JSEE Vol. 22, No. 4, 2020 DOI: 10.48303/jsee.2022.254647

Research Paper

Bayesian and Non Parametric Estimation of ETAS Models Applied to Seismic Recurrence in Ecuador 2016

Fausto Fabian Crespo Fernandez[1] and Carlos Jimenez Mosquera[2*]

1. MSc. Graduate, San Francisco de Quito University, Ecuador
2. Ph.D. San Francisco de Quito University, Ecuador,
*Corresponding Author; email: cjimenez@usfq.edu.ec

Received: 17/10/2017
Revised: 02/05/2018
Accepted: 19/05/2018

ABSTRACT

In this paper, the purpose is to analyze, from the Bayesian point of view, the occurrence rate of earthquakes in Ecuador since March 2016 to July 16, 2016. We implemented the ETAS models, starting with the purely temporal model, then considering the magnitudes, and later the spatio-temporal models (both isotropic and anisotropic), and finally the hypo-central model. We introduced the use of Welzl algorithm to evaluate the log-likelihood of the occurrence rate for spatio-temporal models. We conducted simulations by extracting values from the a posteriori distributions of the models parameters, to obtain 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 from July 16, 2016 to September 2016.

Keywords:
Bayesian analysis; ETAS; Rstan; Welzl algorithm and inter-event times and earthquake; Ecuador

1. 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) and 1979 (magnitude 8.2) [1]. The Paleogene- Neogene faults of Jama-Quininde and Esmeraldas define a 200 km long zone, which was the rupture zone of the 1942 earthquake [1]. This zone is highly seismic.

The earthquake on April 16, 2016 of magnitude 7.8 [2] (according to the published figures by the Geophysics Institute of the National Polytechnic University EPN, the magnitude of this earthquake was 7.4) caused losses of millions of dollars and more than 650 deaths. Hence it is important to study the aftershocks sequence after the big earthquake on April 16, 2016 to make predictions.

Predicting earthquakes has been a difficult problem for researchers for decades. The aftershocks following an earthquake of great magnitude usually occurs in swarms where sometimes the main earthquake can be distinguished and sometimes not. These aftershocks can in turn cause other a ftershocks of smaller magnitude. In addition, earthquakes tend to be correlated both spatially and temporally. There is also the problem of precursor earthquakes that are smaller earthquakes that precede an earthquake of greater magnitude. These precursor earth-quakes are not part of any ongoing aftershock sequence and are followed in a short time by a major earthquake. Aftershock sequences also contain many small events that are followed by larger events. Thus it is difficult to distinguish aftershocks from precursor earthquakes. That is why

it is usual to eliminate aftershocks from the data known as declustering.

One of the simplest ways to model the rate of occurrence of aftershocks is by non-stationary Poisson processes in time or space. Other models used in seismic data analysis have been Marked Point Processes [3] (the marks are the magnitudes of earthquakes), Homogeneous Markov Renewal Processes [4] with a Weibull distribution for the times between earthquakes, model ETAS or Epidemic-Type Aftershock Sequences [5-6], and Poisson doubly stochastic processes [7]. BPT or Brownian Passage Times [8], which is a time-dependent renewal process, and Probabilistic Bayesian Networks [9].

This article analyzes the ETAS models, starting with the purely temporal model then considering the magnitudes of the aftershocks, and then using the spatial-temporal model and finally the hypo-central version to predict the occurrence rate of earthquakes with magnitude higher than one fixed threshold M0. We also make estimations of the models parameters using Bayesian statistics and HMC (Hamiltonian Markov Chains) through Rstan (the R interface of the Stan C++ Package).

The empirical law of Omori [10] and the law of Omori-Utsu (also called Modified Law of Omori) [5] describe the decreasing frequency of aftershocks over time after an earthquake:

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

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

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 lets us model complex aspects immediately after the main earthquake. The value of c is normally less than 0.1 days [11]. In addition, the occurrence rate of events decreases with the inverse of the p power of $t+c$ (power law).

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 close to 1. If we take into account the branching property of the occurrence of the earthquakes, then the value of p increases from 1 to about 1.5 [12].

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:

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

That is, the number of events of magnitude greater than a threshold decreases exponentially with the increase of the magnitude of that threshold by a power law.

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 $H_t : \lambda(t, x, y, M \mid H_t)$ is the expected number of earthquakes in the unit of time, space, and magnitude [13].

Marked Point Processes are stochastic models used to represent a finite number of events located in time and space [14]. This approach where marks are the magnitudes of earthquakes is discussed in [6, 15-16].

An earthquake T is represented by a tuple (x_i, y_i, z_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 model ETAS [6], models the magnitudes and the times of the earthquakes and was extended to the spatio-temporal case by Ogata in 1998 [6]. The ETAS model is a special type of Hawkes point process, which is 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:

$$\lambda(t \mid H_t) = \mu + \sum_{j:t_j<t} \frac{K(p-1)c^{p-1}}{(t-t_j+c)^p} \tag{4}$$

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

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

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 invents 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.

Considering the magnitudes of earthquakes we have the model:

$$\lambda(t \mid H_t) = \mu + \sum_{j:t_j<t} \frac{K(p-1)c^{p-1}Ae^{\alpha(M_j-M_o)}}{(t-t_j+c)^p} \tag{7}$$

where M_o is the threshold magnitude that is chosen arbitrarily as the minimum magnitude that provides completeness of the catalog of earthquakes. Sornette and Werner [17-18] showed that this threshold quantity is not related to the minimum magnitude needed to trigger other events, which is what actually represents in the ETAS model [19]. Besides, it was also shown that a greater M_o causes the reduction of the parameter α and the ratio of branching (or number of descendants earthquakes per father earthquake), that is, more earthquakes seem to be independent [19].

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

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

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

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:

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

with $s(m) = \beta e^{-\beta m}$ the distribution of magnitudes by the Gutenberg-Richter law and the $\lambda(t)$ branching term in Equation (8).

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

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

where β is the parameter of the distribution of the number of earthquakes greater than $m : s(m) = \beta e^{-\beta m}$ and $\beta = b\ln 10$ (b from Gutenberg-Richter Law).

Assuming $M_{\max} = \infty$ the previous formula can be reduced to [18-19]:

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

So n is infinite if $p < 1$ or if

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

$$\int_{M_o}^{M_{\max}} \beta e^{-\beta(m)} dm = 1 \tag{13}$$

implies we need to add $1/\left(\exp(-\beta M_o) - \exp(-\beta M_{\max})\right)$ to the constant β.

The expected event rate as a function of time can be calculated as:

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

where now we have $s(m) = \beta e^{-\beta(m-M_o)}$. Then [19]:

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

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

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

where:

- $\mu(x, y)$ is the background intensity that is a function of latitude and longitude (does not depend on the time).
- $\kappa(M) = Ae^{\alpha(M-M_o)}$ 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 [20-21].

And the spatial probability density of earthquakes may have one of the following forms [22-23]:

$$f(x, y \mid M, H_t) = \frac{1}{2\pi d e^{\alpha(M_j - M_0)}} \exp\left(-\frac{x^2 + y^2}{2de^{\alpha(M_j - M_0)}}\right) \quad (17)$$

$$f(x, y \mid M, H_t) = \frac{(q-1)d^{q-1}}{\pi}\left\{\frac{x^2 + y^2}{e^{\alpha(M-M_0)}} + d\right\}^{-q} \quad (18)$$

$$f(x, y \mid M, H_t) = \frac{(q-1)\, d^{q-1} e^{\alpha(M-M_0)}}{\pi}\left\{x^2 + y^2 + d\right\}^{-q} \quad (19)$$

$$f(x, y \mid M, H_t) = \frac{(q-1)}{\pi d e^{\alpha(M-M_0)}}\left\{1 + \frac{x^2 + y^2}{de^{\alpha(M-M_0)}}\right\}^{-q} \quad (20)$$

The position-dependent intensity $\mu(x, y)$ is calculated by means of bicubic splines [6], kernel functions [24-25], grid average [26], tessellation [27] or by using non-parametric estimation by the Forward Likelihood Predictive method (FLP) [28] that is implemented in the ETAS FLP library for *R*.

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

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

where

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

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

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

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

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

Where

$$g(t, x, y, M) = \frac{Ke^{\alpha(M-M_0)}}{(t+c)^p}\left\{\frac{pS_i p^T}{e^{\alpha(M-M_0)}} + d\right\}^{-q} \quad (25)$$

and $p = (x - x_j, y - y_j)$ 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) and S_j *j* = 1; 2, ... are positive definite symmetric matrices represent-ing 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 previous selected earthquake, are those with latitude and longitude $\pm 3.33 \times 10^{0.5M_j - 2}$ km (Utsu spatial distance) from the latitude and longitude of the selected earthquake and with a time difference therefrom (towards the future) of max $(100, 10^{0.5M_j - 1})$ days [6]. Then the process is repeated with earthquakes that are not yet in any cluster until all earthquakes belong to a cluster.

Zhuang et al. [29] presented a modification of the spatio temporal ETAS model that includes the depths of earthquakes.

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

Where

$$g(t,\ x,\ y, M) = \frac{Ke^{\alpha(M-M_0)}}{(t+c)^p} \times \left\{ \frac{pS_i p^T}{e^{\alpha(M_j-M_0)}} + d \right\}^{-q} h(z-z_i, z_i) \quad (27)$$

and

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

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

2. Stochastic Declustering

We consider events (t_j, x_j, y_j) with $j = 1,\ldots, N$ associated with the probabilities $j = 1,\ldots, N$ and suppose that each event j is removed with probability ρ_j, then the remaining events represent a new point process called thinned process. Now ρ_j, is the probability that event j is induced by the preceding events $\rho_j = \sum_{i=1}^{j-1} \rho_{i,j}$ where $\rho_{i,j}$ is the probability that event j is induced by event i. Then the probability that event j is background event is $1-\rho_{i,j}$. If we delete the event j with probability ρ_j for all $j = 1,\ldots, N$ we obtain a non-homogeneous Poisson process associated with the spatial intensity $\mu(x,y)$. This sub process is called the background process and the com-plementary process is called the clustering process [24].

The total intensity function is:

$$m_1(x,y) = \lim_{\to\infty} \frac{1}{T}\int_0^T \lambda(t,x,y \mid H_t)dt \approx \mu(x,y) + \frac{1}{T}\sum_{j:t_j<t} \kappa(M_j) f(x-x_j, y-y_j \mid M_j) \quad (29)$$

and can be estimated using Gaussian kernel estimation

$$\hat{m}_1(x,y) = \frac{1}{T}\sum_{j=1} k_{d_j}(x-x_j, y-y_j) \quad (30)$$

con

$$k_d(x,y) = \frac{1}{2\pi d}\exp\left(-\frac{x^2+y^2}{2d^2}\right) \quad (31)$$

and d_j is a variable bandwidth for each earthquake j that is calculated as follows: given a positive integer n_p between 10 and 100 we find the smallest disk centered at the epicenter of the event j and with a radius greater than a certain small value (0.02 degrees). The occurrence rate of the cluster and the background sub processes are calculated by [24, 30]:

$$\hat{\Upsilon}(x,y) = \frac{1}{T}\sum_j \rho_j k_{d_j}(x-x_j, y-y_j) \quad (32)$$

and

$$\hat{\mu}(x,y) = \hat{m}_1(x,y) - \hat{\Upsilon}(x,y) = \frac{1}{T}\sum_j (1-\rho_j) k_{d_j}(x-x_j, y-y_j) \quad (33)$$

This method is implemented in the *R* package ETAS, and is used to estimate the background seismicity $\mu(x,y)$ and with these values we can estimate the other parameters of the ETAS models using Bayesian statistics and Rstan.

3. Forward Likelihood Predictive FLP

With the FLP method, the background intensity and the induced intensity of the ETAS model (branching process) are estimated at the same time. It is based on successive increments of the log of the likelihood by adding one event at a time. Let:

$$\log L(\hat{\lambda}_{\psi;H_{t_k}}(z); H_{t_{k+1}}) \quad (34)$$

be the logarithm of the likelihood with the first k+1 events using the estimator $\hat{\lambda}_{\psi;H_{t_k}}(z)$ where ψ are the parameters. The predictive information of the first k events on the k+1 event is:

$$\delta_{k,k+1}(\psi) = \log L(\hat{\lambda}_{\psi;H_{t_k}}(z); H_{t_{k+1}}) - \log L(\hat{\lambda}_{\psi;H_{t_k}}(z); H_{t_k}) \quad (35)$$

This is similar to cross validation but only applied to the events following one specific event.

Therefore $\psi(H_{t_k})$ is chosen such that maximizes

$$FLP_{k_1,k_2(\bar{\psi})} = \sum_{k=k_1}^{k_2} \delta_{k,k+1} \quad (36)$$

with $k_1 = [x/2]$ and $k_2 = n-1$. This method is better than the kernel estimation for spatio-temporal models [28].

4. Bayesian Statistics and Hamiltonian Markov Chains HMC

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 [31].

Besides, Bayesian statistics can consider the uncertainty not only in the model parameters, conditioned on the available catalogue 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 [32].

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 [31]. 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 [33]. 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 non-informative prior distributions thus allowing each parameter to be estimated from the data. But, if non-informative prior distributions were assigned to all the parameters, then the model would fit the data very closely but with scientifically unreasonable parameters [34-35]. Thus we prefer to use weakly informative priors (heavy tailed distributions), although the prior selection sensitivity is an important aspect in Bayesian analysis [36].

We used flexible priors of the parameters of the models. In this case, 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 [31].

Markov chains stabilize in a state of equilibrium from any initial state if the chain is irreducible (from any state we can reach another with probability > 0), aperiodic (the chain is not trapped in cycles) and recurrent positive (the number of steps to get from one state to another is finite).

If θ is the last accepted value of the parameter, a new value is proposed from a jump dis-tribution (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 [37].

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

Hamiltonian Monte Carlo, also called Hybrid Monte Carlo (HMC), is a Markov chain Monte Carlo (MCMC) method to approximate integrals with respect to a target probability distribution π on R^d. It was originally proposed by Duane et al. [38], it was later introduced in statistics by Neal [39] and is now part of the standard toolbox [40], in part due to favorable scaling properties with respect to the dimension d [41-42], com- pared to random walk Metropolis {Hastings. Hamiltonian Monte Carlo is at the core of the No-U-Turn sampler (NUTS [43]) used in the software Stan [44].

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 - 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 $r=(r_1,r_2,...,r_n)$ with n is the number of components of θ, and $M_{n\times n}$ is 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 $\varepsilon/2$ are used for the evolution of in the first dr/dt and last step respectively.

The final values of the Leapfrog steps are the proposed values (θ^*,r^*) and are accepted or rejected similarly to the Metropolis, but now:

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

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:

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

where Z is a normalization constant. Then [37]:

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

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, but the "No-UTurn Sampler" or NUTs [44], optimizes these adaptively.

5. Parameters Estimation

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

$$\log L(\theta)=\sum_j \log\lambda(t_j\mid H_t)\int_0^{T_{\max}}\lambda(t)dt \tag{43}$$

and for spatio temporal models

$$\log L(\theta)=\sum_j \log\lambda(t_j\mid H_t)\int_0^{T_{\max}}\iint_S\lambda(t,x,y)dxdydt \tag{44}$$

where θ is the vector of parameters and S is the region that contains the earthquake data, which is commonly the rectangle $[\text{long}_{\min},\ \text{long}_{\max}]\times[\text{lat}_{\min},\ \text{lat}_{\max}]$.

For the models 3 and 6, the logarithms of the likelihood have respectively the following analytic forms:

$$\log L(\mu,k,p,c)=\sum_{i=1}^{N}\log(\lambda(t_i))-\mu T_{\max}-k\sum_{i=1}^{N}\left(1-\frac{c^{p-1}}{(T_{\max}-t_i+c)^{p-1}}\right) \tag{45}$$

$$\log L(\mu,k,p,c,A,\alpha)=\sum_{i=1}^{N}\log(\lambda(t_i))-\mu T_{\max}-kA\sum_{i=1}^{N}e^{\alpha(M_i-M_0)}\left(1-\frac{c^{p-1}}{(T_{\max}-t_i+c)^{p-1}}\right) \tag{46}$$

For the case of the spatio temporal ETAS models, the logarithm of the likelihood does not have a closed analytical form. As mentioned in [45] 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 peaks in three dimensions and this integral has to be calculated for each tuple of parameters during the optimization [45]. Schoenberg mentions that the approximations of this integral, work poorly with optimization routines. In [6], 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 [45].

In [45], the author proposed for the spatio-temporal model the approximation:

$$\log L(\mu,k,p,c,\alpha,\Upsilon)\approx\sum_{i=1}^{N}\log(\lambda(t_i))-\mu T_{\max}-k\sum_{i=1}^{N}e^{\alpha(M_i-M_0)} \tag{47}$$

This approximation assumes that the time $T_{\max} \to \infty$ and the size of the observed area of the earthquakes tend 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 [45] and has the property that for any j and any values of the parameters of the ETAS model:

$$Ke^{\alpha(M_j - M_0)} = \int_{IR}\int_{IR}\int_{IR} g(t-t_j, x-x_j, y-y_j)dtdxdy \geq \iiint_S g(t-t_j, x-x_j, y-y_j)dtdxdy \quad (48)$$

That is, if certain values of the parameters are poor, then $K\sum_{i=1}^{N} e^{\alpha(M_j - M_0)}$ will be larger than $\iiint_S g(t-t_j, x-x_j, y-y_j)dtdxdy$ and the optimization routine will avoid these parameter values [45].

However, this approach is analyzed in [46], and it is shown that the assumption that $t \to \infty$ is too crude, which causes the introduction of systematic biases [46].

Because the earthquakes analyzed in the provinces of Manabi 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 of the spatio-temporal model of isotropic cluster making the region S the minimum covering circle containing the positions of earthquakes. This is done by the Welzl algorithm [47] and then we can approximate the integral by polar coordinates centered at the position of each earthquake with:

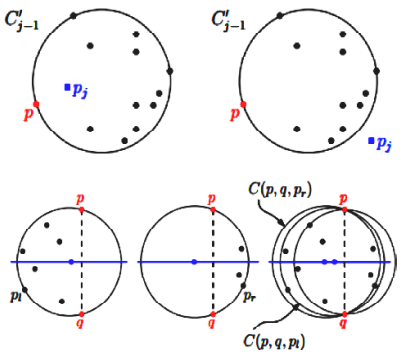
$$\log L(\mu, k, p, c, d, \alpha, \Upsilon) \approx \sum_{i=1}^{N} \log(\lambda(t_i)) - \mu T_{\max}\pi r^2 - k\alpha\sum_{i=1}^{N} e^{\alpha(M_j - M_0)} \times \left(1 - \frac{c^{p-1}}{(T_{\max} - t_i + c)^{p-1}}\right)\left(1 - \frac{d^{q-1}}{\left(\frac{r_i^2}{e^{\alpha(M_i - M_0)}} + d\right)^{q-1}}\right) \quad (49)$$

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 depends on the angle θ. Instead, we approximate the maximum ρ, in any direction, as the maximum distance between the earthquake i and other earthquakes).

In this approximation, it is not assumed that $t \to \infty$ but that 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 $O(n)$. In summary, the algorithm is as follows:

- The points $p_1, p_2, ..., p_{i-1}$ are randomized and let C_j be the minimum covering circle for $p_1, p_2, ..., p_i$ $(j \leq i-1)$ with point p at its boundary
- We already have C'_{j-1} and we want to add p_j. If p_j is inside of C'_{j-1} then $C_{j-1} = C_j$, and if it is outside, then C_j will have in its boundary the point p_j and the point p.
- We draw the line l through $q = p_j$ and p and the perpendicular bisector of l, and we can assume without loss of generality that the line l is vertical, then we find points p_l and p_r to the left and right of l respectively, such that the centers of the circles that pass through $q = p_j, p, p_l : C(p, q, p_l)$ and $p_j = q, p, p_r : C(p, q, p_r)$ be as far as possible from l (to the left and right respectively). Then we choose the circle of smallest radius of the previous three circles.



For the anisotropic model, the Euclidean distance is changed by the metric defined by the standardized variance covariance matrix S_j 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:

$$\log L(\theta) = \sum_j \log \lambda(t_j, x_j, y_j \mid H_t) - \int_0^{T_{\max}} \iint_S \int_0^Z \lambda(t, x, y) dz dx dy dt \tag{50}$$

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 $\log(L_{3d}) = \log(L_{2d}) - N \log(Z)$ [28].

For the case of models 3 and 6 with variable background seismicity (instead of μ we have $\mu * \mu(x, y)$) the logarithm of the likelihood is:

$$\log L(\mu, k, p, c) = \sum_{i=1}^N \log(\lambda(t_i)) - \mu t_{\max} \sum_{i=1}^N \mu(x, y) - k \sum_{i=1}^N \left(1 - \frac{c^{p-1}}{(t_{\max} - t_i + c)^{p-1}}\right) \tag{51}$$

$$\log L(\mu, k, p, c, A, \alpha) = \sum_{i=1}^N \log(\lambda(t_i)) - \mu t_{\max} \sum_{i=1}^N \mu(x, y) - kA \sum_{i=1}^N e^{\alpha(M_i - M_0)} \left(1 - \frac{c^{p-1}}{(t_{\max} - t_i + c)^{p-1}}\right) \tag{52}$$

And similarly for the temporal and hypo-central 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 sources 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 Monte Carlo MCMC chains [33], in this case, Hamiltonian Monte Chains HMC which is implemented in the stan language. Also for the case of the temporal models 3 and 6 we compared the obtained results with Rstan with those obtained by the program ETAS.exe of Ogata available in [49] that minimizes by the method of Davidon-Fletcher-Powell [48].

6. Preprocessing

To achieve efficiency, a preprocessing was performed and we do not pass the values of the times, magnitudes and latitudes of earthquakes directly to the Rstan library (as is done for the temporal model in Stackoverow [50-51], because then, in each iteration, it would be necessary to calculate for each earthquake, the differences of time, latitude and longitude, 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 $N(N-1)/2$. 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:

$$start = N * (j-1) - (j * (j-1))/2 + 1$$
$$end = j \ \ N - (j * (j+1))/2$$

For the case of the anisotropic cluster model, the earthquake data of each cluster were adjusted with four normal bivariate models [6, 22, 52]:

$$N\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} \tilde{\sigma}^2 & 0 \\ 0 & \tilde{\sigma}^2 \end{pmatrix}\right) \tag{53}$$

$$N\left(\begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}, \begin{pmatrix} \hat{\sigma}^2 & 0 \\ 0 & \hat{\sigma}^2 \end{pmatrix}\right) \tag{54}$$

$$N\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} \tilde{\sigma}_1^2 & \tilde{\rho}\,\tilde{\sigma}_1\tilde{\sigma}_2 \\ \tilde{\rho}\,\tilde{\sigma}_1\tilde{\sigma}_2 & \tilde{\sigma}_2^2 \end{pmatrix}\right) \tag{55}$$

$$N\left(\begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}, \begin{pmatrix} \hat{\sigma}_1^2 & \hat{\rho}\,\hat{\sigma}_1\hat{\sigma}_2 \\ \hat{\rho}\,\hat{\sigma}_1\hat{\sigma}_2 & \hat{\sigma}_2^2 \end{pmatrix}\right) \tag{56}$$

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 [6]:

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

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

$$\tilde{\sigma}_1^2 = \left[\sum_j (x_j, x_1)^2\right] / n \quad (59)$$

$$\tilde{\sigma}_2^2 = \left[\sum_j (y_j, y_1)^2\right] / n \quad (60)$$

$$\tilde{\rho} = \left[\sum_j (x_j, x_1)(y_j, y_1)\right] / (n\tilde{\sigma}_1\tilde{\sigma}_2) \quad (61)$$

$$\hat{\sigma}_1^2 = \left[\sum_j (x_j, \bar{x})^2\right] / n \quad (62)$$

$$\hat{\sigma}_2^2 = \left[\sum_j (y_j, \bar{y})^2\right] / n \quad (63)$$

$$\hat{\rho} = \left[\sum_j (x_j, \bar{x})(y_j, \bar{y})\right] / (n\hat{\sigma}_1\hat{\sigma}_2) \quad (64)$$

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 [6]. 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 [6]. Then the selected matrix of variance covariance is normalized:

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

In addition, as part of the preprocessing 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.

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 [53, 24]. 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 [54]:

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

Similarly the probability that the event j is produced by the event i is:

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

We can also obtain the expected number of direct aftershocks from the earthquake i as $\sum_j \rho_{ij}$ [40].

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: $S(t) = \sum_{i:t_i<t} \phi(i)$ [30].

7. Residual Analysis and Goodness of Fit

Analysis of residuals and goodness of fit of the models can be done as mentioned in [30], 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 in 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.

8. Time between Events

The relationship between inter-event times and ETAS models was analyzed analytically by Saichev and Sornette [55]. In [56], it is analyzed the existence of a universal law of escalation [57] for the probability density function of the recurrence times or times between earthquakes (time between two successive events) τ.

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

where the function $f(x)$ has been found practically the same in different regions and λ the average rate

of events observed in the analyzed region [55].

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 [56] is:

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

where θ is the parameter of the Omori-Utsu or Modified Omori law form, $\Phi(t)=\theta c^{\theta}/(c+t)^{1+\theta}$, is the average number of de-scendants per earthquake, and $P(\tau)\approx\varphi(x,\varepsilon)=\exp(-(1-n)x-(n\varepsilon^{\theta}/1-\theta)x^{1-\theta})$ is the probability that there are no events in $[t,t+\tau]$.

9. Data Analysis and Results

The data for this study is available in http://www.igepn.edu.ec/ultimos-sismos. Table (1) shows a summary by province and month of the 908 earthquakes that occurred in Ecuador from March 18, 2016 to July 16, 2016. Out of the 908 earthquakes, 810 had their epicenter closest to the provinces of Manabi and Esmeraldas. On the other hand, the cities that had the greatest number of earthquakes with near epicenters were: Jama 241 earthquakes; Muisne 222 earthquakes; Pedernales 77 earthquakes; Manta 66 earthquakes; Puerto Lopez 64 earthquakes and Bahia de Caraquez 49 earthquakes.

Table (2) shows the monthly average of the earthquakes magnitude in the provinces Manabi and Esmeraldas and the total average considering both provinces per month. Here we see that the average magnitude in April (when the 7.4 earthquake occurred) is not the greater value because the number of earthquakes in that month was larger than in other months, and then an earthquake

Table 1. Earthquake distribution in Ecuador 18/03/2016-16/07/2016.

Closest Province	March	April	May	June	July	Total
Azuay			1			1
Bolivar		4				4
Canar	1	4				5
Carchi	1	1				2
Cotopaxi	1	1				2
El Oro	2	3	4	2		11
Esmeraldas		171	47	16	35	269
Galapagos		3				3
Guayas		3	5			8
Imbabura	1	1				2
Loja	1	3	1	1		6
Los Rios		1		1		2
Manabi	1	399	92	27	22	541
Los Rios		1		1		2
Morona Santiago		6	7	2	1	16
Napo		1	1			2
Pastaza		4	1	1		6
Pichincha	2	5				7
Santa Elena		3	3	1		7
Sto Domingo de los Tsachilas		4				4
Tungurahua		5	1			6
Zamora Chinchipe		2	2			4
Total	10	624	165	51	58	908

Table 2. Average monthly magnitude of earthquakes in Manabi and Esmeraldas.

Province	March	April	May	June	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

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 Manabi and Esmeraldas was 8.44 km and 8.07 km respectively.

The Figure (1) on the right shows the position of the earthquakes with respect to the position of the nearest tectonic plates.

Figure (2) shows the cumulative number of earthquakes over time and vertical segments proportional to their magnitude. It is observed that the magnitudes have been decreasing after April 16 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 16, 2016 is the characteristic of big events. In Figure (3), 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 16 April).

To verify the Gutenberg-Richter law, we plot in Figure (4) the number of events with magnitudes greater than a given threshold magnitude vs the threshold magnitudes and in Figure (5), 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 $a = 5.064$ ($sd = 0.039$) and $b = -0.694$($sd = 0.007$) were obtained with a p-value for both values of 2.2e-16 i.e very significant. 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

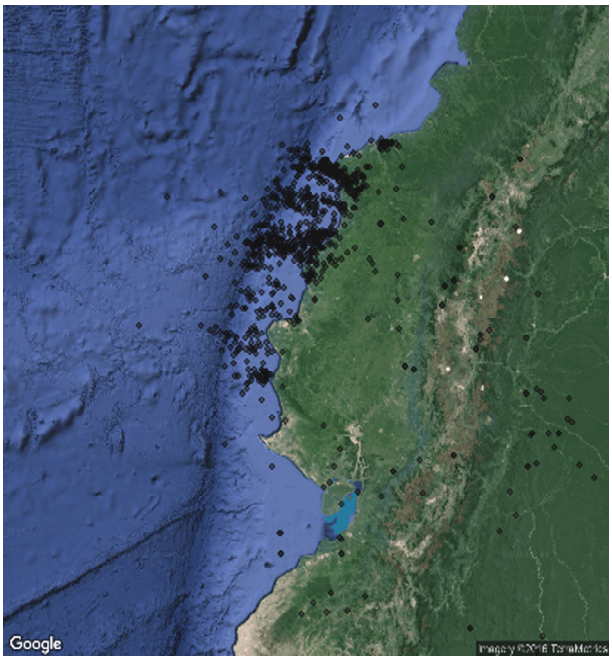


Figure 1. Earthquake distribution (Google Maps).

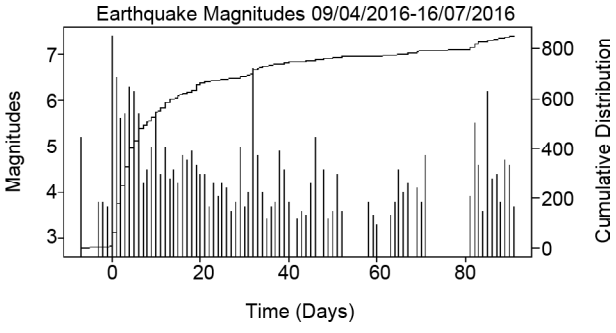


Figure 2. Cumulative number of earthquakes and mag-nitude of events over time.

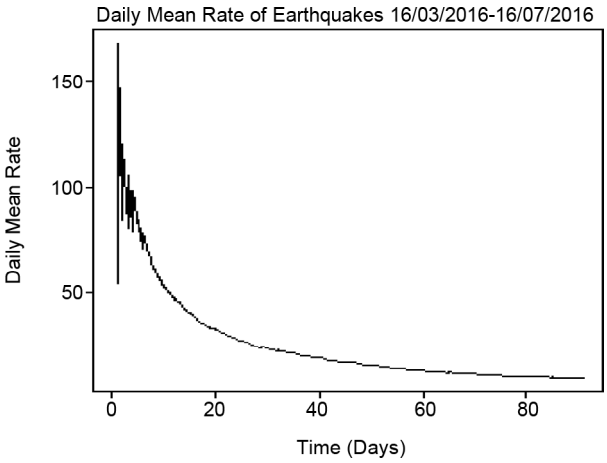


Figure 3. Average daily rate of earthquakes over time.

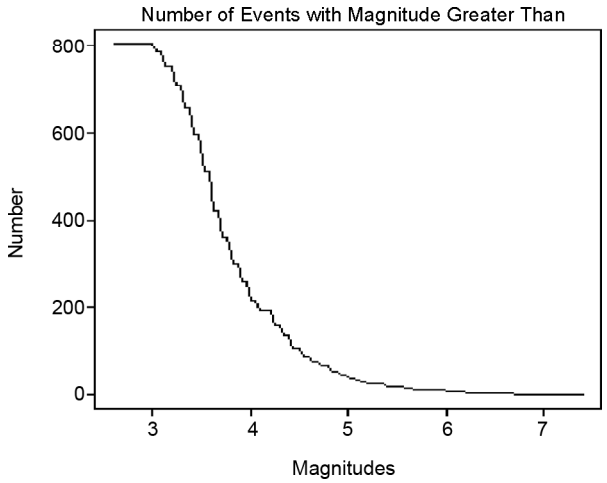


Figure 4. Number of events with magnitude greater than or equal to a given threshold magnitude as a function of that threshold magnitude.

$\beta = bln(10) = 1,598.$

By adjusting the earthquake catalog in Manabi and Esmeraldas provinces using the ETAS *R* package with 10 iterations, we obtained estimates of the background seismicity. Figure (6) 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 μ=1.0227, A=0.4789, c=0.0434, α = 0.6947, p = 1.4058, D = 0.0054, q = 1.7413, γ = 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.

Figure (7) shows the distribution of magnitude of earthquakes, latitude and longitude in time.

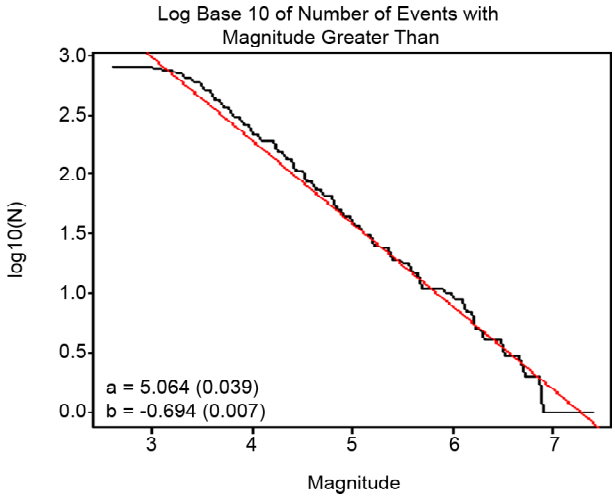


Figure 5. Decimal logarithm of the number of events with magnitude greater than or equal to a given threshold magni-tude as a function of that threshold magnitude.

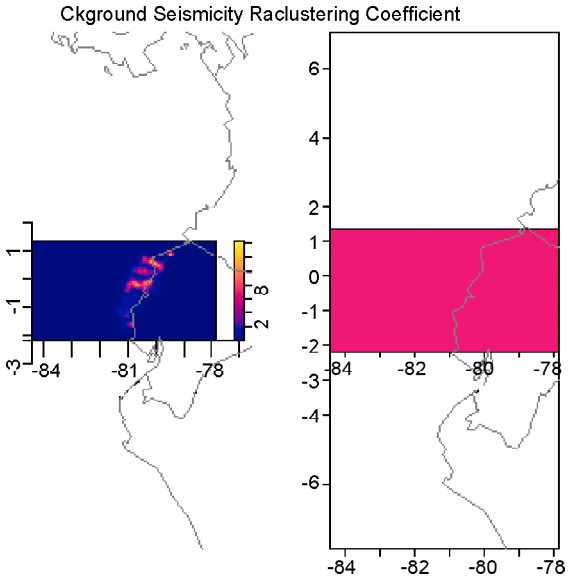


Figure 6. Stochastic declustering result in the ETAS spackage.

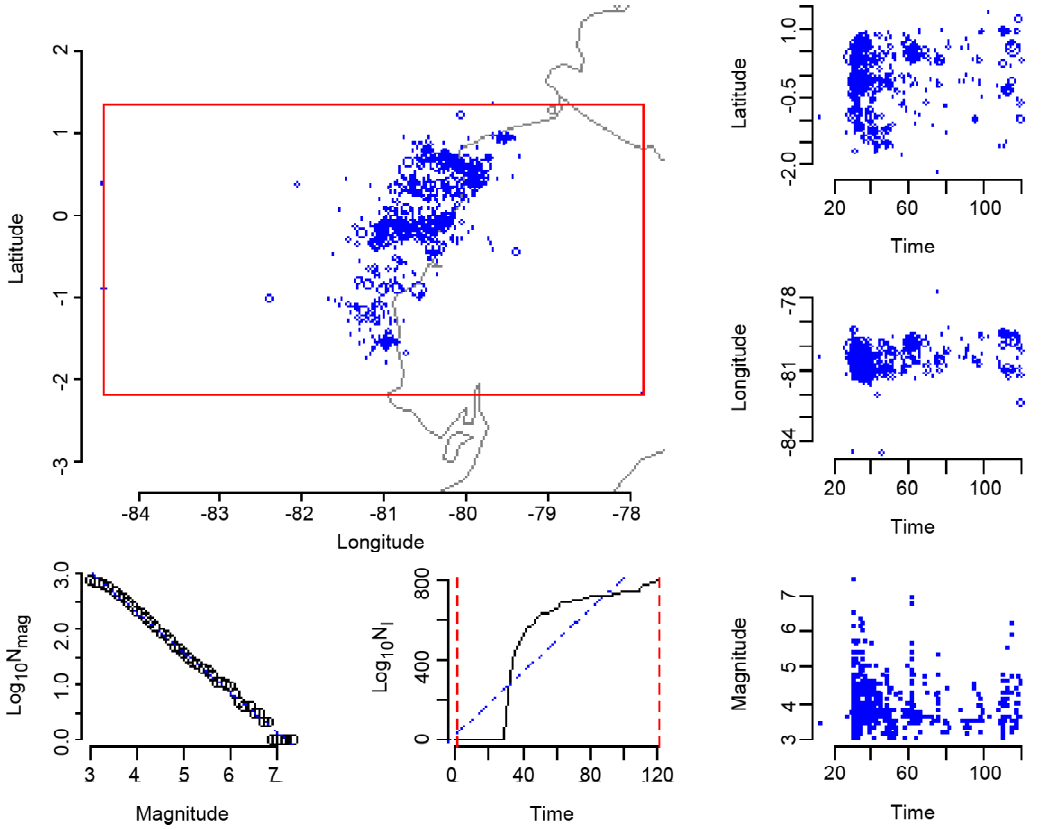


Figure 7. Distribution of magnitudes, latitudes, and longitudes of earthquakes over time.

The results of adjusting the Flp or Forward Likelihood Predictive model using the R package EtasFlp, are shown in Figures (8) to (13). Figure (8) shows the background seismic intensity and the areas within the provinces of Manabi and Esmeraldas where it is greater.

Figure (9) shows the seismicity induced or triggered by previous earthquakes. In graphs of Figures (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.

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

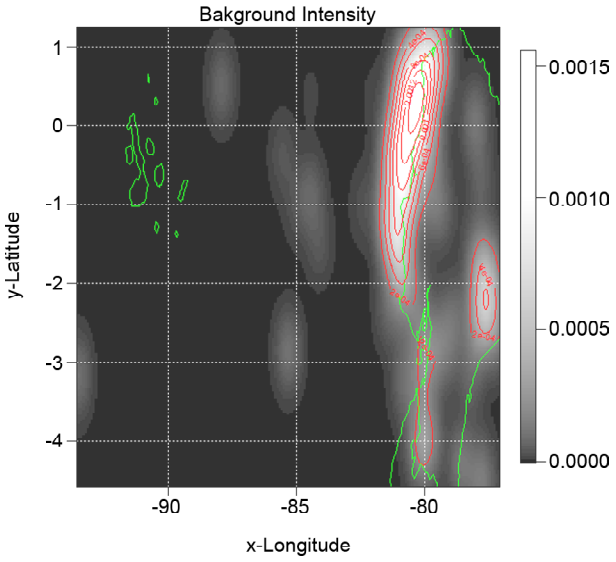


Figure 8. Background seismic intensity.

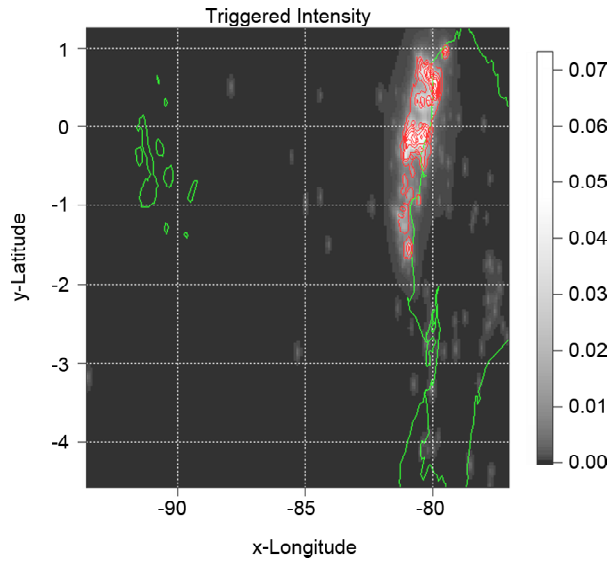


Figure 9. Triggered seismic Intensity.

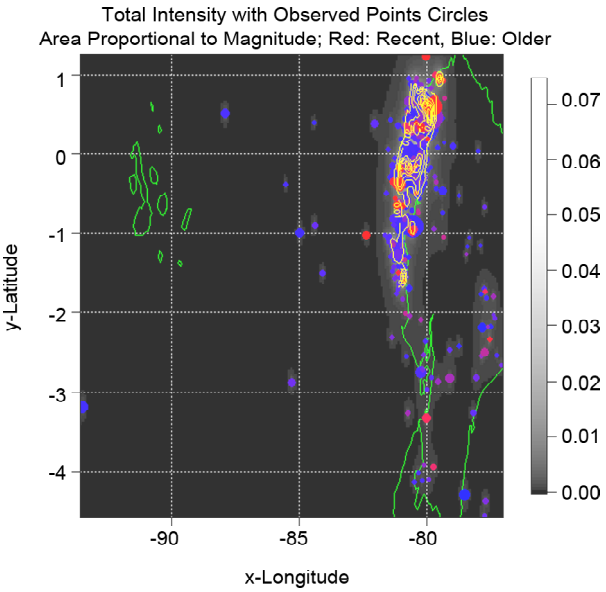


Figure 10. Total seismic intensity with observed points.

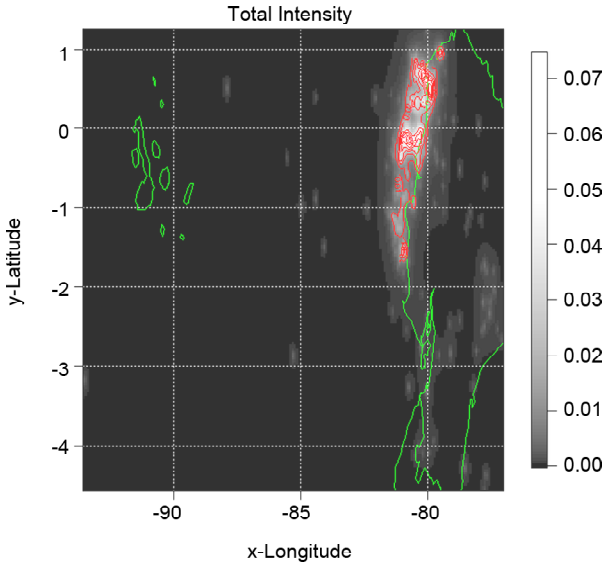


Figure 11. Total seismic intensity..

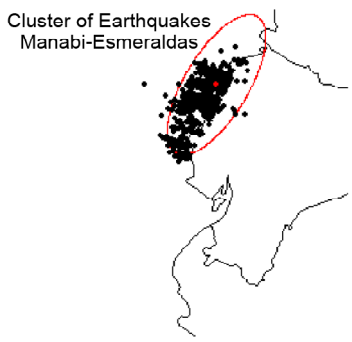


Figure 12. Cluster associated with the earthquake of magnitude 7.4, April 2016 (1).

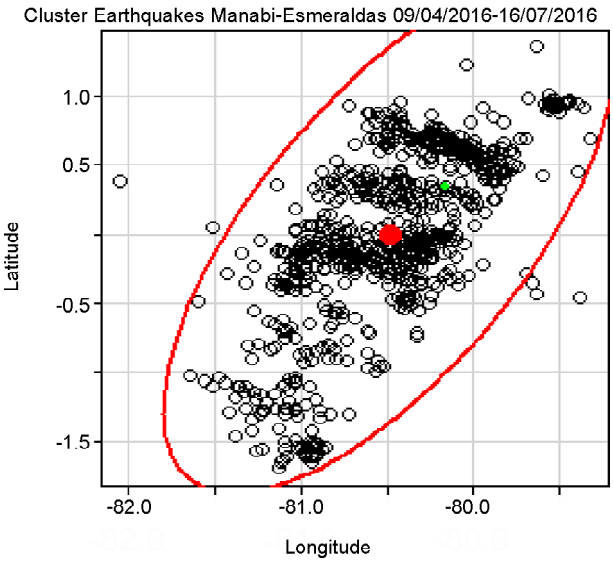


Figure 13. Cluster associated with the earthquake of magnitude 7.4, April 2016 (2).

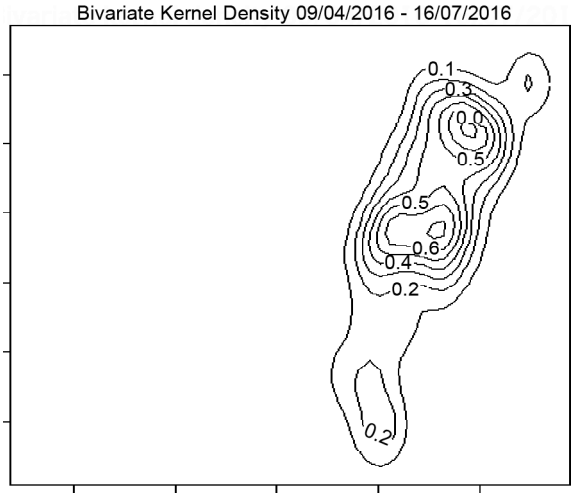


Figure 14. Estimated bivariate kernel.

ETAS models have a finite number of parameters and each with a precise meaning; therefore, we think we can take the limit as the number of parameters goes to infinity.

Figures (14) and (15) show the anisotropic distribution of earthquakes. This is done by adjusting the four bivariate models to the 810 earthquakes with epicenters near Manabi and Esmeraldas and choosing the lowest *AIC* and then the axes of the ellipse are the Eigen vectors of the normalized variance-covariance matrix.

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

$$\begin{pmatrix} 1.469 & -0.696 \\ -0.696 & 0.709 \end{pmatrix} \tag{68}$$

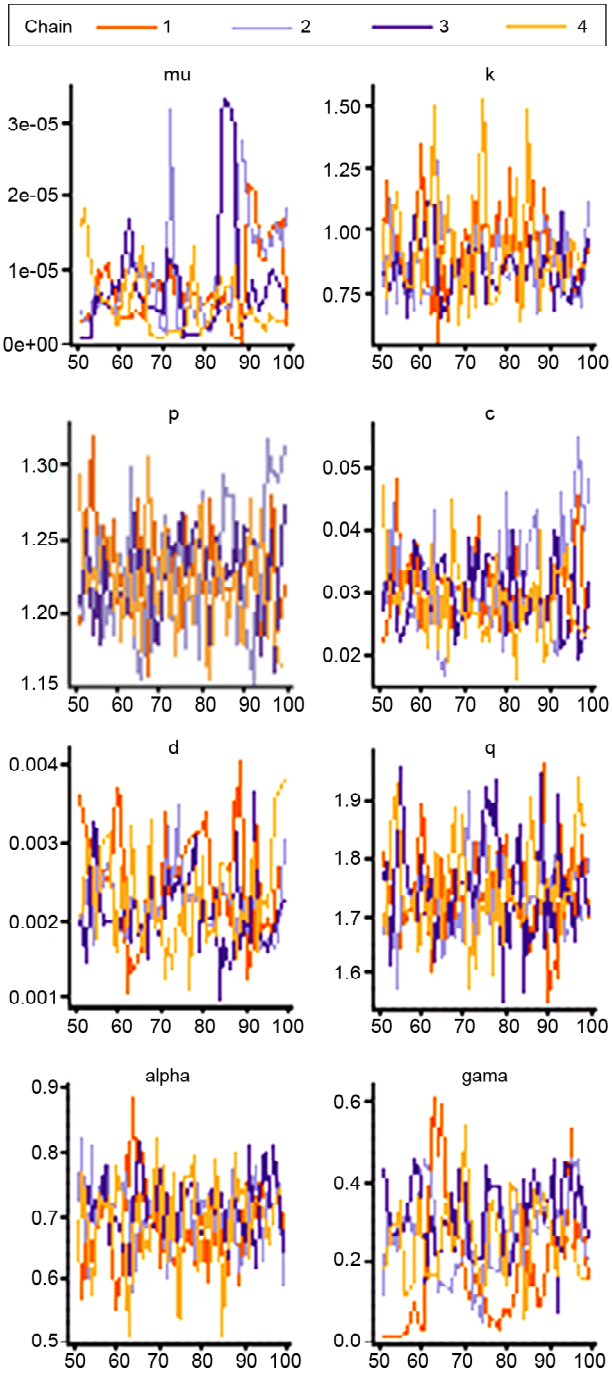


Figure 15. Chains evolution for anisotropic ETAS model with constant background seismicity.

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

The results of the parameter estimates of the models using the HMC algorithm in Rstan are shown in Tables (3) to (11). 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 hypo-central space models, $mu = 0.25 * N(t\max * pi * r^2)$ [6, 16] 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 earthquakes with the epicenter close to the provinces of Manabi and Esmeraldas, R=1.70479.

Table 3. Results Rstan temporal model without magnitudes and constant background seismicity (100 iterations, 4 chains).

	Mean	sd (mean)	sd	Confidence Interval (95%)	Neff	Rhat
mu	0.26	0.04	0.12	[0.08 ,0.52]	7	1.20
k	1.01	0.01	0.05	[0.91,1.11]	101	1.03
p	1.41	0.01	0.06	[1.30,1.52]	69	1.04
c	0.03	0.00	0.01	[0.17,0.46]	60	1.03
lp	1790.74	0.23	1.61	[1787.26,1792.78]	47	1.06

Table 4. Results Rstan temporal model without magnitudes and variable background seismicity (100 iterations, 4 chains).

	Mean	sd (mean)	sd	Confidence Interval (95%)	Neff	Rhat
mu	5.35e-05	6.6e-06	4.3e-05	[6.5e-06, 1.6e-04]	43	1.07
k	1.09	3.5e-03	5e-02	[1.01,1.20]	200	1.02
p	1.32	4.2e-03	4.3e-02	[1.24,1.41]	102	1.04
c	0.02	5.3e-04	5.8e-03	[0.14,0.37]	122	1.03
lp	1773.29	0.23	1.61	[1769.71,1774.99]	39	1.08

Table 5. Results Rstan temporal model with magnitudes and constant background seismicity (100 iterations, 4 chains).

	Mean	sd (mean)	sd	Confidence Interval (95%)	Neff	Rhat
mu	0.31	0.07	0.33	[0.10 ,1.42]	20	1.13
k	0.83	0.05	0.41	[0.30,1.69]	51	1.06
p	1.41	0.01	0.09	[1.15,1.59]	44	1.07
c	0.46	0.00	0.02	[0.02,0.13]	31	1.08
alpha	0.83	0.01	0.11	[0.64,1.04]	88	1.02
A	0.65	0.04	0.31	[0.22,1.41]	61	1.03
lp	1832.84	8.80	40.47	[1689.23,1842.83]	21	1.14

Table 6. Results Rstan temporal model with magnitudes and variable background seismicity (100 iterations, 4 chains).

	Mean	sd (mean)	sd	Confidence Interval (95%)	Neff	Rhat
mu	6.9e-05	9.6e-06	4.6e-05	[1.2e-05, 1.7e-04]	23	1.10
k	0.75	0.05	0.37	[0.25,1.60]	52	1.03
p	1.32	0.00	0.05	[1.24,1.44]	156	0.99
c	0.32	0.00	0.01	[0.02,0.05]	171	0.99
alpha	0.81	0.00	0.08	[0.65,0.95]	125	1.02
A	0.78	0.05	0.37	[0.29,1.73]	64	0.99
lp	1800.76	0.16	1.59	[1797.36,1802.91]	93	1.05

Table 7. Results Rstan spatio temporal model with isotropic clusters and constant background seismicity (100 iterations, 4 chains).

	Mean	sd (mean)	sd	Confidence Interval (95%)	Neff	Rhat
mu	0.012	0.000	0.005	[0.004 ,0.024]	88	0.99
k	0.864	0.019	0.1651	[0.572,1.205]	74	1.04
p	1.22	0.004	0.035	[1.16,1.290]	57	1.05
c	0.025	0.000	0.005	[0.015,0.035]	50	1.08
d	0.006	0.000	0.001	[0.003,0.010]	27	1.14
q	1.926	0.0017	0.145	[1.694, 2.251]	68	1.01
alpha	0.704	0.008	0.071	[0.58,0.845]	78	1.03
gamma	0.253	0.041	0.164	[0.000,0.575]	15	1.39
lp	1394.355	0.563	2.574	[1387.357,1398.070]	20	1.20

Table 8. Results Rstan spatio temporal model with anisotropic clusters and constant background seismicity (100 iterations, 4 chains).

	Mean	sd (mean)	sd	Confidence Interval (95%)	Neff	Rhat
mu	0.008	0.000	0.003	[0.002 ,0.014]	54	1.04
k	0.875	0.015	0.166	[0.604,1. 246]	119	1.014
p	1.213	0.005	0.044	[1.136,1.305]	85	1.051
c	0.028	0.000	0.008	[0.015,0.047]	77	1.042
d	0.002	0.000	0.000	[0.001,0.003]	55	1.070
q	1.782	0.006	0.081	[1.645,1.929]	140	1.017
alpha	0.711	0.005	0.064	[0.573,0.847]	139	1.002
gamma	0.304	0.017	0.129	[0.085,0.592]	59	1.049
lp	1857.145	0.186	1.842	[1852.712,1859.596]	98	0.995

Table 9. Results Rstan spatio temporal model with anisotropic clusters and variable background seismicity (100 iterations, 4 chains).

	Mean	sd (mean)	sd	Confidence Interval (95%)	Neff	Rhat
mu	6.4e-06	0.000	0.000	[7.7e-07, 1.6e-05]	52	1.082
k	0.915	0.013	0.160	[0.651. 294]	147	0.999
p	1.227	0.023	0.032	[1.164,1.297]	200	0.999
c	0.030	0.000	0.007	[0.019,0.046]	200	0.988
d	0.002	0.000	0.000	[0.001,0.004]	130	1.011
q	1.753	0.007	0.093	[1.596,1.956]	200	0.995
alpha	0.687	0.005	0.065	[0.550,0.798]	144	1.010
gamma	0.238	0.014	0.134	[0.042,0.513]	86	1.032
lp	1799.904	0.217	1.942	[1794.635,1801.923]	80	1.022

Table 10. Results Rstan hypocentral model with anisotropic clusters and constant background seismicity (100 iterations, 4 chains).

	Mean	sd (mean)	sd	Confidence Interval (95%)	Neff	Rhat
mu	2.9e-04	1.9e-05	0.000	[1e-04, 5.5e-04]	47	0.983
k	0.872	0.219	0.219	[0.577,1.335]	100	0.994
p	1.199	0.048	0.040	[1.119,1.287]	71	0.999
c	0.035	0.013	0.010	[0.018,0.056]	61	1.002
d	0.003	0.001	0.000	[0.002,0.005]	39	0.990
q	1.864	0.087	0.087	[1.690,2.040]	100	0.995
alpha	0.735	0.084	0.084	[0.584,0.892]	100	0.989
gamma	0.378	0.021	0.140	[0.144,0.631]	42	1.013
Eta	6.384	0.551	0.552	[5.385,7.377]	100	0.991
lp	-709.462	5.085	2.163	[-714.144,-706.497]	18	1.087

Table 11. Results Rstan hypocentral model with anisotropic clusters and variable background seismicity (100 iterations, 4 chains).

	Mean	sd (mean)	sd	Confidence Interval (95%)	Neff	Rhat
mu	2.9e-07	2.7e-08	1.7e-07	[7.3e-08, 6.9e-07]	40	1.090
k	0.912	0.015	0.151	[0.656,1.216]	97	0.988
p	1.199	0.003	0.032	[1.145,1.264]	150	0.993
c	0.036	0.001	0.009	[0.023,0.055]	118	1.000
d	0.003	0.000	0.001	[0.001,0.004]	90	1.001
q	1.770	0.009	0.102	[1.568,1.977]	126	0.994
alpha	0.714	0.006	0.060	[0.604,0.820]	93	0.989
gamma	0.352	0.018	0.119	[0.141,0.598]	43	1.015
Eta	6.121	0.039	0.476	[5.218,6.957]	150	0.993
lp	-719.413	0.231	1.766	[-723.034,-716.626]	58	1.047

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 use four chains with 100 iterations for each one. Tables (3) to (11) show the Rhat which is a way to measure the convergence of the chains: when it is close to 1 indicates convergence.

Figures (15) to (18) show the evolution of the chains for the anisotropic ETAS model with con- stant background seismicity.

From the following tables we observe that the models with variable background seismicity have a value of the constant considerably lower than the models with constant background seismicity. This is because for the variable background seismicity, multi-plied by the sum of the values of over the

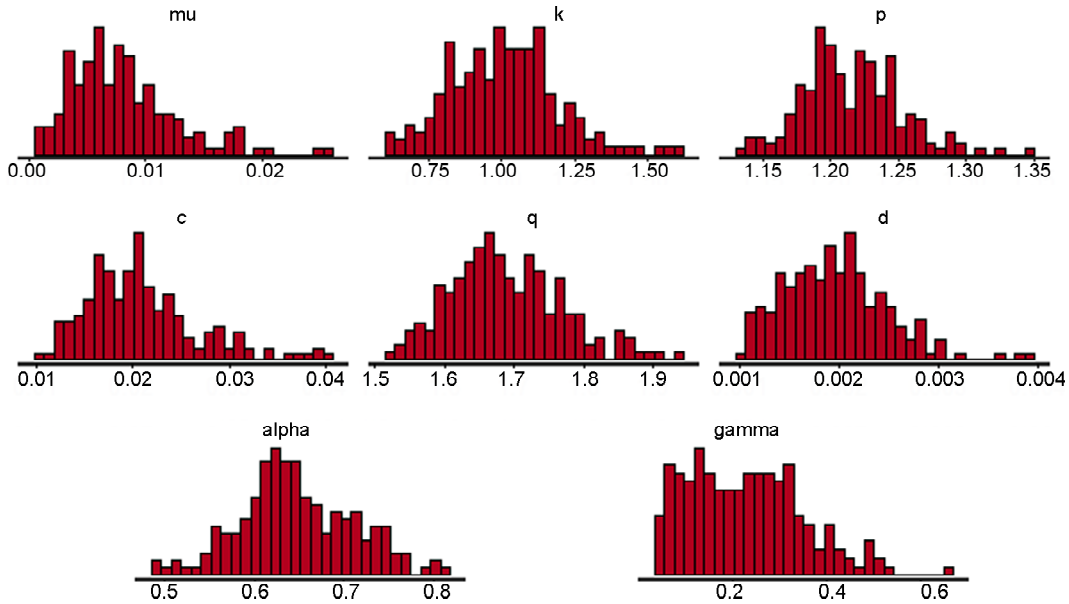


Figure 16. A posteriori parameter distributions for anisotropic ETAS model with constant background seismicity.

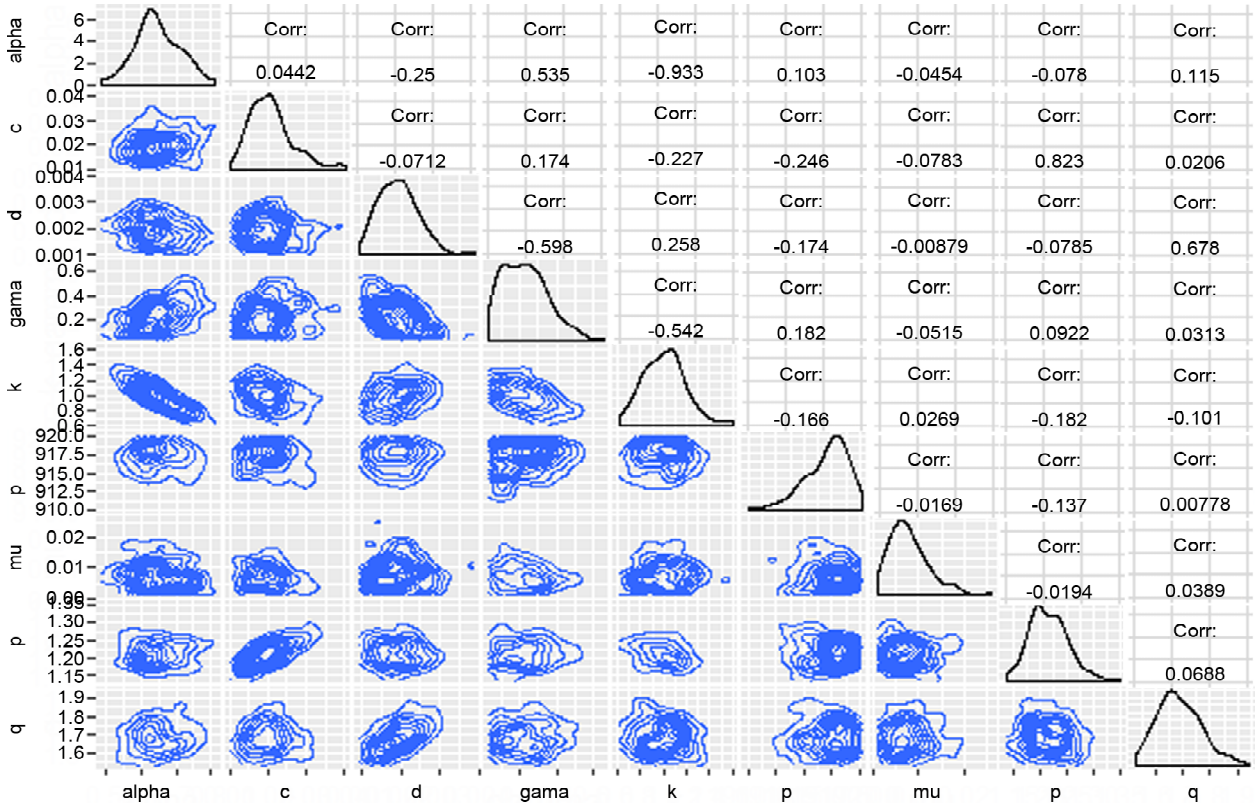


Figure 17. Correlation of parameter values in the chains for anisotropic ETAS model with constant background seismicity.

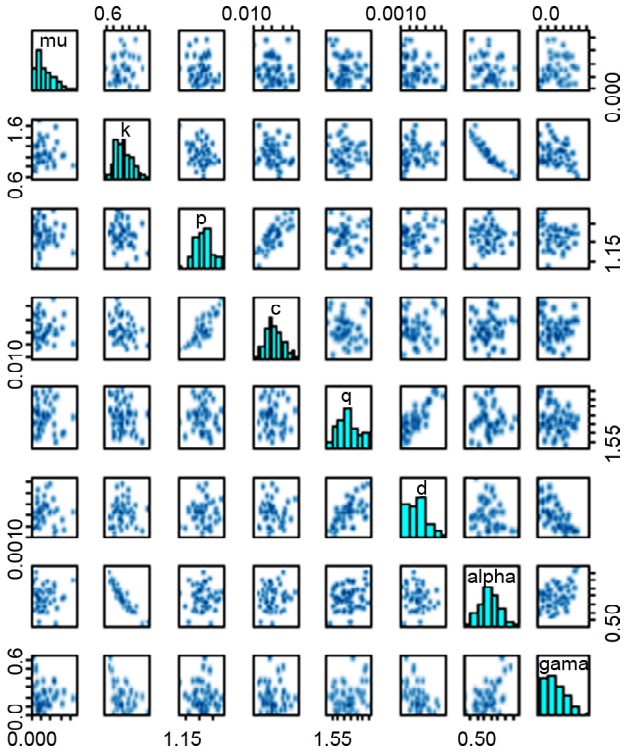


Figure 18. Correlation of parameter values in the chains for anisotropic etas model with constant background seismicity.

event points must be approximately equal to 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.

The possible cause of this is that the area is small (the radius of the minimum covering circle is 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 fit 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 two dimensions to the hypocentral model with constant background seismicity is:

$$\ln(L_{2d}) = \ln(L_{3d}) - N\ln(Z) = -709.462 + 789 * \log(30) = 1974.083 \quad (69)$$

Figure (17) shows the estimated bivariate density of earthquakes in Ecuador using the *kde* library of *R*.

$$AIC = -2 * 1974.083 + 2 * 9 = -3930.165 \quad (70)$$

(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 (19) 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 16 have a high probability of not being provoked by the background seismicity.

Table 12. ETAS models comparison.

ETAS Model	ln(L)	Int. ln(L)		# param.	AIC	int. 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

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 earth-quake of magnitude 7.4 on April 16.

Figure (20) 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. It is noted that very few events after April 16, 2016 have a high probability of being triggered by background seismicity although such events exist.

Figure (21) 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 hypo-central model with constant background seismicity, and evaluating for each tuple of parameters the seismicity rate function at the times of the events from 16/07/2016 to 11/09/2016 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 (22) 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

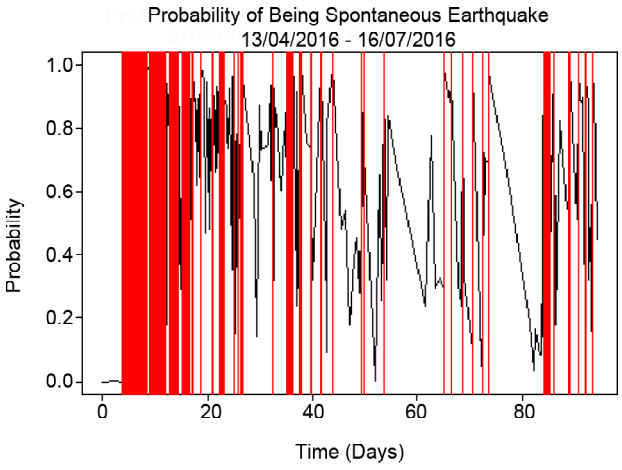


Figure 19. Probability of being a spontaneous earthquake.

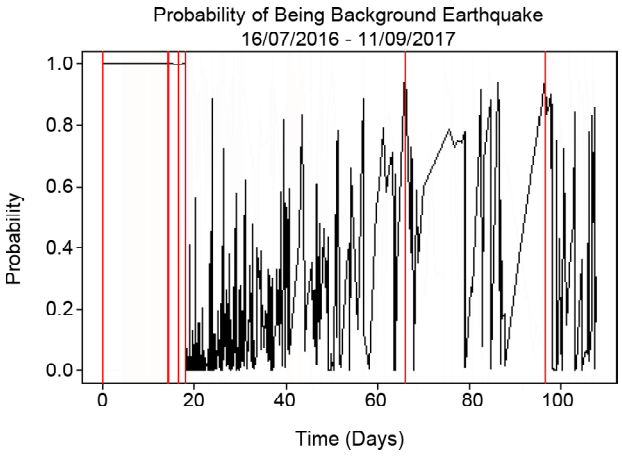


Figure 20. Probability of being background earthquake.

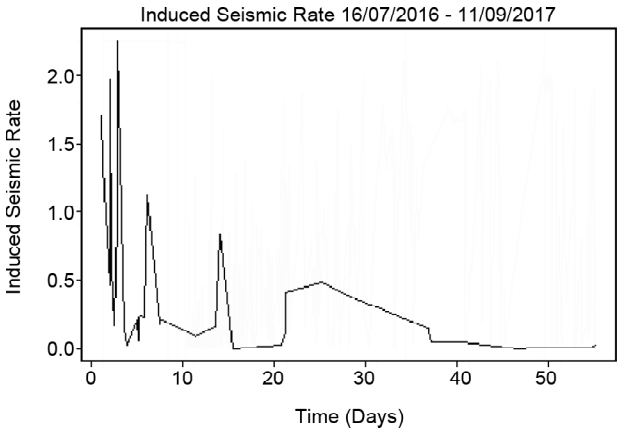


Figure 21. Induced seismic rate.

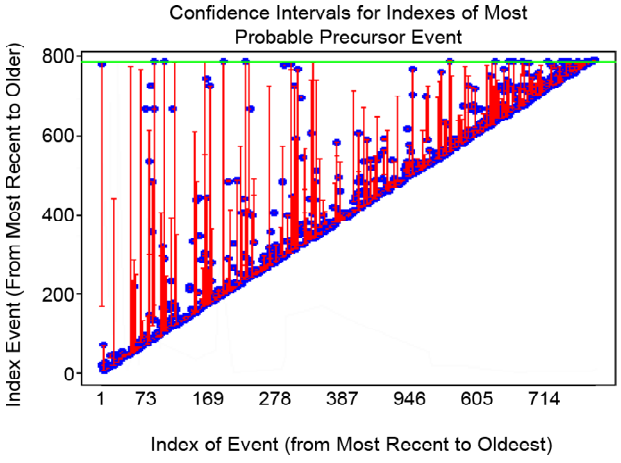


Figure 22. Indexes earthquakes more likely predecessors (fathers).

blue points), that is, these events are induced by the April 16 earthquake with probability 95%.

To demonstrate the advantages of the Bayesian approach to the frequentist approach, Table (13) shows the earthquakes most likely to be father earthquakes of the greatest number of earthquakes, taking the values of the medians of the posterior distributions of the parameters obtained by Rstan. Table (14) shows the earthquakes most likely to be triggered by the earthquake of magnitude 7.4 of April 16 by means of a frequentist approach. However, using a Bayesian approach, we can do simulations of the parameter values and for each tuple of values calculate the probability of being an earthquake induced by the magnitude 7.4 of April 16. With this we can see that out of the 16 earthquakes that the frequentist 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 16. These are shown in Table (15).

The following figures show the residual analysis

Table 13. Earthquakes with more than three direct aftershocks (descendants).

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
6.2	21/04/2016 22:03	0.18 S	80.77 W	10	Jama	8
6.6	18/05/2016 2:57	0.44 N	79.95 W	7	Muisne	7
5.7	19/04/2016 17:22	0.57 N	80.07 W	5	Muisne	7
6.1	17/04/2016 2:13	0.40 S	80.40 W	7	San Vicente	7
6	21/04/2016 22:20	0.18 S	80.88 W	4	Jama	6
5.5	07/07/2016 23:28	0.39 S	81.07 W	5	Bahia de Caraquez	5
6.1	20/04/2016 3:33	0.39 N	79.89 W	55	Muisne	5
6.5	17/04/2016 16:35	0.91 S	80.56 W	10	Jaramijo	5
4.3	16/04/2016 21:14	0.29 S	80.48 W	8	Jama	5
5.1	16/04/2016 19:11	0.07 S	80.23 W	11	Jama	5
4.4	12/07/2016 13:32	0.95 N	79.53 W	2	Esmeraldas	4
5.6	26/04/2016 16:58	0.14 S	80.99 W	10	Jama	4
5.4	21/04/2016 22:06	0.14 S	80.37 W	14	Jama	4
4.8	20/04/2016 6:23	0.11 S	80.48 W	6	Jama	4
5.6	18/04/2016 13:38	0.93 S	80.99 W	0	Manta	4
5.1	18/04/2016 8:25	0.25 S	80.42 W	9	Jama	4
4.1	17/04/2016 21:09	0.70 S	81.08 W	5	Manta	4
4.3	17/04/2016 2:28	0.06 N	80.94 W	10	Jama	4
5.2	16/04/2016 21:17	0.18 S	80.66 W	6	Jama	4

Table 14. Earthquakes most likely to be triggered by earthquake of magnitude 7.4 of April 16 (frecuentist approach).

Magn.	Date	Lat.	Long.	Depth.	Closest City	Probability
3.5	05/06/2016 2:32	0.30 N	80.09 W	5	Pedernales	0.1646961
3.4	28/05/2016 14:43	0.21 N	80.21 W	7	Pedernales	0.114522
3.1	05/05/2016 18:18	0.46 N	80.14 W	10	Muisne	0.1209538
4.5	30/04/2016 20:49	0.32 N	80.18 W	5	Pedernales	0.2839891
3	19/04/2016 8:07	0.37 N	80.08 W	10	Muisne	0.1556532
4.8	17/04/2016 19:38	0.39 N	80.24 W	5	Muisne	0.257433
3.2	17/04/2016 18:17	0.19 N	80.25 W	5	Pedernales	0.1783448
3.3	17/04/2016 0:59	0.32 N	80.16 W	8	Pedernales	0.5395906
3.5	17/04/2016 0:01	0.43 S	79.64 W	9	El Carmen	0.1781111
4.4	16/04/2016 23:04	0.53 N	79.96 W	3	Muisne	0.3284828
4.2	16/04/2016 21:23	0.28 N	80.16 W	9	Pedernales	0.4430858
3.8	16/04/2016 20:25	0.24 N	80.04 W	6	Pedernales	0.5594365
4.7	16/04/2016 19:59	0.32 N	80.22 W	5	Pedernales	0.8844682
5.7	16/04/2016 19:29	0.21 S	80.66 W	6	Jama	0.4358462
4.6	16/04/2016 19:16	0.46 S	79.38 W	10	El Carmen	0.7047577
5.1	16/04/2016 19:11	0.07 S	80.23$^{i\circ}$ W	11	Jama	0.7823534

of the models and the prediction of the accumulated number of events, evaluated at the times of the earthquakes. This is achieved through 1000 simulations of the posterior distributions of the parameters. The green vertical lines represent the first month after July 16, 2016 which is the date of the last event that was used for the estimation. As we can see the temporal models with constant and variable bottom seismicity do not differ much.

Using the model ETAS considering the magnitudes we can perform simulations of the parameters and with this we can calculate the average number of descendants 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) (Figures 23-34).

Table 15. Earthquakes with probability greater than 95% of being triggered by earthquake of magnitude 7.4 of April 16 (Bayesian approach).

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

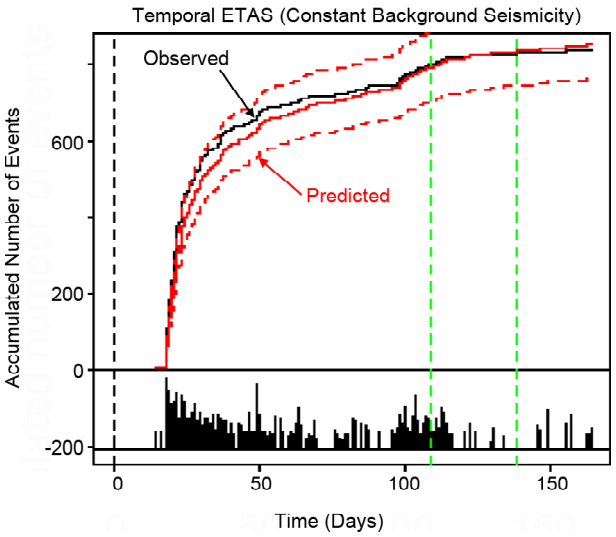


Figure 23. Acumulated number of events for temporal ETAS model with constant background seismicity.

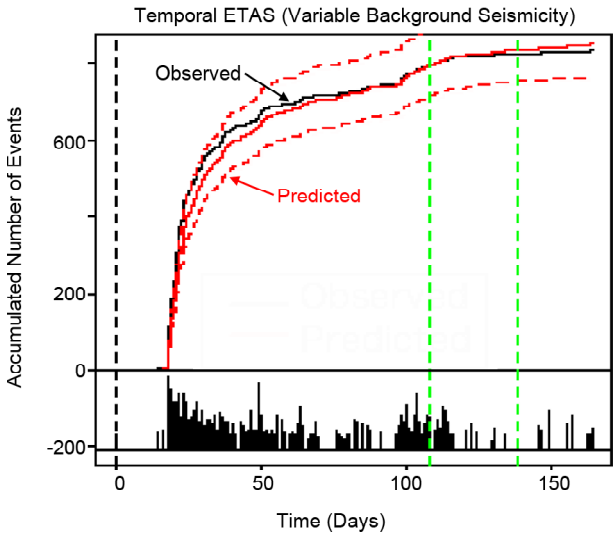


Figure 25. Acumulated number of events for temporal ETAS model with variable background seismicity.

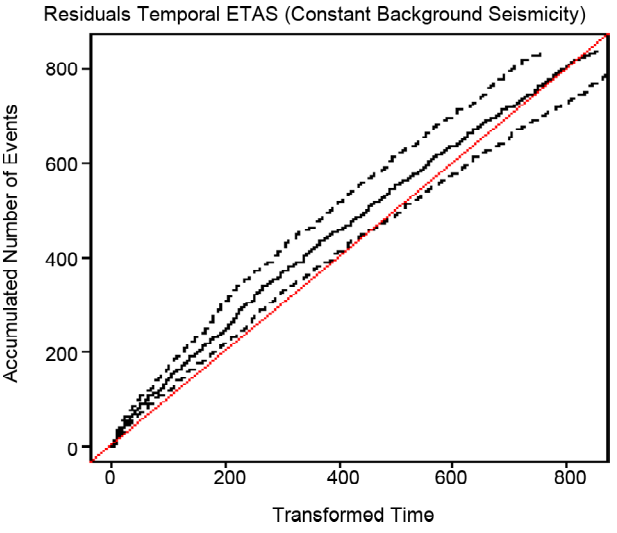


Figure 24. Residuals for ETAS temporal temporal ETAS model with constant background seismicity.

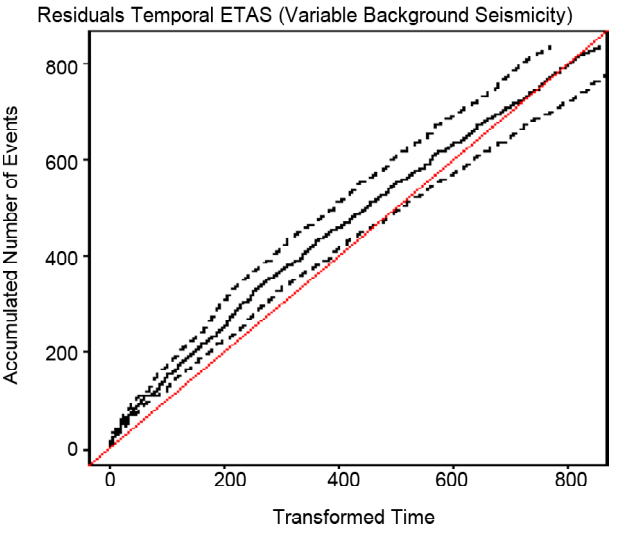


Figure 26. Residuals for temporal ETAS model with variable background seismicity.

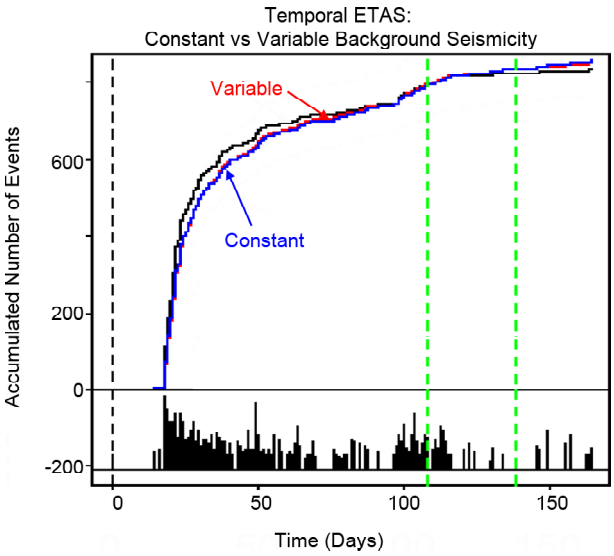


Figure 27. Acumulated number of events for temporal ETAS model with constant background seismicity vs variable background seismicity.

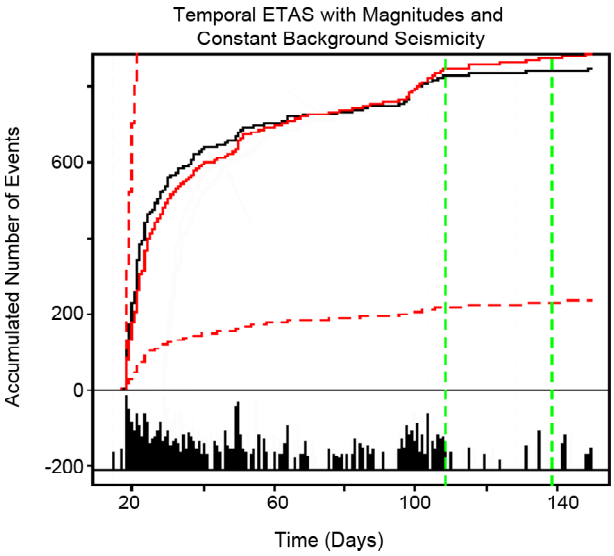


Figure 29. Acumulated number of events for temporal ETAS with magnitudes model and constant background seismicity.

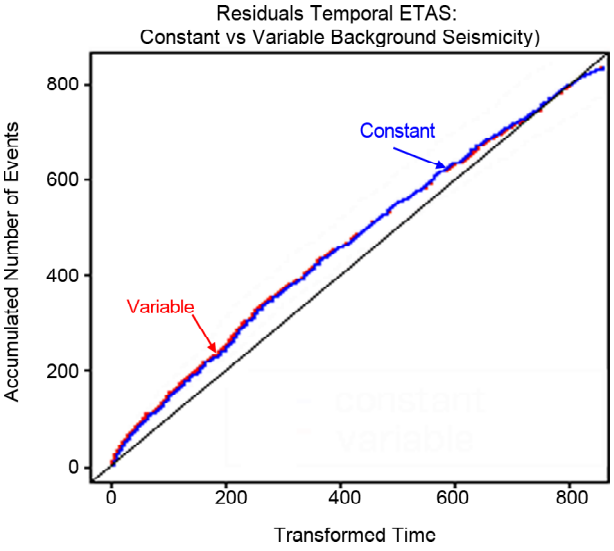


Figure 28. Residuals for temporal ETAS model with constant background seismicity vs variable background seismicity.

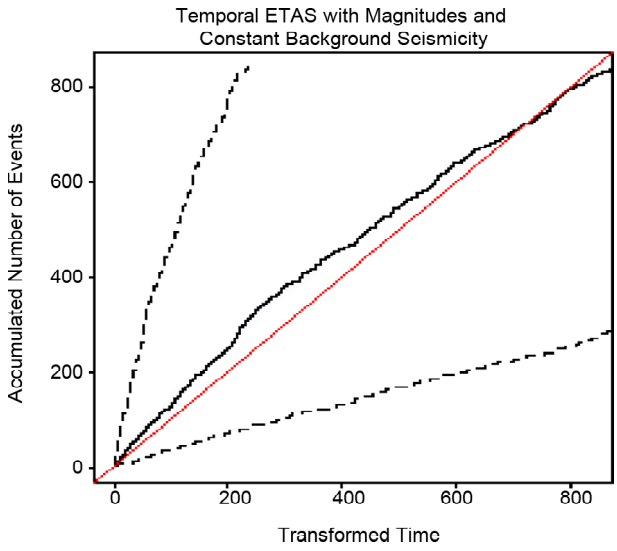


Figure 30. Residuals for temporal ETAS with magnitudes model and constant background seismicity.

Figure (27) 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 (28) to better distinguish the graphs.

It can be observed that the fit is acceptable although there are considerable differences between the model and the observed.

10. Conclusions

Bayesian analysis is a useful tool for parameters estimation by maximizing the logarithm of the likelihood. Instead of point estimates, Bayesian analysis provides a posteriori probability distributions for the parameters, allow in 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 the 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

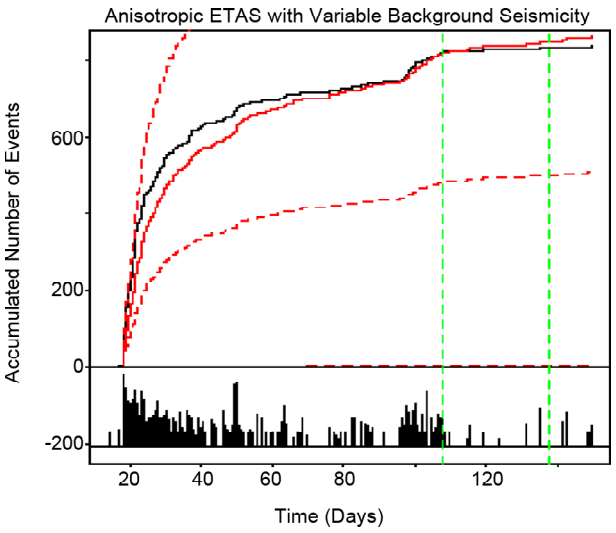


Figure 31. Acumulated number of events for spatio temporal anisotropic ETAS model and variable background seismicity.

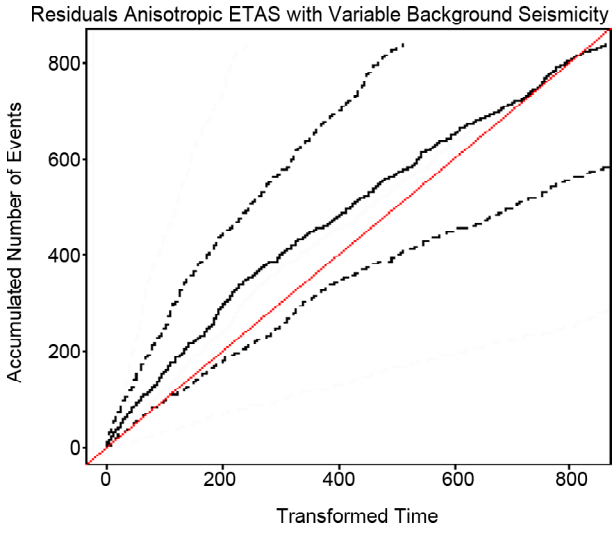


Figure 32. Residuals for spatio temporal anisotropic ETAS model and variable background seismicity.

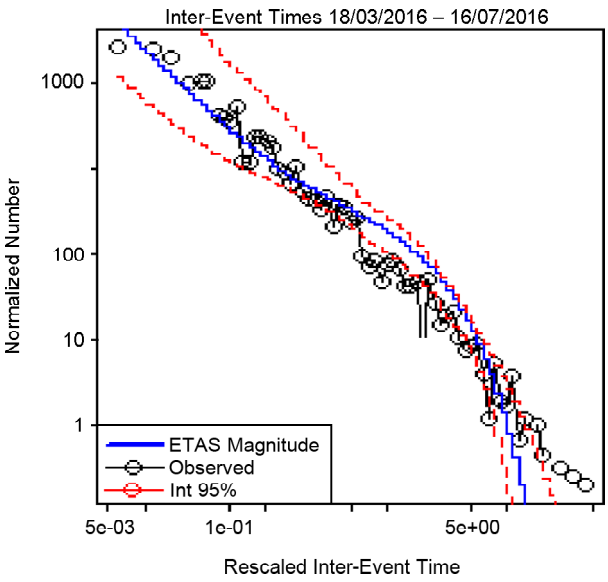


Figure 33. Time inter event modelled with Saichev-Sornette formula.

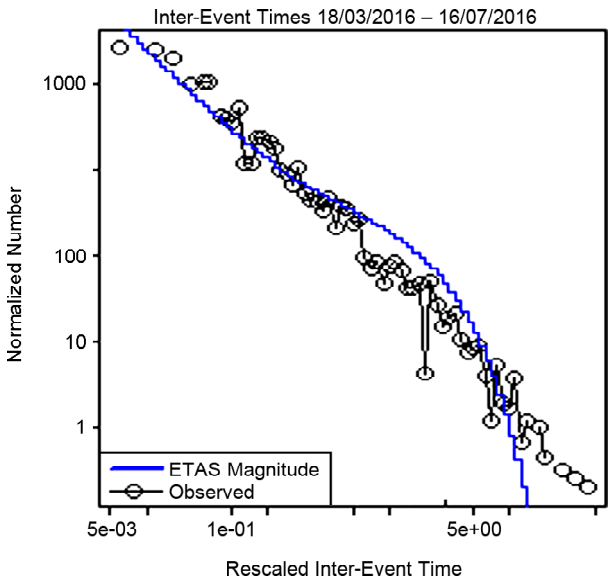


Figure 34. Time inter event modelled with Saichev-Sornette formula.

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. Collot, J-Y., Marcaillou, B., Sage, F., Michaud, F., Agudelo, W., Charvis, P., Graindorge, D., Gutscher, M., and Spence, G. (2004) Are rupture zone limits of great subduction earthquakes controlled by upper plate structures? Evidence from multichannel seismic refection data acquired across the northern Ecuador-southwest Colombia margin. *Journal of Geophysical Research*, **109**, B11103, 109: http://onlinelibrary.wiley.com/doi/10.1029/2004JB003060/references.

2. Lotto, G. and Stein, R. (2016) Ecuador earthquakes: What happened and what's next? http://temblor.net/earthquakeinsights/ecuador-earthquakes what-happened-and-what-is-next-986.

3. Cressie, N. (1993) *Statistics for Spatial Data*. Ed Wiley, New York.

4. Pieavatolo, L., Epifany, A., and Ladelli, I. (2013) Bayesian estimation for a parametric Markov renewal model applied to seismic data. *Electronic Journal of Statistics*, **8**(2).

5. Utsu, T., Ogata, Y., and Matsu'ura, R.S. (1995) The centenary of the Omori formula for a decay law of aftershock activity. *Journal of Physics of the Earth*, **43**, 1-33.

6. Ogata, Y. (1998) Space-time point process models for earthquake occurrences, annals of the institute of statistical mathematics. *Annals of the Institute of Statistical Mathematics*, **50**(2), 379-402.

7. Varini, E. and Ogata, Y. (2014) Bayesian estimation of doubly stochastic Poisson processes for detection of seismicity phases.

8. Working Group on California Earthquake Probabilities Working Group (1999) Earthquake Probabilities in the San Francisco Bay Region: 2000 to 2030 - A Summary of Findings. U.S. Geological Survey, Open-File Report 03-214, https://pubs.usgs.gov/of/2003/of03.214/WG02_OFR.03.214_Chapter5.pdf.

9. Bayraktarli, Y., Baker, J., and Faber, M. (2011) Uncertainty treatment in earthquake modeling using Bayesian probabilistic networks. *Georisk*, **5**, 44-58.

10. Omori, O. (1894) On the aftershocks of earthquakes. *Journal of the College of Science, Imperial University of Tokyo*, 7, 111-200.

11. The Headquarters for Earthquake Research Promotion (1998) Aftershock Probability Evaluation Methods. Earthquake Research Committee, Prime Minister's Office, Government of Japan.

12. Kagan, Y. and Knopoff, L. (1981) Stochastic synthesis of earthquake catalogs. *Journal of Geophysical Research Solid Earth*, **86**, 2853-2862.

13. Daley, D.J. and Vere-Jones, D. (2002) An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods. Second Edition.

14. Holden, L., Natvig, B., Sannan, S., and Bungum, H. (2000) Modeling spatial and temporal dependencies between earthquakes. *Natural and Anthropogenically Induced Hazards*.

15. Vere-Jones, D. (1995) Forecasting earthquakes and earthquake risk. *International Journal of Forecasting*, **11**, 503-538.

16. Ogata, Y. (1988) Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Statist. Assoc.*, **83**, 9-27.

17. Sornette, D. and Werner, M.J. (2005a) Apparent clustering and apparent background earthquakes biased by undetected seismicity. *Journal of Geophysical Research Solid Earth*, **110**, 243-261.

18. Sornette, D. and Werner, M.J. (2005b) Constraints on the size of the smallest triggering earthquake from the epidemic type aftershock sequence model, bath's law, and observed aftershock sequences. *Journal of Geophysical Research Solid Earth*, **110**.

19. Touati, S. (2011) *Complexity, Aftershock Sequences, and Uncertainty in Earthquake Statistics*. University Edingbourgh, Ph.D. Thesis.

20. Utsu, T. (1970) Aftershocks and earthquake statistics (1) - some parameters which characterize an aftershock sequence and their interrelations. *Journal of the Faculty of Science, Hokkaido University, Series 7, Geophysics*, **3**(3), 129-195.

21. Yamanaka, Y. and Shimazaki, K. (1990) Scaling relationship between the number of aftershocks and the size of the main shock. *J. Phys. Earth*, **38**, 305-324.

22. Ogata, Y. and Zhuang, J. (2006) Space-time ETAS models and an improved extension. *Tectonophysics*, **413**(1-2), 13-23.

23. Hernandez Vargas, N.A. (2014) Bayesiann point process modelling of earthquake occurrences. 2012. M. D. Hooman and A. Gelman. The no-u-turn sampler: Adaptively setting path lengths in Hamiltonian Montecarlo. *Journal of Machine Learning Research*, **15**(1), 1593-1623.

24. Zhuang, J., Ogata, Y., and Vere-Jones, D. (2002) Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, **97**(458).

25. Helmstetter, A., Kagan,Y., and Jackson, D. (2007) High-resolution Time-independent Grid-based Forecast for M >= 5 Earthquakes in California. *Seismological Research Letters, Seismological Society of America*, **78**(1), 78-86.

26. Tsukakoshi, Y. and Shimazaki, K. (2006) Temporal behavior of the background seismicity rate in central Japan, 1998 to mid-2003. *Tectonophysics*, **417**(1-2), 155-168.

27. Chiodi, M. and Adelfio, G. (2011) Forward likelihood-based predictive approach for space-time processes. *Environmetrics*, **22**, 749-757.

28. Zhuang, J. Guo, Y., and Zhou, S. (2015) A hypocentral version of the space-time ETAS model. *Geophysical Journal International*, DOI:10.1093/gji/ggv319.

29. Zhuang, J., Ogata, Y., and Vere-Jones, D. (2004) Diagnostic analysis of space-time branching processes for earthquakes.

30. Ross, G.J. (2016) Bayesian estimation of the etas model for earthquake occurrences. http://www.gordonjross.co.uk/bayesianetas.pd.

31. Ebrahimian, H. and Jalayer, F. (2017) Robust seismicity forecasting based on Bayesian parameter estimation for epidemiological spatio temporal aftershock clustering models. DOI: 10.1038/s41598.017.09962.z.

32. Rasmussen, J.G. (2011) Bayesian inference for hawkes processes. *Methodology and Computing in Applied Probability*, **15**(15), 1-20, http://dx.doi.org/10.1007/s11009.011.9272.5.

33. Gelman, A. (2002) 'Prior distribution'. In: *Encyclopedia of Environmetrics*, **3**, 1634-1637, http://www.stat.columbia.edu/gelman/research/published/p039-o.pdf.

34. Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D.B. (2013) Bayesian data analysis. Third edition. Chapman Hal-l/CRC Texts in Statistical Science, 3.

35. Mueller, U.K. (2012) Measuring prior sensitivity and prior informativeness in large Bayesian models. *Journal of Monetary Economics*, **59**, 581-597.

36. Barrows, D. (2016) *Estimation and Inference of Nonlinear Stochastic Time Series*. Master Thesis: McMaster University.

37. Ogata, K. (2004) *System Dynamics*. 4th Edition.

38. Duane, S., Kennedy, A.D., Pendleton, B.J., and Roweth, D. (1987) Hybrid Monte Carlo. *Physics Letters B*, **195**(2), 216-222.

39. Neal, R.M. (1993) *Probabilistic Inference Using Markov Chain Monte Carlo Methods*. Department of Computer Science, University of Toronto.

40. Brooks, S., Gelman, A., Jones, G.L., and Meng, X.L. (2011) *Handbook of Markov Chain Monte Carlo*. CRC Press.

41. Beskos, A., Pillai, N.S., Roberts, G.O., Sanz-Serna, J.M. and Stuart, A.M. (2010) The acceptance probability of the hybrid Monte Carlo method in high? Dimensional problems. *AIP Conference Proceedings 1281*, https://doi.org/10.1063/1.3498436.

42. Carpenter, B., Gelman, A., Homan, M., Lee, D., Goodrich, B., et al. (2011) Stan: A probabilistic programming language. *Journal of Statistical Software*, **20**, 1-37.

43. Beskos, A., Pillai, N., Roberts, G., Sanz-Serna, J. M., and Stuart, A. (2013) Optimal tuning of the hybrid Monte Carlo algorithm. *Bernoulli*, **19**(5A), 1501-1534, DOI: 10.3150/12-BEJ414.

44. Hoffman, M.D. and Gelman, A. (2014) The no-u-turn sampler: adaptively setting path lengths in hamiltonian Monte Carlo. *Journal of Machine Learning Research*, **15**, 1593-1623.

45. Schoenberg, F.P. (2013) Facilitated estimation of etas. *Bulletin of the Seismological Society of America*, **103**(1), 601-605.

46. Lippiello, E. Arcangelis, L.D. Giacco, F., and Godano C. (2014) Parameter estimation in the

etas model: Approximations and novel methods. *Bulletin of the Seismological Society of America*, DOI: 10.1785/0120130148.

47. Welzl, E. (1991) Smallest enclosing disks (balls and ellipsoids) occurrences. H. Maurer (Ed.), New Results and New Trends in Computer Science, Lecture Notes in Computer Science, 555, 359-37, 1991.

48. Fletcher, R. and Powell, M.J.D. (1963) A rapidly convergent descent method for minimization. *Comput. J.*, **6**, 163-168.

49. http://www.ism.ac.jp/ogata/statsei4/programm.html.c.

50. http://stackoverow.com/questions/37033535/stan-programming-foretas-model. a.

51. http://stackoverow.com/questions/37699772/stan-code-for-largedata-set.b.

52. Ogata, Y. (2011) Significant improvements of the space-time etas model for forecasting of accurate baseline seismicity. *Earth Planets Space*, **63**, 217-229.

53. Kagan, Y.Y. and Knopoff, L. (1980) Dependence of seismicity on depth. *Bulletin of the Seismological Society of America*, **70**(5).

54. Zhuang, J., Christophersen, A., Savage, M., Vere-Jones, D., Ogata, Y., and Jackson, D. (2008) Differences between spontaneous and triggered earthquakes: Their influences on foreshock probabilities. *Journal of Geophysical Research*, **113**, B11302, doi:10.1029/2008JB005579.

55. Saichev, A. and Sornette, D. (2007) Theory of earthquake recurrence times. *Journal of Geophysical Research*, **112**, B04313: doi: 10.1029/2006JB004536.

56. Saichev, A. and Sornette, D. (2006) Universal distribution of inter-earthquake times explained. *Phys. Rev. Letts.*, **97**, 078501.

57. Corral, A. (2004) Universal local versus unified global scaling laws in the statistics of seismicity. *Physica A*, **340**, 590-597.