihood of the occurrence rate for spatio-temporal models. Finally, we conducted simulations by extracting values from the a posteriori distribution of the model's parameters. We obtained estimations of the accumulated number of earthquakes (with magnitude greater than a threshold) and the behavior of inter-time events. The estimations are validated with the observed data from July 16th, 2016 to September 2016.
The Stan code for the ETAS models can be found in https://github.com/crespofabian8012/StanCon2018.

KEYWORD
Bayesian analysis, ETAS, MCMC, HMC, Rstan, Welzl algorithm, inter-event times and earthquake, Ecuador

INTRODUCTION

Ecuador is located in the so-called Ring of Fire and shows an intense seismic and volcanic activity. The subduction of the Nazca plate in the sea border between Ecuador and Colombia has caused four mega-earthquakes in the last century: 1906 (magnitude 8.8), 1942(magnitude 7.8), 1958(magnitude 7.7) y 1979 (magnitude 8.2) [Collot2004]. The Paleogene-Neogene faults of Jama-Quininde and Esmeraldas define a 200 kilometers long zone which was the rupture zone of the 1942 earthquake. [Collot2004].

The earthquake on April 16th, 2016, of magnitude 7.8 [Lotto2016] caused losses of millions of dollars and more than 650 deaths.

The empirical law of Omori [Omori1894] and the law of Omori-Utsu (also called Modified Law of Omori) {Utsu1995} describe the decreasing frequency of aftershocks over time after an earthquake:

$$N(t) = \frac{K}{(t+c)} \quad (1)$$

and

$$N(t) = \frac{K}{(t+c)^p} \quad (2)$$

where $N(t)$ is the occurrence rate of events, t is the time since the earthquake and K, c, p are constants.

The constant c allows us to use the value $t = 0$ in the formula and let us model complex aspects immediately after the main earthquake. The value of c is normally less than 0.1 days [Japan1998]. Omori's law is usually used to model the dependence of short-term earthquakes. The values of p that have appeared most in practice are values between 1 and 1.5. [Kagan1981].

Another empirical law commonly used is the Gutenberg-Ritcher law, which relates the magnitude to the frequency of occurrence of earthquakes with magnitudes greater than M:

$$log_{10}N(\geq M) = a - bM \tag{3}$$

In general, seismicity is modeled with two components: background seismicity and seismicity triggered by previous seismic events. These models are formulated according to the conditional intensity in the past history H_t [DaleyVere-Jones2002]: $\lambda(t, x, y, M \mid H_t)$ is the expected number of earthquakes in the unit of time, space, and magnitude.

An earthquake T is represented by a tuple $(x_i, y_i, z_i, t_i, M_i)$ where x_i, y_i, z_i are the longitude, latitude, and depth respectively (hypo-central coordinates), t_i is the time of occurrence of the earthquake and M_i is the magnitude of the earthquake i on the Richter scale.

The ETAS model [Ogata1998] is a special type of Hawkes point process, which are sometimes called branching models or self-excitation models. These models were called epidemics by Ogata in 1988 since each earthquake causes aftershocks and these, in turn, produce their own aftershocks. The simplest ETAS model is the temporal model with constant background seismicity:

$$\lambda(t \mid H_t) = \mu + \sum_{j:t_j<t} g(t - t_j) \tag{4}$$

where μ is the background intensity, assumed to be constant (measured in events/day) and

$$g(t) = \frac{K(p-1)c^{p-1}}{(t+c)^p} \tag{5}$$

is the probability density function of the occurrence times of the events triggered by previous earthquakes in time.

Assuming that the background intensity is not constant, but depends only on the longitude and latitude x, y (but not on the time) we have:

$$\lambda(t \mid H_t) = \mu(x, y) + \sum_{j:t_j<t} \frac{K(p-1)c^{p-1}}{(t-t_j+c)^p} \tag{6}$$

where $\mu(x, y)$ is now measured in events per day per unit of longitude and per unit of latitude. It is generally assumed that $\mu(x, y) = \mu\, u(x, y)$ where μ on the right side is a constant value now (different from $\mu(x, y)$) and $u(x, y)$ is the function that depends on the position.
Considering the magnitudes of earthquakes, we have the model

$$\lambda(t \mid H_t) = \mu + \sum_{j:t_j<t} g(t - t_j)\kappa\big(M_j\big) \tag{7}$$

where $\kappa(M) = Ae^{\alpha(M-M_0)}$ is the expected number of events triggered by an event of magnitude M.

M_0 is the threshold magnitude that is chosen arbitrarily as the minimum magnitude that provides completeness of the catalog of earthquakes. In [Sornette2005] and [Sornette2005b], it was shown that this threshold quantity is not related to the minimum magnitude needed to trigger other events, which is what actually represents M_0 in the ETAS model [Touati2011]. It was also shown that a greater M_0 causes the reduction of the parameter α and the ratio of branching (or the number of descendants earthquakes per father earthquake). That is, more earthquakes seem to be independent [Touati2011].

Analogously if we assume that the background intensity is not constant, but depends on the longitude and latitude x, y we have:

$$\lambda(t \mid H_t) = \mu(x,y) + \sum_{j:t_j<t} \frac{K(p-1)c^{p-1}Ae^{\alpha(M_j-M_0)}}{(t-t_j+c)^p} \tag{8}$$

The event rate in the ETAS models may explode. Stability depends on the branching ratio n (expected number of descendants of a parent event). We have

$$n = \int_0^\infty dt \int_{M_O}^{M_{max}} s(m)\, \lambda(t \mid H_t) dm \tag{9}$$

with $s(m) = \beta e^{-\beta m}$ the distribution of magnitudes by the Gutenberg-Richter law and $\lambda(t \mid H_t)$ the branching term in (7)

For the temporal ETAS model with magnitudes and $p > 1$ we have

$$n = \frac{Kc^{1-p}\beta}{(p-1)(\beta-\alpha)} \frac{(1-e^{-(\beta-\alpha)(M_{max}-M_0)})}{(1-e^{-\beta(M_{max}-M_0)})} \tag{10}$$

where β is the parameter of the distribution of the number of earthquakes greater than m, $s(m)$ as before and β =bln10 (b from Gutenberg-Richter Law). Assuming M_{max} is ∞ the previous formula can be reduced to: [Sornette2005] and [Touati2011]

$$n = \frac{Kc^{1-p}}{(p-1)(\beta-\alpha)} \tag{11}$$

If we assume $p < 1$ or if $\alpha > \beta$, the double integral in (9) is not convergent and n is infinite. If each event induces another event: n =1 then the process propagates indefinitely. This justifies normalizing the functions that appear in the summation over the preceding events. For example

$$\int_0^\infty \frac{K}{(t+c)^p} dt = 1 \tag{12}$$

implies that we need to add $(p-1)c^{p-1}$to the constant K and similarly

$$\int_{M_0}^{M_{max}} \beta e^{-\beta}\, dm = 1 \tag{13}$$

implies we need to add $\frac{1}{(e^{-\beta M_0} - e^{-\beta M_{max}})}$to the constant β.

The expected event rate as a function of time can be calculated as [Touati2011]

$$\bar{n}(t) = \int_{M_0}^{\infty} n(t)s(m)dm \tag{14}$$

where now we have $s(m) = \beta e^{-\beta(m-M_0)}$. Then

$$\bar{n}(t) = \int_{M_0}^{\infty} \frac{Ke^{\alpha(m-M_0)}}{(t+c)^p} \beta e^{-\beta(m-M_0)} dm = \frac{K\beta}{(t+c)^p} \int_{M_0}^{\infty} e^{(\alpha-\beta)(m-M_0)} dm = \frac{\beta}{\beta-\alpha} \frac{K}{(t+c)^p} \tag{15}$$

In the ETAS spatio-temporal model, the conditional intensity of earthquakes is

$$\lambda(t, x, y \mid H_t) = \mu(x, y) + \sum_{j:t_j<t} \kappa\big(M_j\big)\, g\big(t - t_j\big) f\big(x - x_j, y - y_j \big| M_j\big) \tag{16}$$

where:

- $\mu(x, y)$ is the background intensity that is a function of latitude and longitude (does not depend on the time).
- $\kappa(M) = Ae^{\alpha(M-M_0)}$ is the expected number of events triggered by an event of magnitude M
- $g(t) = \frac{(p-1)c^{p-1}}{(t+c)^p}$ is the probability density function of the occurrence times of events triggered by previous events. [Utsu1969], [Yamanaka1990].

And the spatial probability density of earthquakes may have the following form: [Ogata2006] and [Hernandez2012]

$$f\big(x, y \big| M_j\big) = \frac{(q-1)d^{q-1}}{\pi} \{\frac{x^2+y^2}{e^{\alpha(M_j-M_0)}} + d\}^{-q} \tag{17}$$

The position-dependent intensity $\mu(x, y)$ is calculated by means of bi-cubic splines [Ogata1998], kernel functions [Zhuang2002] and [Helmstetter et al., 2007], grid average [Tsukakoshi and Shimazaki, 2006], tessellation (Ogata 2004) or by using non-parametric estimation by the Forward Likelihood Predictive method (FLP) [Adelfio2011](implemented in the R package etasFLP).

In 1998 Ogata proposed a modified version of the spatio-temporal ETAS model:

$$\lambda(t, x, y) = \mu + \sum_{j:t_j<t} g\big(t - t_j, x - x_j, y - y_j, M_j\big) \tag{21}$$

where $g(t, x, y, M) = \frac{Ke^{\alpha(M-M_0)}}{(t+c)^p} \{\frac{x^2+y^2}{e^{\gamma(M-M_0)}} + d\}^{-q}$ (22)

and the background intensity μ is constant. We can normalize $g(t, x, y, M)$ as

$$g(t, x, y, M) = \frac{K(p-1)c^{p-1}(q-1)d^{q-1}\alpha e^{(\alpha-\gamma)(M-M_0)}}{\pi(t+c)^p} \{\frac{x^2+y^2}{e^{\gamma(M-M_0)}} + d\}^{-q} \tag{23}$$

The previous formulas model the spatial correlation between earthquakes, conditioned on the magnitude of the main event, through the Euclidean distance between earthquakes: $D = x^2 + y^2$(isotropic models), however Ogata himself proposed anisotropic models where the clusters have elliptic forms

$$\lambda(t, x, y \mid H_t) = \mu + \sum_{j:t_j<t} g\left(t - t_j, x - x_j, y - y_j, M_j\right) \tag{24}$$

where now $g\left(t - t_j, x - x_j, y - y_j, M_j\right) = \frac{Ke^{\alpha(M_j - M_0)}}{(t+c)^p}\{\frac{pS_ip^T}{e^{\alpha(M_j - M_0)}} + d\}^{-q}$ (25)

and $p = \left(x - x_j, y - y_j\right)$is a row vector, x_j, y_j are the coordinates of the earthquake j, preceding the earthquake with epicenter x, y (both in the same cluster i) and S_i, $i = 1,2, \ldots$, are positive definite symmetric matrices representing the normalized covariance matrix of the earthquake cluster obtained by applying MBC or Magnitude Based Cluster algorithm.

This method is based on selecting the greatest magnitude earthquake(with magnitude $M_j$between those that are not in any cluster yet (if there are two with equal magnitude the oldest one is chosen) and then the earthquakes of the cluster associated with the previously selected earthquake, are those with latitude and longitude $\pm 3.33 * 10^{0.5M_j - 2}$ km(Utsu spatial distance)from the latitude and longitude of the selected earthquake and with a time difference therefrom(towards the future) of $\max\left(100, 10^{0.5M_j - 1}\right)$ days [Ogata1998]. Then the process is repeated with earthquakes that are not yet in any cluster until all earthquakes belong to a cluster.

The article [Guo2015] introduced a modification of the spatio-temporal ETAS model that includes the depths of earthquakes

$$\lambda(t, x, y, z \mid H_t) = \mu + \sum_{j:t_j<t} g\left(t - t_j, x - x_j, y - y_j, z - z_j, M_j\right) \tag{26}$$

where $g\left(t - t_j, x - x_j, y - y_j, z - z_j, M_j\right) = \frac{Ke^{\alpha\left(M_j - M_0\right)}}{(t+c)^p}\left\{\frac{pS_ip^T}{e^{\alpha\left(M_j - M_0\right)}} + d\right\}^{-q} h(z - z_j, z_j)$ (27)

and $h(z, z') = \frac{(\frac{z}{Z})^{\eta\frac{z'}{Z}}\left(1-\frac{z}{Z}\right)^{\eta\left(1-\frac{z'}{Z}\right)}}{ZB(\eta\frac{z'}{Z}+1, \eta\left(1-\frac{z'}{Z}\right)+1)}$ (28)

with Z: the thickness of the seismogenic layer and $B(p, q) = \int_0^1 t^{p-1}(1-t)^{q-1}$.

BAYESIAN STATISTICS AND HAMILTONIAN MARKOV CHAINS

Most studies of ETAS use point estimates for the model parameters, which ignores the inherent uncertainty that arises from estimating these from historical earthquake catalogs, resulting in misleadingly optimistic forecasts. In contrast, Bayesian statistics allows parameter uncertainty to be explicitly represented and fed into the forecast distribution. [Gordon2016].
Besides, Bayesian statistics can consider the uncertainty not only in the model parameters, conditioned on the available catalog of events occurred before the forecasting interval but also the uncertainty in the sequence of events that are going to happen during the forecasting interval as shown in [Ebrahimian2017].

Bayesian statistics does not work with only a single point estimate of a parameter but instead, consider the whole posterior distribution (given the data) which represents the uncertainty based on both the observed earthquake catalog and any prior knowledge we have based on previous studies [Gordon2016]. The posterior distribution in the ETAS model is highly complex and does not allow us to find the mean or the maximum of the posterior for the parameters analytically. Instead, we can use a Markov Chain Monte Carlo (MCMC) algorithm [Rasmussen2011]. The Metropolis-Hastings algorithm updates one parameter at a time and uses truncated normal distributions as proposal distributions for the evolution of the Markov chains.

The key issues in setting up a prior distribution are:
-what information is going into the prior distribution and
-the properties of the resulting posterior distribution.

When there is no previous knowledge, researchers choose noninformative prior distributions thus allowing each parameter to be estimated from the data. But, if noninformative prior distributions were assigned to all the parameters, then the model would fit the data very closely but with scientifically unreasonable parameters [Gelman2002] and [Gelman2013]. Thus we prefer to use flexible and weakly informative priors (heavy-tailed distributions). Although the prior selection sensitivity is an important aspect of the Bayesian analysis. [Mueller2012]. We used exponential prior distributions for the positive parameters of ETAS models. The exponential distribution is one of the less informative distributions for positive parameters.

MCMC or Monte Carlo Markov Chains is a general class of methods for extracting values from the posterior distributions of the parameters. [Gordon2016]

If θ is the last accepted value of the parameter, a new value θ^* is proposed from a jump distribution (proposal distribution) $q(.|.)$ that is $\theta^* = q(.|\theta)$ and then we calculate

$$r = \frac{L(\theta^*)p(\theta^*)q(\theta^*|\theta)}{L(\theta)p(\theta)q(\theta|\theta^*)} \tag{37}$$

where L is the likelihood function, p is the a priori distribution.

For the case of the Metropolis algorithm, the jump distribution is symmetric. Then a value u is generated from a uniform distribution and if $u < \min(1, r)$we accept the value θ^* and if not, it is rejected. [Barrows2016]

The Metropolis-Hastings algorithm may not efficiently explore the parameters space. In fact, when you don't have conjugate priors it is hard to optimize acceptance rates and proposals.

Hamiltonian Monte Carlo, also called Hybrid Monte Carlo (HMC), is a Monte Carlo Markov Chain (MCMC) method to approximate integrals with respect to a target probability distribution π on $\mathbb{R}^d$. It was originally proposed by Duane et al. [1987], it was later introduced in statistics by Neal [1993] and is now part of the standard toolbox [Brooks et al., 2011, Lelièvre et al., 2010], in part due to favorable scaling properties with respect to the dimension d [Beskos et al., 2010, 2013],

compared to random walk Metropolis-Hastings. Hamiltonian Monte Carlo is at the core of the No-U-Turn sampler (NUTS, [Hoffman2014]) used in the software Stan [Carpenter2016].

In the HMC algorithm, the parameter estimates are treated as physical particles that move on the surface of the likelihood. The potential energy is analogous to minus the logarithm of the likelihood $U(\theta) = -\log(L(\theta)p(\theta))$. The kinetic energy is $K(r) = \frac{1}{2}r^T M^{-1} r$ where $r = (r_1, r_2, \dots, r_n)$with n the number of components of θ and M_{nxn} a square matrix that can be the identity but can also be the correlation of the components of θ. The Hamiltonian of the system is $H(\theta, r) = U(\theta) + K(r)$ and the dynamics of the system is given by

$$\frac{d\theta}{dt} = M^{-1} r \tag{38}$$

$$\frac{dr}{dt} = -\nabla U(\theta) \tag{39}$$

The Euler method for the above system causes instability and Leapfrog method is used instead where step lengths ϵ and $\epsilon/2$ are used for the evolution of $\frac{dr}{dt}$ in the first and last step respectively. The final values of the Leapfrog steps are the proposed values $(\theta^*.r^*)$and are accepted or rejected similarly to the Metropolis, but now

$$r = \exp\left[H(\theta, r) - H(\theta^*.r^*)\right] \tag{40}$$

This comes from the definition of the Hamiltonian as an energy function. The distribution of the total potential energy as a function of the Hamiltonian is

$$P(\theta, r) = \frac{1}{Z}\exp\left(-H(\theta, r)\right) \tag{41}$$

where Z is a normalization constant. Then ([Barrows2016])

$$r = \frac{P(\theta^*.r^*)}{P(\theta,r)} = \exp\left[H(\theta, r) - H(\theta^*.r^*)\right] \tag{42}$$

The advantages of HMC are:

- It can produce high dimensional proposals that are accepted with high probability without having to spend time tuning.
- Has inbuilt diagnostics to analyze the MCMC output.
- Built in C++ so runs quickly but outputs to R.

Although, HMC requires some tuning: the number and size of the leapfrog steps, the "No-U-Turn Sampler" or NUTs (Hoffman and Gelman (2014)), optimizes these adaptively.

PARAMETER ESTIMATION

The parameters of the ETAS model are generally estimated by maximizing the logarithm of the likelihood, for temporal models:

$$logL(\mu, K, p, c) = \sum_j \log\lambda\left(t_j \mid H_t\right) + \int_0^{T_{max}} \lambda(t \mid H_t)\mathrm{dt} \tag{43}$$

And for spatio-temporal models

$$logL(\mu, K, p, c, d, \alpha, q) = \sum_j \log\lambda\left(t_j, x_j, y_j \middle| H_t\right) + \int_0^{T_{max}} \iint_S \lambda(t, x, y \mid H_t)\mathrm{dxdydt} \tag{44}$$

where S is the region that contains the earthquake data, which is commonly the rectangle $[long_{min}, long_{max}] \times [lat_{min}, lat_{max}]$
For the models (4) and (7), the logarithms of the likelihood have respectively the following analytic forms:

$$logL(\mu, K, p, c) = \sum_{i=1}^{N} \log\left(\lambda\left(t_j \mid H_t\right)\right) - \mu T_{max} - K \sum_{i=1}^{N} \left(1 - \frac{c^{p-1}}{(T_{max} - t_i + c)^{p-1}}\right) \quad (45)$$

$$logL(\mu, K, p, c, A, \alpha) = \sum_{i=1}^{N} \log\left(\lambda\left(t_j \mid H_t\right)\right) - \mu T_{max} - KA \sum_{i=1}^{N} e^{\alpha(M_i - M_o)} \left(1 - \frac{c^{p-1}}{(T_{max} - t_i + c)^{p-1}}\right) \quad (46)$$

For the case of the spatio-temporal ETAS models, the logarithm of the likelihood does not have a closed analytical form. As mentioned in [Schoenberg2013], the first sum in the logarithm of the likelihood formula can be easily calculated but the integral in general is difficult to approximate. The numerical approximations of this integral can be computationally expensive since the function has local maximums in 3 dimensions and this integral has to be calculated for each tuple of parameters during the optimization. [Schoenberg2013]. Schoenberg mentions that the approximations of this integral, work poorly with optimization routines. In [Ogata1998] the space around each earthquake is divided into *K* quadrants and the integral in each quadrant is calculated, however, this is computationally costly and the choice of the number *K* is problematic. [Schoenberg2013].

In [Schoenberg2013], the author proposed for the spatio-temporal model the approximation:

$$logL(\mu, K, p, c, d, \alpha, q) = \sum_{i=1}^{N} \log\left(\left(t_j \mid H_t\right)\right) - \mu T_{max} - K \sum_{i=1}^{N} e^{\alpha(M_i - M_o)} \quad (47)$$

This approximation assumes that the time $T_{max} \to \infty$and the size of the observed area of the earthquakes tends to infinity. The approximation is accurate if for each earthquake, the region of the aftershocks caused by this earthquake is contained completely within the observed area[Schoenberg2013], and has the property that for any *j* and any values of the parameters of the ETAS model

$$Ke^{\alpha(M_j - M_o)} = \iiint_{\mathbb{R}^3} g(t - t_j, x - x_j, y - y_j)\, dtdxdy \quad (48)$$

That is, if certain values of the parameters are poor then $K \sum_{i=1}^{N} e^{\alpha(M_i - M_o)}$ will be larger than $\iiint_{\mathbb{R}^3} g(t - t_j, x - x_j, y - y_j)\, dtdxdy$ and the optimization routine will avoid these parameter values.[Schoenberg2013]

However, this approach is analyzed in [Lippiello2014] and it is shown that the assumption that $T_{max} \to \infty$is too crude, which causes the introduction of systematic biases. [Lippiello2014].

Because the earthquakes analyzed in the provinces of Manabí and Esmeraldas are in a reduced area (compared to the radius of the earth), we can neglect the curvature of the earth in that area and find the minimum circle that covers all the earthquakes. Then we can approximate the double integral of the log of the likelihood. This paper explores the possibility of calculating the integral in the spatio-temporal model by making the region *S* the minimum covering circle containing the

positions of earthquakes. This is done by the Welzl algorithm [Welzl991] and then we can approximate the integral by polar coordinates centered at the position of each earthquake with:

$$logL(\mu,k,p,c,d,\alpha,q) \approx \sum_{i=1}^{N} \log\left(\lambda\left(t_j \mid H_t\right)\right) - \mu T_{max}\pi r^2 - K\alpha \sum_{i=1}^{N} e^{\alpha(M_i-M_o)}\left(1-\frac{c^{p-1}}{(T_{max}-t_i+c)^{p-1}}\right)\left(1-\frac{d^{q-1}}{\left(\frac{r_i^2}{e^{\alpha(M_i-M_o)}}+d\right)^{q-1}}\right) \quad (49)$$

where r is the radius of the minimum circle that contains all the earthquakes and r_i the greatest distance between the coordinates of the earthquake i and the previous earthquakes according to the metric defined in the cluster to which they belong. (In general, when we solve the double integral over the area of earthquakes in polar coordinates centered at the i earthquake, the distance ρ depends on the angle θ. Instead, we approximate the maximum ρ, in any direction, as the maximum distance between the earthquake i and the other earthquakes).

In this approximation, it is not assumed that $T_{max} \to \infty$ because the integral with respect to the time goes from 0 to T_{max}.

Welzl's algorithm is based on the fact that the minimum covering circle containing the points must contain at most 3 at its boundary (only two points if they are on a diameter). This algorithm is randomized and incremental and at each step maintains the minimum covering circle and then adds a point: if the point is within the current minimum circle, the circle it is not updated, but if it is outside, then the new minimum covering circle passes through the new point. This algorithm has an expected order $O(n)$.
For the anisotropic model, the Euclidean distance is changed by the metric defined by the standardized variance-covariance matrix S_i where now r_i is calculated as the maximum distance (according to the previous metric) and the rest of the previous earthquakes in the cluster.
For the hypocentral ETAS model, the logarithm of the likelihood is

$$logL(\mu,k,p,c,d,\alpha,q,Z) = \sum_j \log\lambda\left(t_j, x_j, y_j, z_j \middle| H_t\right) + \int_0^{T_{max}} \iint_S \int_0^Z \lambda(t,x,y,z \mid H_t)\mathrm{dzdxdydt}$$

And, in this case, we use the same approximation proposed for the double integral over S. To be able to compare the hypocentral model with the previous models we use the correction of the logarithm of the likelihood of the hypocentral model [Guo2015]:

$$logL_{3d} = logL_{2d} - Nlog(Z) \quad (50)$$

For the case of formulas (45) and (46) with variable background seismicity (instead of μ we have $\mu * \mu(x,y)$).
The logarithm of the likelihood is

$$logL(\mu,K,p,c) = \sum_{i=1}^{N} \log\left(\lambda\left(t_j\right)\right) - \mu T_{max} \sum_{i=1}^{N} \mu(x,y) - K\sum_{i=1}^{N}\left(1-\frac{c^{p-1}}{(T_{max}-t_i+c)^{p-1}}\right) \quad (51)$$

$$logL(\mu,K,p,c,A,\alpha) = \sum_{i=1}^{N} \log\left(\lambda\left(t_j\right)\right) - \mu T_{max} \sum_{i=1}^{N} \mu(x,y) - KA\sum_{i=1}^{N} e^{\alpha(M_i-M_o)}\left(1-\frac{c^{p-1}}{(T_{max}-t_i+c)^{p-1}}\right) \quad (52)$$

And similarly, for the temporal and hypocentral space models.

The uncertainties of the ETAS model come from two main sources: the model itself (for example in the temporal model, only the times of occurrence of the aftershocks are considered, so the model itself has uncertainties) and on the other hand, the error that comes from the approximation of the likelihood or the ETAS models (for example the use of Weyl algorithm introduces an error). Employing MCMC and HMC methods we robustly account for the various source of noisy sources of noise in the data.

We obtained estimates of the parameters of the ETAS models using FLP and compared them with the estimates obtained using MCMC, in this case, HMC which is implemented in the stan language. Also, for the case of the temporal models, we compared the results obtained with Rstan with those obtained by the program ETAS.exe of Ogata available in [Stasei4] that minimizes $-logL$ by the method of Davidon-Fletcher-Powell [Fletcher1963].

PREPROCESSING

To achieve efficiency, a pre-processing was performed and we don't pass the values of the times, magnitudes and latitudes of earthquakes directly to the Rstan library (as is done for the temporal model in Stackoverflow {Stackoverflow1} and {Stackoverflow2} because then, in each iteration, it would be necessary to calculate for each earthquake, the differences of time, latitude, and latitude, with all the previous earthquakes($O(n^2)$). Instead, we sorted the data by time descending and these differences were calculated in three arrays of size $\frac{N(N-1)}{2}$. Besides, for each earthquake j, we know the initial and final positions where the differences of its time, latitude and longitude with respect to the previous earthquakes are:

$start = N(j-1) - \frac{j(j-1)}{2} + 1$

$end = jN - \frac{j(j+1)}{2}$

For the case of the anisotropic cluster model, the earthquake data of each cluster were adjusted with four normal bivariate models([Ogata1998], [Ogata2006], [Ogata2011]):

$N(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} \tilde{\sigma}^2 & 0 \\ 0 & \tilde{\sigma}^2 \end{pmatrix})$ (53) $N(\begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}, \begin{pmatrix} \hat{\sigma}^2 & 0 \\ 0 & \hat{\sigma}^2 \end{pmatrix})$ (54)

$N(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} {\tilde{\sigma}_1}^2 & \tilde{\rho}\tilde{\sigma}_1\tilde{\sigma}_2 \\ \tilde{\rho}\tilde{\sigma}_1\tilde{\sigma}_2 & {\tilde{\sigma}_2}^2 \end{pmatrix})$ (55) $N(\begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}, \begin{pmatrix} \hat{\sigma}_1^2 & \hat{\rho}\hat{\sigma}_1\hat{\sigma}_2 \\ \hat{\rho}\hat{\sigma}_1\hat{\sigma}_2 & \hat{\sigma}_1^2 \end{pmatrix})$ (56)

where (x_1, y_1) are the latitude and longitude of the cluster's main earthquake, $(\bar{x}\ ,\ \bar{y}\ )$are the coordinates of the centroid of the cluster and the parameters correspond to[Ogata1998]:

$\tilde{\sigma}^2 = \left[\sum_j (x_j - x_1)^2 + \sum_j (y_j - y_1)^2\right]/(2n)$ (57)

$\hat{\sigma}^2 = \left[\sum_j (x_j - \bar{x})^2 + \sum_j (y_j - \bar{y})^2\right]/(2n)$ (58)

$\tilde{\rho} = [\sum_j (x_j - x_1)(y_j - y_1)]/(n\tilde{\sigma}_1\tilde{\sigma}_2)$ (59)

$\hat{\rho} = [\sum_j (x_j - \bar{x})(y_j - \bar{y})]/(n\hat{\sigma}_1\hat{\sigma}_2)$ (60)

${\tilde{\sigma}_1}^2 = [\sum_j (x_j - x_1)^2]/n$ (61)

${\tilde{\sigma}_2}^2 = [\sum_j (y_j - y_1)^2]/n$ (62)

${\hat{\sigma}_1}^2 = [\sum_j (x_j - \bar{x})^2]/n$ (63)

${\hat{\sigma}_2}^2 = [\sum_j (y_j - \bar{y})^2]/n$ (64)

and we selected the model with the lowest $AIC = -nln(\det(S)) + 2k$ where S is the variance-covariance matrix of each of the four models and \$k\$ is the corresponding number of parameters[Ogata1998]. For the case that the number of earthquakes in the cluster is less than 6, the smaller AIC is selected from the first two models. [Ogata1998]. Then the selected variance-covariance matrix is normalized:

$$\left(\frac{1}{\sqrt{1-\rho^2}}\right)\begin{pmatrix} \frac{\sigma_2}{\sigma_1} & -\rho \\ -\rho & \frac{\sigma_1}{\sigma_2} \end{pmatrix} \quad (65)$$

In addition, as part of the pre-processing, for each cluster, all the distances from each earthquake to each of the previous one in the same cluster, for both the Euclidean distance and the anisotropic metric, were calculated.

Once we have the parameters of the ETAS models, we can estimate the probability that a given event is spontaneous or is triggered by others [Kagan1980] [Zhuang2002]. The contribution of the spontaneous seismicity rate to the occurrence of an event i can be taken as the probability that the event i is spontaneous [Zhuang2008]:

$$\phi(i) = \frac{\mu(x_i,y_i)}{\lambda(t_i,x_i,y_i)} \quad (66)$$

Similarly, the probability that the event j is produced by the event $i\ is$: [Zhuang2008]

$$\rho_{ij} = \frac{\kappa(m)g(t_j-t_i)f(x_j-x_i,y_j-y_i,m_i)}{\lambda(t_i,x_i,y_i)} \quad (67)$$

We can obtain the expected number of direct aftershocks from the earthquake $i\ as\ \sum_j \rho_{ij}$.

We can also verify the assumption that the background seismic intensity $\mu(x,y)$ is stationary, i.e it does not depend on the time calculating for each t: [Zhuang2004]

$S(t) = \sum_{j:t_j<t} \rho_i$

RESIDUAL ANALYSIS AND GOODNESS TO FIT

Analysis of residuals and goodness of fit of the models can be done as mentioned in [Zhuang2004], but in this case, we take samples of the posterior distributions of the parameters instead of just use single values of them.

We can evaluate the seismic intensity as a function on the time, magnitude, latitude, longitude, and depth of each earthquake (according to the corresponding ETAS model) and compare the accumulated curve obtained from the model with the observed one. In this case, since we have the values of the parameters obtained from the posterior distribution, we can do simulations and for each tuple of parameters, evaluate the seismic intensity formula and then we calculate the median of these values and the 2.5% and 97.5% percentiles and 0.975 (95 % credibility interval).

To analyze the differences between the predictions of the model and the observed data, we graph the transformed time: $\lambda(t)$ in the x-axis, and in the y-axis the observed values.

TIME BETWEEN EVENTS

The relationship between inter-event times and ETAS models was analyzed analytically by[Saichev2007]. In [Saichev2006], it is analyzed the existence of a universal law of escalation [Corral2004] for the probability density function of the recurrence times or times between earthquakes (time between two successive events) τ

$$H(\tau) \approx \lambda f(\lambda\tau) \quad (68)$$

where the function $f(x)$ has been found practically the same in different regions [Saichev2007] and λ the average rate of events observed in the analyzed region. [Saichev2007].

The scaling factor of times between earthquakes is taken as the inverse of its mean.

The form of the function $f(x)$ which is demonstrated in [Saichev2007] is

$$f(x) = \left(n\epsilon^{\theta}x^{-1-\theta} + \left[1 - n + n\epsilon^{\theta}x^{-\theta}\right]^2\right) * \varphi(x,\epsilon) \quad (69)$$

where θ is the parameter of the Omori-Utsu or Modified Omori law.

$$\Phi(\mathrm{t}) = \frac{\theta c^{\theta}}{(c+t)^{1+\theta}} \quad , x = \lambda\tau, \epsilon = \lambda c \quad (70)$$

is the rate of triggered events of first from a given earthquake (average number of direct descendants per earthquake).

Also

$$P(\tau) \approx \varphi(x,\epsilon) = \exp\left(-(1-n)x - \frac{n\epsilon^{\theta}}{1-\theta}x^{1-\theta}\right) \quad (71)$$

is the probability that there are no events in $[t, t+\tau]$.

DATA ANALYSIS AND RESULTS

The data used for this study is available in http://www.igepn.edu.ec/ultimos-sismos and the Stan code for the ETAS models can be found in https://github.com/crespofabian8012/StanCon2018.

Table 1 shows that, out of the 908 earthquakes in Ecuador from March 18th-July16th, 2016, 810 had their epicenter closest to the provinces of Manabí and Esmeraldas. On the other hand, the cities with the larger number of near earthquake epicenters were: Jama 241 earthquakes; Muisne 222 earthquakes; Pedernales 77 earthquakes; Manta 66 earthquakes; Puerto López 64 earthquakes and Bahía de Caráquez 49 earthquakes. **Table 1**

Closest province	March	April	May	June	July	Total
Esmeraldas		171	47	16	35	269
Manabi	1	399	92	27	22	541
Total Ecuador	10	624	165	51	58	908

Table 1 shows the monthly average of the earthquake's magnitude in the provinces Manabí and Esmeraldas and the average considering both provinces per month. Here we see that the average magnitude in April (when the earthquake of magnitude 7.4 occurred) is not the greater value because the number of earthquakes in that month was larger than in others months, and then an earthquake of greater magnitude in May, June or July increase the month average of that month more than the corresponding increase in other months.

The average depth of the earthquakes of Manabí and Esmeraldas was 8.44 km and 8.07 km respectively. **Table 2.**

Province	March	Abril	May	junio	July	Total
Esmeraldas		3.76	4.05	3.59	3.85	3.82
Manabi	3.40	3.87	3.68	3.75	4.12	3.84
Total	3.53	3.83	3.80	3.74	3.95	3.83

In figure 1 we have the distribution of earthquakes in Ecuador from March to July, 2016 by depth: in green the earthquakes with depths greater than 30 km and in red, earthquakes with depths less than 30 km. We can see that the earthquakes in the provinces of Manabí and Esmeraldas (coast of

Ecuador), that are the ones analyzed in this article, occurred with depths less than 30 km. So the thickness of the seismogenic layer is $Z = 30$ km for the hypocentral ETAS model.

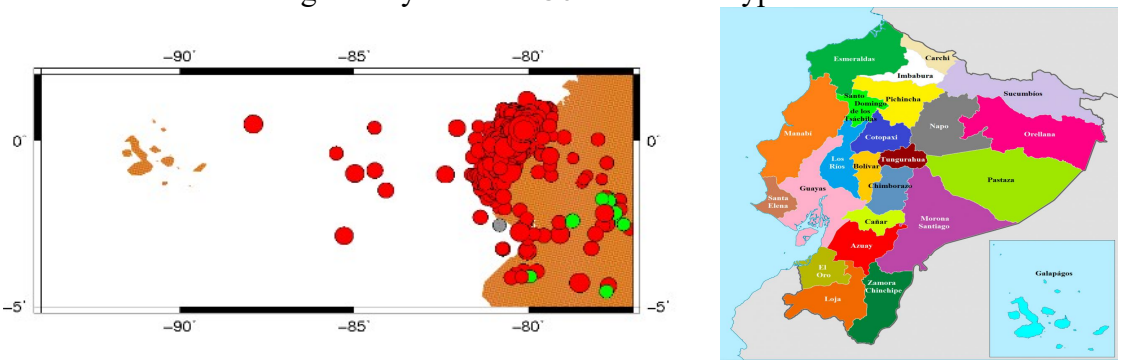


FIGURE 1. Earthquake distribution by depth: The red points are events with depth less than 30 km and the green points are events with a depth greater than 30 km.

Figure 2 shows the distribution of earthquakes with circles proportional to their magnitude, in red the most recent earthquakes and in blue the older ones.

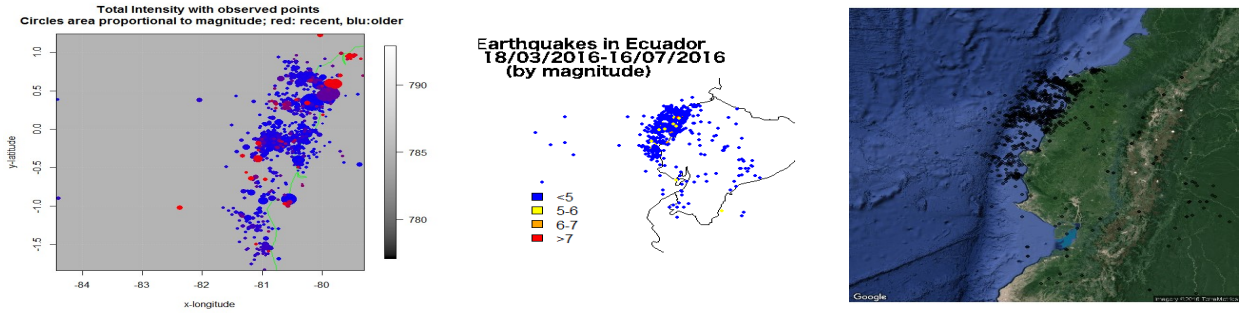


FIGURE 2 and 3. Earthquake distribution by magnitude and time. FIGURE 4. Earthquake distribution(Google Maps).

In figure 3, the earthquakes are shown by magnitude intervals and we can observe that the number of earthquakes of greater magnitude is smaller which agrees with the law of Gutenberg-Richter. The figure 4 on the right shows the distribution of the earthquakes with respect to the position of the nearest tectonic plates.

The following graph (figure 5) shows the cumulative number of earthquakes over time and also vertical segments for each event (the segment length is proportional to the magnitude). It is observed that the magnitudes have been decreasing after April 16th, although there have been earthquakes of magnitude more than 6 degrees Richter. The abrupt change in the curve of the cumulative number of events after the earthquake of 7.4 on April 16th, 2016, is characteristic of big events. In figure 6 we can see how the average daily earthquake rate has been decreasing after reaching a daily average rate of more than 150 earthquakes (immediately after the earthquake of April 16th).

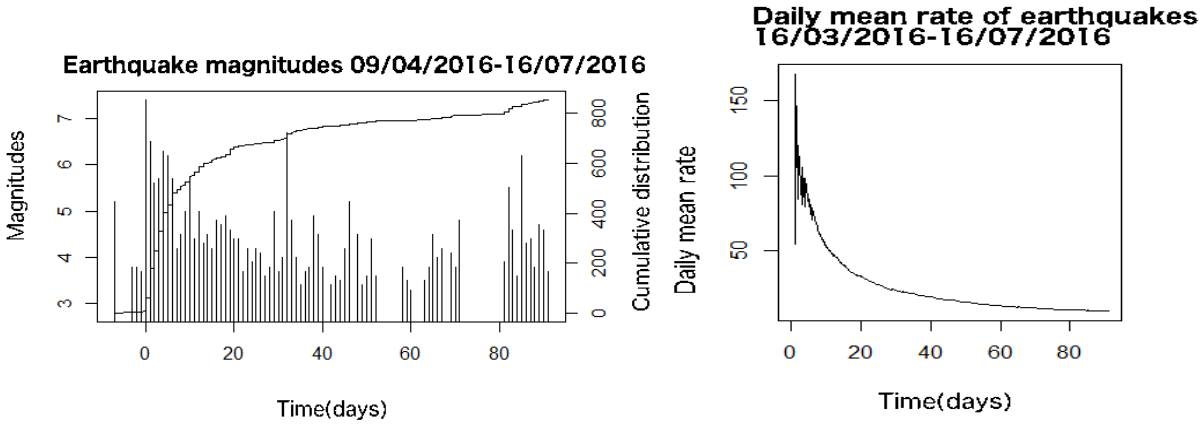


FIGURE 5 and(6). The cumulative number of earthquakes and magnitude of events over time. (6) Average daily rate of earthquakes over time.

To verify the Gutenberg-Richter law, we plot in figure 7 the number of events with magnitudes greater than a given threshold magnitude vs the threshold magnitudes and in figure 8, we plot the logarithm in base 10 of the number of events with magnitude greater than a given threshold magnitude vs the threshold magnitudes. Then a linear regression model log(N)= a + bM was fitted and the values of a = 5.064(sd = 0.039) and b=-0.694 (sd = 0.007) were obtained with a p-value for both values of 2.2e-16. The standard residual error was 0.1041 and the adjusted R^2 was 0.9888. With these values, the b in Gutenberg-Richter's law is 0.694 and $\beta = \text{bln}\,(10) = 1.598$.

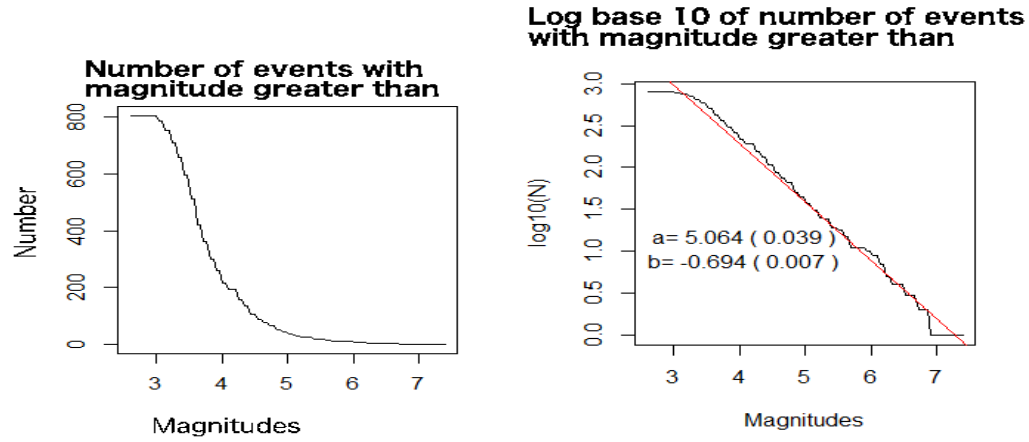


FIGURE 8. The decimal logarithm of the number of events with magnitude greater than or equal to a given threshold magnitude as a function of that threshold magnitude.

By adjusting the earthquake catalog in Manabí and Esmeraldas provinces using the Etas R package with 10 iterations, we obtained estimates of the background seismicity. Figure 9 shows the distribution of background seismicity in the two provinces of the Ecuadorian coast. The parameters of the ETAS model in the etas package (Zhuang Etas model that is combined with the stochastic declustering method) obtained by maximizing the likelihood were $\mu = 1.0227$, $A = 0.4789$, $c = 0.0434$, $\alpha = 0.6947$, $p = 1.4058$, $d = 0.0054$, $q = 1.7413$, $\gamma = 0.1783$, the logarithm of the likelihood $ln(L) = 1414.451$ and the Akaike Information Criteria $AIC = -2ln(L) + 2k = -2 * 1414.451 + 2 * 8 = -2812.902$. The results of adjusting the Flp:

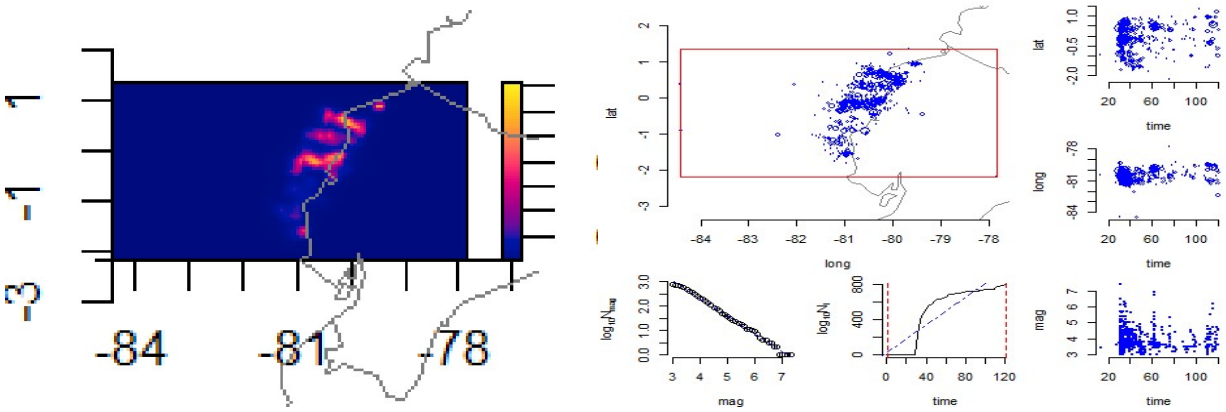


FIGURE 9 and 10. Stochastic declustering result in the etas package. Distribution of magnitudes, latitudes, and longitudes of earthquakes over time using EtasFlp package.

Forward Likelihood Predictive model using the R package EtasFlp, are shown in Figures 12 to 15. Figure 12 shows the background seismic intensity and the areas within the provinces of Manabí and Esmeraldas where it is greater. Figure 13 shows the seismicity induced or triggered by previous earthquakes. In graphs 16 and 17, we can observe that there are areas where the standardized residuals (which have mean 0 and variance 1) between theoretical and observed, are considerable.

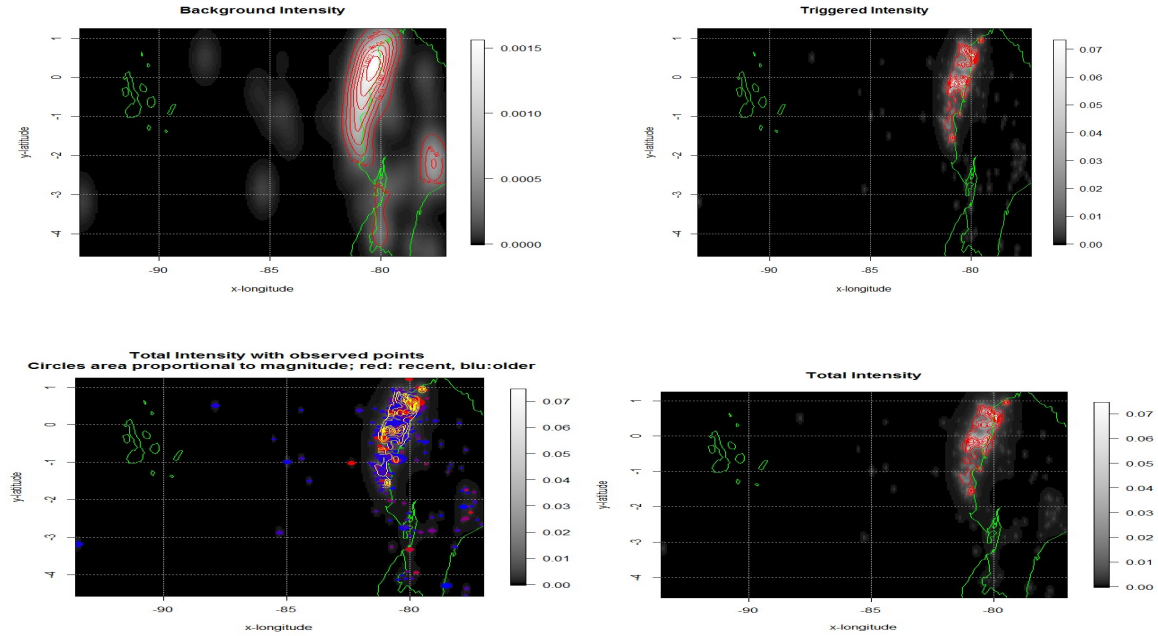
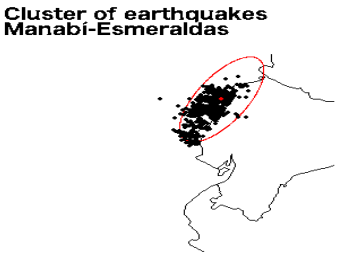


FIGURE 12-15. (12) Background seismic intensity. (13) Triggered seismic intensity. (14) The total seismic intensity with observed events. (15) Total intensity.

We used non-parametric models such as FLP but we preferred Bayesian statistics and HMC because it gave us the parameter posterior distributions. With this posterior distributions, we carried out simulations to obtain 95 % credibility intervals.

Figures 18 and 19 show the anisotropic distribution of earthquakes. This is achieved by adjusting the four bivariate models to the 810 earthquakes with epicenters near Manabí and Esmeraldas and choosing the lowest AIC and then the axes of the ellipse are the eigenvectors of the normalized variance-covariance matrix.



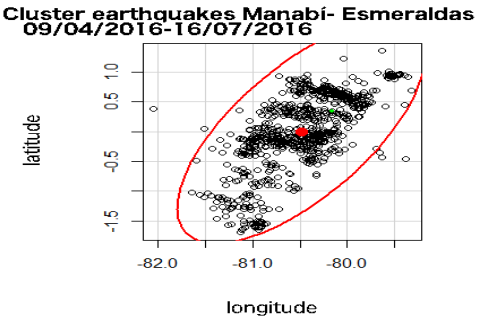


FIGURE 16 and 17. Cluster associated with the big earthquake of April 16th, 2016.

The normalized variance-covariance matrix for the 804 earthquake cluster related to the main earthquake of April 16(magnitude 7.4) is

$$\begin{pmatrix} 1.469 & -0.696 \\ -0.696 & 0.709 \end{pmatrix})$$

and corresponds to the fourth model of the adjusted bivariate models.

For all parameters of the ETAS models (which must be positive), a weakly a priori exponential distributions were used. For p and q, a value close to and greater than 1: 1.000005 was used as their minimum value. For the initial value of mu in the temporal and hypocentral space models, $mu = 0.25\, N / (Tmax * \pi * r^2)$ [Ogata1998] and [Ogata1998b] where N is the number of events, $Tmax$ is the maximum time of the events(in days), measured from the oldest event, and r is the radius of the minimum covering circle of earthquake epicenters close to the provinces of Manabí and Esmeraldas, r = 1.70479.

For the other parameters, the initial values used were as follows $k = 0.3$, $p = 1.3$, $c = 0.01$, $d = 0.01$, $q = 1.7$, $\alpha = 1.005$, $\gamma = 0.9$, $\eta = 0.7$.

In all cases, we used 4 chains with 100 iterations for each one. We use the Rhat value which is a way to measure the convergence of the chains: when it is close to 1 indicates convergence. Figures 19-22 show the evolution of the chains for the anisotropic with constant background seismicity.

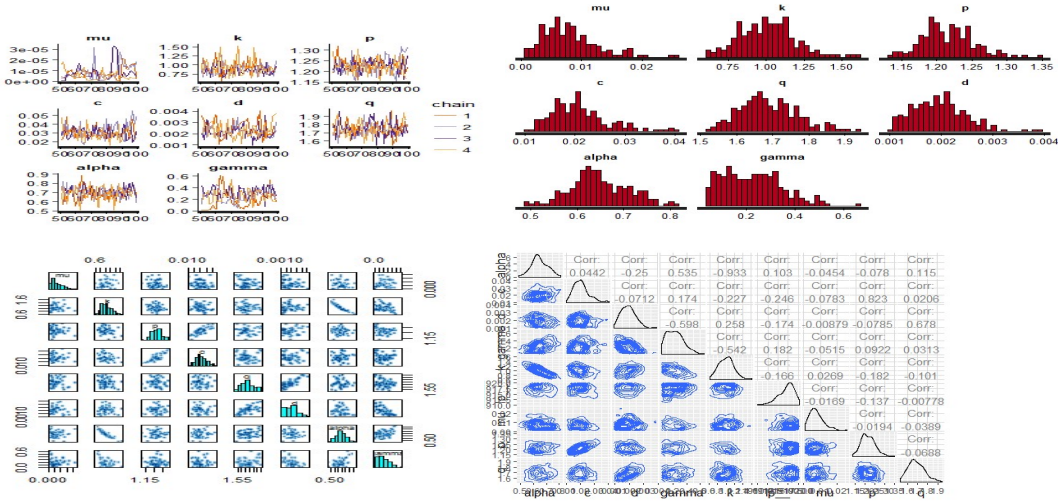


FIGURE 19 -22. Chains evolution for anisotropic etas model with a constant background seismicity. (19) Evolution of 4 chains for each parameter. (20) A posteriori parameter distributions. (21) Correlation of parameter values in the chains. (22) Correlation of parameter values in the chains.

From the following tables, we observe that the models with variable background seismicity (where μ is replaced with $\mu * \mu(x, y)$) have a value of the constant μ considerably lower than the models

with constant background seismicity. This is because, for the variable background seismicity, μ multiplied by the sum of the values of $\mu(x, y)$ over the event points (x, y)must be approximately equal to μ for the constant background seismicity.
As shown in Table 12, the model that best fits the data is the anisotropic hypocentral model with constant background seismicity. We can also observe that in our case, the models with variable seismicity do not provide better fits than the analogous models with constant background seismicity, although the methods of stochastic and Flp disaggregation showed that the background seismicity is not constant.

Table 12: Etas models comparison

Etas Model	ln(L)	Int. 95% ln(L)		# param.	AIC	int.95% AIC	
Temp. sis. cst	1790.7	1787.3	1792.8	4	-3573.5	-3577.6	-3566.5
Temp. seis. var.	1773.3	1769.7	1775.0	4	-3538.6	-3542.0	-3531.4
Magn. seis. cst	1832.8	1689.2	1842.8	6	-3653.7	-3673.7	-3366.5
Magn. seis. var.	1800.8	1797.4	1802.9	6	-3589.5	-3593.8	-3582.7
Spa. temp. iso. seis. cst	1394.4	1387.4	1398.1	8	-2772.7	-2780.1	-2758.7
Spa. temp. ani. seis. cst	1857.1	1852.7	1859.6	8	-3698.3	-3703.2	-3689.4
Spa. temp. ani. seis. var.	1799.9	1794.6	1801.9	8	-3583.8	-3936.8	-3573.3
Hypo. ani. seis. cst.	1974.1	1969.4	1977.0	9	-3930.2	-3936.1	-3920.8
Hypo. ani. seis. var.	1923.3.1	1919.4	1926.1	9	-3828.6	-3834.2	-3821.4

The possible cause of this is that the area is small, (the radius of the minimum covering circle is approximately 1.70479 degrees) which makes that the variable background seismicity can be replaced by a constant background seismicity (although in the case of models with constant background seismicity, the values of the other parameters may be biased). However, the model fits better the data when we consider the metric defined by the variance-covariance matrix, to measure the distances between the epicenters of the earthquakes (the anisotropic model) and also consider the depths of the earthquakes. We have that the logarithm of the likelihood equivalent in 2 dimensions to the hypocentral model with constant background seismicity is

$$\ln(L_{2d}) = \ln(L_{3d})_{Nln(Z)} = -709.462 + 789 * \ln(30) = 1974.083$$

$$\text{AIC} = -2 * 1974.083 + 2 * 9 = -3930.165$$

(there were 789 earthquakes in the cluster associated with the earthquake of magnitude 7.4 on April 16 with valid depth values) and similarly for the hypocentral model with variable background seismicity (in this case we had $N = 777$ available background seismicity values).
Using the values obtained from the posterior distributions of the parameters of the best fit model: anisotropic hypocentral with constant background seismicity, we can neglect the values of the first half of the evolution of the chains and perform simulations by extracting values (from the second half) of the parameters and estimate the probability that an earthquake is induced (not caused by background seismicity). Figure 23 shows the probability of being an induced earthquake as a function of time and in red the events with probability greater than 0.95 of being an induced event. We can observe that the events on April 16th have a high probability of being triggered (not from background seismicity). Equally, after 80 days since 04/13/2016, there are still events with a high probability of being induced by previous earthquakes, especially due to the earthquake of magnitude 7.4 on April 16th.

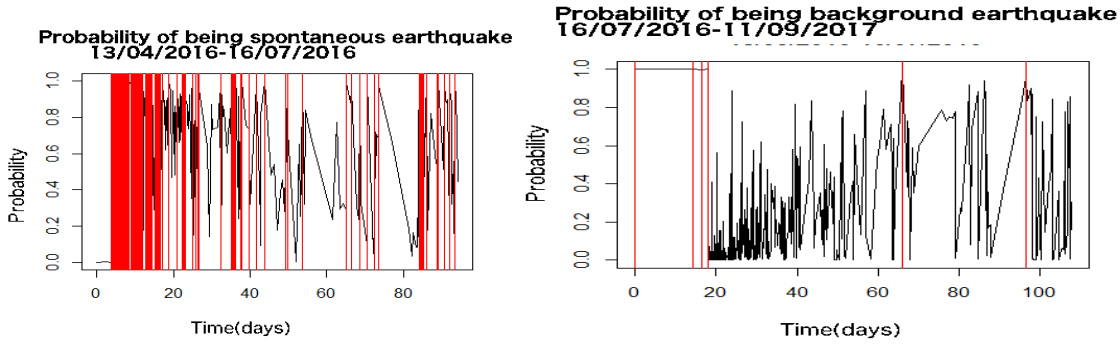


FIGURE 23 and 24. (23)The probability of being a spontaneous earthquake.(24) The probability of being background earthquake.

Figure 24 shows the probability of being a background earthquake as a function of time, and in red the events with probability greater than 0.95 of being a background seismic event. We can see that very few events after April 16th, 2016, have a high probability of being triggered by background seismicity, although such events exist.

Figure 25 shows the induced seismicity rate versus time after the date of the last event used to adjust the models: 07/16/2016, this is achieved by performing simulations of the anisotropic hypocentral model with constant background seismicity, and evaluating for each tuple of parameters the seismicity rate function at the times of the events from 07/16/2016 to 09/11/2016 and then taking the median by all the tuples of parameters. It is observed that the induced rate has been decreasing and is close to extinction.

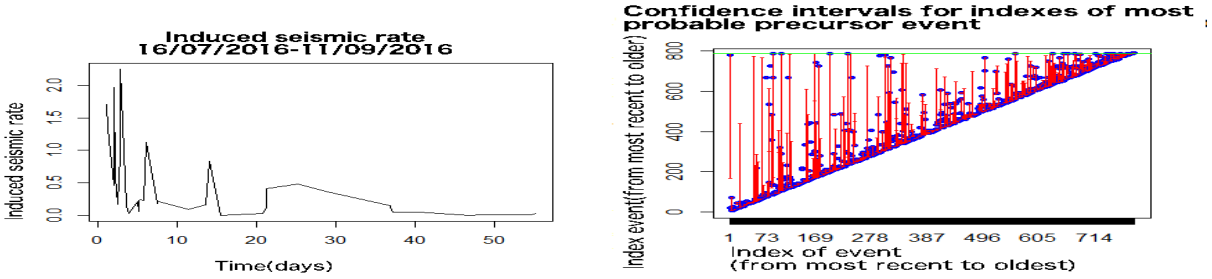


FIGURE 25. Induced seismic rate. FIGURE 26. Indexes earthquakes more likely predecessors(fathers).

Figure 26 shows the confidence intervals (or credibility in this case since we use Bayesian approach) of the most probable father earthquake of each earthquake. On the x-axis, we have the indexes of earthquakes in the catalog, from the most recent to the oldest, i.e from 07/16/2017 to the oldest recorded earthquake available in the month of March, and the vertical segments are the credibility intervals of the index of the most probable father earthquake. For this, we performed simulations of the parameters, and for each tuple we calculate for each event j (index on x-axis), the probability that it will be induced by the event $i : \rho_{ij}$ ($i > j$ then the one with the highest probability was chosen and the median and the quantiles for the i index were calculated for all tuples of simulated parameters. The green horizontal line corresponds to the magnitude 7.4 earthquake index of April 16. As can be seen, there are credibility intervals that are only formed by the median (the blue points), that is, these events are induced by the April 16th earthquake with probability 95 %.

Taking the median values of the parameter posterior distributions we Table 13. This table shows the earthquakes most likely to be father earthquakes of more than 8 descendant earthquakes. However, we can do simulations of the parameter values and for each tuple of values calculate the

probability of being induced by the magnitude 7.4 of April 16th. With this, we can see that out of the 16 earthquakes that the first approach report as direct children, only five of them have a probability greater than 0.95 of being direct descendants of the earthquake of magnitude 7.4 of April 16th. These are shown in Table 15.

Table 13: Earthquakes with more than 8 direct descendants (date in dd/mm/YYYY)

Magn.	Date	Lat.	Long.	Depth.	Closest City	No aftershocks
7.4	16/04/2016 18:58	0.35 N	80.16 W	17	Pedernales	16
6.7	18/05/2016 11:46	0.46 N	79.84 W	9	Muisne	9
6.3	20/04/2016 3:35	0.68 N	80.22 W	4	Muisne	9

Table 15: Earthquakes with probability greater than 95% of being triggered by earthquake of magnitude 7.4 of April 16th.

Magnitude	Date	Latitude	Longitude	Depth	Closest City
3.3	17/04/2016 0:59	0.32 N	80.16 W	8	Pedernales
3.8	16/04/2016 20:25	0.24 N	80.04 W	6	Pedernales
4.7	16/04/2016 19:59	0.32 N	80.22 W	5	Pedernales
4.6	16/04/2016 19:16	0.46 S	79.38 W	10	El Carmen
5.1	16/04/2016 19:11	0.07 S	80.23 W	11	Jama

The following figures show the residual analysis and the prediction for the accumulated number of events, evaluated at the times of the earthquakes. We used 1000 simulations from the posterior distributions of the parameters. The green vertical lines represent the first month after July 16th, 2016(the date of the last event that was used for the estimation)

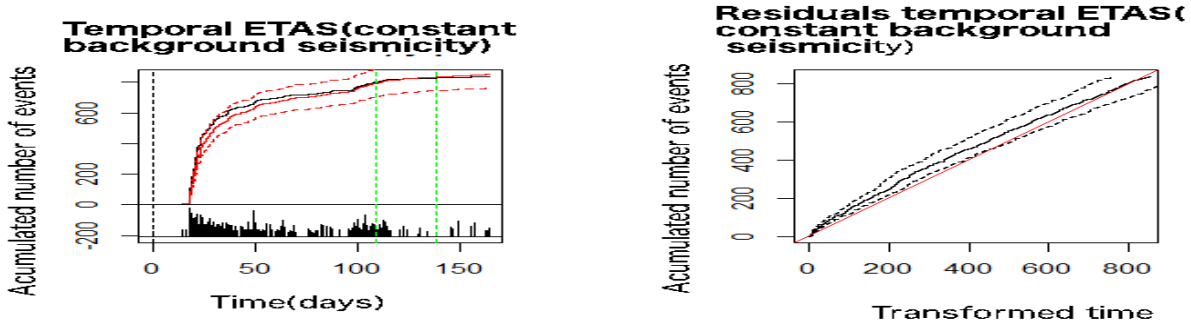
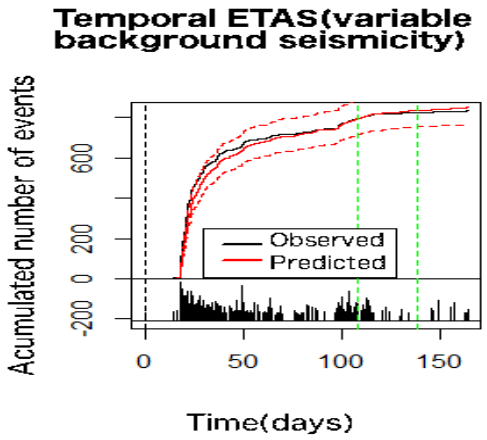


FIGURE 27 and 28. (27) The accumulated number of events for temporal Etas model with constant background seismicity. (28) Residuals for Etas temporal temporal Etas model with constant background seismicity.



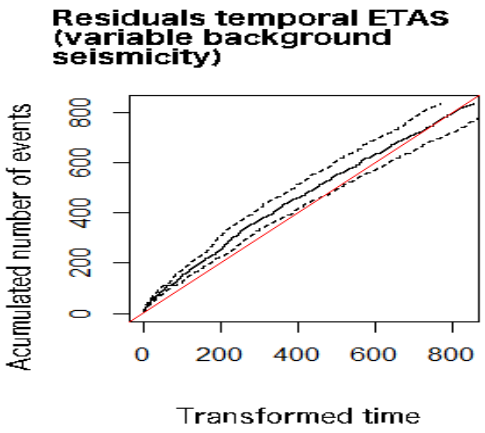
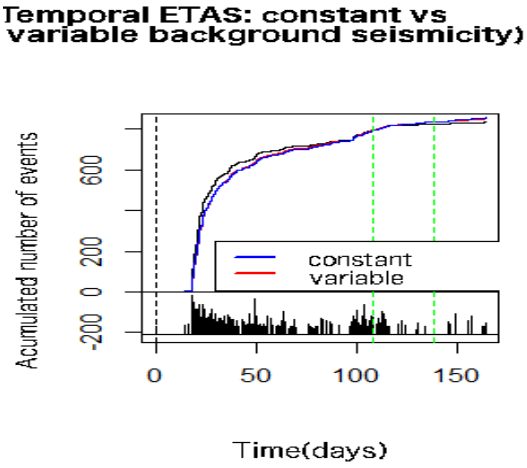


FIGURE 29 and 30. (29) The accumulated number of events for temporal Etas model with variable background seismicity. (30) Residuals for temporal Etas model with variable background seismicity.



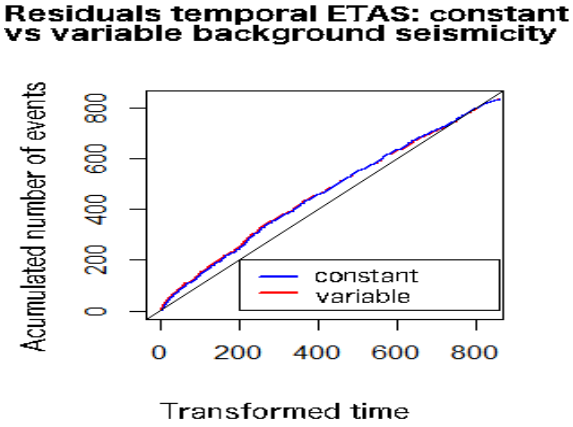
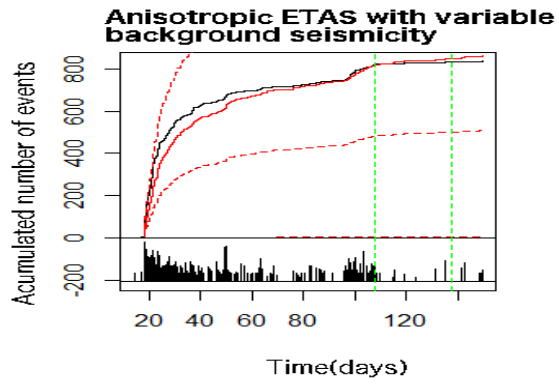


FIGURE 31 and 32. Comparison between Constant vs variable background seismicity in temporal ETAS model.



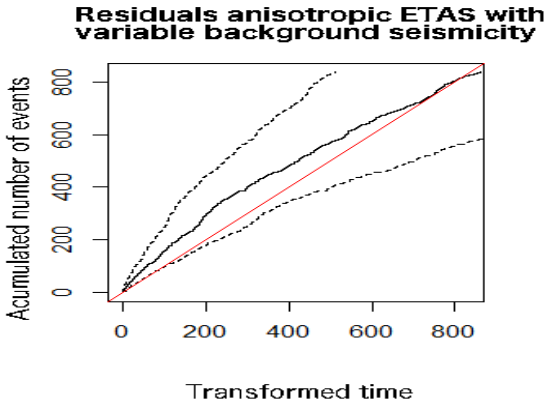


FIGURE 33 and 34. (33) The accumulated number of events for anisotropic Etas model with variable background seismicity. (34) Residuals for Etas model with variable background seismicity

Using the model Etas considering the magnitudes we can perform simulations of the parameters and with this we can calculate the average number of descendant earthquakes by father earthquake n and with this we can estimate by the formula of Saichev-Sornette, the distribution of the logarithm of the times between events (scaled by the inverse of the average time between earthquakes). Figure 31 shows the estimates with 1000 simulations and the credibility intervals at 95 % of the model prediction vs the observed. The credibility interval is not shown in Figure 32 to better distinguish the graphs.

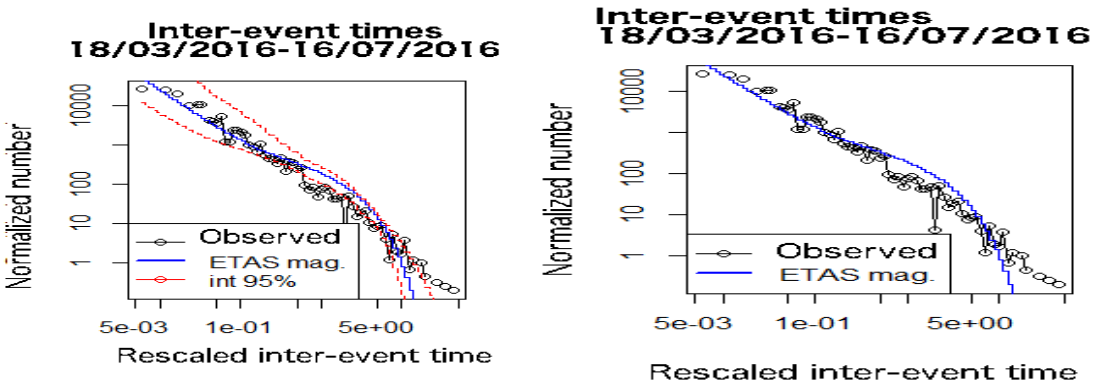


FIGURE 35 and 36. Time between events modelled with Saichev-Sornette formula. We can see the observed rescaled inter event times are inside the predicted 95% confidence interval.

CONCLUSIONS

Bayesian analysis is a useful tool for parameters estimation by maximizing the logarithm of the likelihood. Instead of point estimates, the Bayesian analysis provides a posteriori probability distribution for the parameters, allowing simulations (through extractions). But, Bayesian analysis can be sensitive on the prior distribution selection and further research is needed to assess the effect of the prior distribution selection on a posteriori distribution for each ETAS parameter.

The study of seismicity is a complex problem because of the number of factors to consider, fortunately, there are flexible models that allow the incorporation of different aspects such as ETAS models.

We introduce the use of the Welzl Algorithm to evaluate the double integral in the log likelihood formula. This is feasible since the size of the analyzed area is small compared to the earth radius. The preprocessing was crucial to reduce the number of operations of the evolution of the HMC algorithm.

The topic of times between events is a subject that requires more advanced studies to elucidate the validity of a universal law that is independent of the region analyzed.

REFERENCES

1. http://stackoverflow.com/questions/\newline 37033535/stan-programming-for-etas-model.
2. http://stackoverflow.com/questions/\newline 37699772/stan-code-for-large-data-set.
3. http://www.ism.ac.jp/ ogata/Statsei4/\newline programme.html.
4. Adelfio G, Chiodi M. Forward Likelihood-based predictive approach for space-time processes. *Environmetrics.* 2011;22:749-757.
5. Adelfio G, Chiodi M. Kernel intensity for space-time point processes with application to seismological problems. Classification and multivariate analysis for complex data structures. *Springer-Verlag Berlin Heidelberg.* 2011:401-408.
6. Alder BJ, Wainwright TE. Studies in Molecular Dynamics. I. General Method. *The Journal of Chemical Physics.* 1959;31 (2). AIP: 459–66.

7. Bak P, Christensen K, Danon L, Scanlon T. Unified scaling law for earthquakes. *Phys Rev Lett, 88, 178501.* 2002;88:178501.
8. Barrows D. ESTIMATION AND INFERENCE OF NONLINEAR STOCHASTIC TIME SERIES. 2016;Master Thesis:McMaster-University.
9. Bayraktarli Y, Baker J, Faber M. Uncertainty treatment in earthquake modeling using Bayesian probabilistic networks. *Georisk.* 2011;5:44-58.
10. Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, others. Stan: A probabilistic programming language. *Journal of Statistical Software.* 2011;20:1-37.
11. Chu A, Schoenberg FP, Bird P, Jackson DD, Kagan YY. Comparison of ETAS parameter estimates across different global tectonic zones. *BullSeism Soc Am.* 2011;101(5):2323-2339.
12. Collot JY, Marcaillou B, Sage F, et al. Are rupture zone limits of great subduction earthquakes controlled by upper plate structures? Evidence from multichannel seismic reflection data acquired across the northern Ecuador–southwest Colombia margin. *JOURNAL OF GEOPHYSICAL RESEARCH, VOL 109, B11103.* 2004;109:http://onlinelibrary.wiley.com/doi\newline-/10.1029/2004 JB003060/#references.
13. Corral A. Universal local versus unified global scaling laws in the statistics of seismicity. *Physica A.* 2004;340:590-597.
14. Corral A. Dependence of earthquake recurrence times and independence of magnitudes on seismicity history. *Tectonophysics.* 2006;424:177-193.
15. Cressie N. Statistics for spatial data. *Ed Wiley, New York.* 1993.
16. Daley DJ, Vere-Jones D. An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods, Second Edition. 2002.
17. Epifany AIaLLP. Bayesian estimation for a parametric Markov Renewal model applied to seismic data. 2014.
18. Fletcher R, Powell MJD. A rapidly convergent descent method for minimization. *Comput J.* 1963;6:163-168.
19. Gelman A. Prior distribution. *http://wwwstatcolumbiaedu/ gelman/research/published/p039-_opdf.* 2002;3:1634-1637.
20. Gelman A, others. Bayesian data analysis. third edition. *Chapman & Hall/CRC Texts in Statistical Science.* 2013;3.
21. Government of Japan H. AFTERSHOCK PROBABILITY EVALUATION METHODS. *The Headquarters for Earthquake Research Promotion Earthquake Research Committee Prime Minister's Office Government of Japan.* 1998.
22. Guo JYZ, Zhou S. A hypocentral version of the space-time ETAS model. *Geophysical Journal International.* 2015;DOI: 10.1093/gji/ggv319.
23. H E, Jalayer F. Robust seismicity forecasting based on Bayesian parameter estimation for epidemiological spatio-temporal aftershock clustering models. 2017:DOI:10.1038/s41598-41017-09962-z.
24. Hawkes AG. Point spectra for some mutually exciting point process. *J Roy Statist Soc Ser B.* 1971;33:438-443.
25. Hawkes AG, Adamopoulos L. Cluster models for earthquakes- Regional comparisons. *Bull Int Statist Inst.* 1973;45(3):454-461.
26. Hernandez Vargas NA. Bayesian Point Process Modelling of Earthquake Occurrences. 2012.
27. Hoffman MD, Gelman A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research.* 2014;15 (1):1593-1623.
28. Holden L, Natvig B, Sannan S, Bungum H. Modeling Spatial and Temporal Dependencies

between Earthquakes. *Natural and Anthropogenically Induced Hazards.* 2000.
29. Holschneider M, Narteau C, Shebalin P, Peng Z, Schorlemmer D. Bayesian analysis of the modified Omori law. *JOURNAL OF GEOPHYSICAL RESEARCH.* 2012;117:B06317,-doi:06310.01029/02011JB009054.
30. Kagan YY, Knopoff L. Dependence of seismicity on depth. *Bulletin of the Seismological Society of America.* 1980;vol. 70 no. 5.
31. Kagan Y, Knopoff L. Stochastic synthesis of earthquake catalogs. *Journal of Geophysical Research Solid Earth.* 1981;86:2853-2862.
32. Lippiello E, Arcangelis LD, Giacco F, Godano C. Parameter Estimation in the ETAS Model: Approximations and Novel Methods. *Bulletin of the Seismological Society of America.* 2014:DOI:-10.1785/0120130148.
33. Lotto G, Stein R. Ecuador earthquakes: What happened and what's next? 2016;http://temblor.net/earthquake-insights/ecuador-earthquakes-what-happened-and-what-is-next-986/.
34. Mueller UK. Measuring prior sensitivity and prior informativeness in large Bayesian models. *Journal of Monetary Economics.* 2012;59 (2012):581-597.
35. Neal RM, others. MCMC Using Hamiltonian Dynamics." Handbook of Markov Chain Monte Carlo 2. *Journal of Machine Learning Research.* 2011;113–62.
36. Ogata Y. The asymptotic behavior of maximum likelihood estimators for point processes. *Ann Int Statist Math.* 1978;30:243-261.
37. Ogata Y. STIMATION OF THE PARAMETERS IN THE MODIFIED OMORI FORMULA FOR AFTERSHOCK FREQUENCIES BY THE MAXIMUM LIKELIHOOD PROCEDURE. *J Phys Earth.* 1983;31:115-124.
38. Ogata Y. Statistical models for earthquake occurrences and residual analysis for point processes. *J Amer Statist Assoc.* 1988;83:9-27.
39. Ogata Y. Space-time point-process models for earthquake occurrences, Annals of the Institute of Statistical Mathematics. *Annals of the Institute of Statistical Mathematics.* 1998;50(2):379-402.
40. Ogata Y. Significant improvements of the space-time ETAS model for forecasting of accurate baseline seismicity. *Earth Planets Space.* 2011;63:217-229.
41. Ogata Y, Tsuruoka H. Statistical monitoring of aftershock sequences: a case study of the 2015 Mw7.8 Gorkha, Nepal, earthquake. *Earth, Planets and Space.* 2016;68:44:DOI-10.1186/s40623-40016-40410-40628.
42. Ogata Y, Zhuang J. Space–time ETAS models and an improved extension. 2006.
43. Omi T, Ogata Y, Hirata H, Aihara K. Forecasting large aftershocks within one day after the main shock. *SCIENTIFIC REPORTS.* 2013;68:DOI:-10.1038/srep02218.
44. Omori O. On the aftershocks of earthquakes. *Journal of the College of Science, Imperial University of Tokyo.* 1894;7:111-200.
45. Rasmussen JG. Bayesian inference for Hawkes processes. *Methodology and Computing in Applied Probability.* 2011;15(15):pp.-1–20. http://dx.doi.org/10.1007/s11009-11011-19272-11005.
46. Ross GJ. Bayesian Estimation of the ETAS Model for Earthquake Occurrences. *http://wwwgordonjrosscouk/ bayesianetaspdf.* 2016.
47. Saichev A, Sornette D. Universal Distribution of Inter-Earthquake Times Explained. *Phys Rev Letts.* 2006;97, 078501.
48. Saichev A, Sornette D. Theory of earthquake recurrence times. *JOURNAL OF GEOPHYSICAL RESEARCH.* 2007;112 , B04313:doi:10.1029/2006JB004536,-002007.

49. Schoenberg FP. Facilitated estimation of ETAS. *Bulletin of the Seismological Society of America.* 2013;103(1):601-605 doi:610.1785/0120120146.
50. Sornette D, Werner MJ. Apparent clustering and apparent background earthquakes biased by undetected seismicity. *Journal of Geophysical Research Solid Earth.* 2005;110:243-261.
51. Sornette D, Werner MJ. Constraints on the size of the smallest triggering earthquake from the epidemic type aftershock sequence model, Båth's law, and observed aftershock sequences. *Journal of Geophysical Research Solid Earth.* 2005;110.
52. Touati S. Complexity, aftershock sequences, and uncertainty in earthquake statistics. *University Edingbourgh.* 2011;Phd. Thesis.
53. Utsu T. Aftershocks and earthquake statistics (I) - Some parameters which characterize an aftershock sequence and their interrelations. *J Fac Sci Hokkaido Univ.* 1969;Ser.VII 3:121-195.
54. Utsu T, Ogata Y, Matsu'ura RS. The centenary of the Omori formula for a decay law of aftershock activity. *Journal of Physics of the Earth.* 1995;43:1-33.
55. Varini E, Ogata Y. Bayesian estimation of doubly stochastic Poisson processes for detection of seismicity phases. 2014.
56. Vere-Jones D. Forecasting earthquakes and earthquake risk. *International Journal of Forecasting.* 1995;11:503-538.
57. Welzl E. Smallest Enclosing Disks (Balls and Ellipsoids) Occurrences. *H Maurer (Ed), New Results and New Trends in Computer Science, Lecture Notes in Computer Science.* 1991;555:359-337.
58. WorkingGroup WGOCEP. Earthquake Probabilities in the San Francisco Bay Region: 2002–2031. *US Geological Survey.* 2003;Open-File Report 03-214:https://pubs.usgs.gov/of/2003/of2003-2214/WG2002_OFR-2003-2214_Chapter2005.pdf.
59. Yamanaka Y, Shimazaki K. Scaling Relationship between the Number of Aftershocks and the Size of the Main Shock. *J Phys Earth.* 1990;38:305-324.
60. Zhuang J, Christophersen A, Savage M, Vere-Jones D, Ogata Y, Jackson D. Differences between spontaneous and triggered earthquakes: Their influences on foreshock probabilities. *JOURNAL OF GEOPHYSICAL RESEARCH.* 2008;113,, B11302:doi:10.1029/2008JB005579.
61. Zhuang J, Ogata Y. Properties of the probability distribution associated with the largest event in an earthquake cluster and their implications to foreshocks. *PHYSICAL REVIEW.* 2006;E 73:046134.
62. Zhuang J, Ogata Y, Vere-Jones D. Stochastic Declustering of Space-Time Earthquake Occurrences. *Journal of the American Statistical Association.* 2002;97, No. 458.
63. Zhuang J, Ogata Y, Vere-Jones D. Diagnostic analysis of space-time branching processes for earthquakes. 2004.