APPROXIMATION OF BAYESIAN HAWKES PROCESSES WITH INLABRU

Francesco Serafini
School of Geosciences
University of Edinburgh
francesco.serafini@ed.ac.uk

Finn Lindgren
School of Mathematics
University of Edinburgh
Finn.Lindgren@ed.ac.uk

Mark Naylor
School of Geosciences
University of Edinburgh
Mark.Naylor@ed.ac.uk

April 2, 2022

ABSTRACT

Hawkes processes are very popular mathematical tools for modelling phenomena exhibiting a *self-exciting* behaviour. Typical examples are earthquakes occurrence, wild-fires, crime violence, trade exchange, and social network activity. The widespread use of Hawkes processes in different fields calls for fast, reproducible, reliable, easy-to-code techniques to implement such models. We offer a technique to perform approximate Bayesian inference of Hawkes process parameters based on the use of the R-package Inlabru. Inlabru, in turn, relies on the INLA methodology to approximate the posterior of the parameters. The approximation is based on a decomposition of the Hakwes process likelihood in three parts, which are linearly approximated separately. The linear approximation is performed with respect to the mode of the posterior parameters distribution, which is determined with an iterative gradient-based method. The approximation of the posterior parameters is therefore deterministic, ensuring full reproducibility of the results. The proposed technique only required the user to provide the functions to calculate the different parts of the decomposed likelihood, while the optimization is performed through the R-package Inlabru. The limitations of this approach include the functional form of the different likelihood parts, which needs to be as linear as possible with respect to the parameters of the model. Moreover, care should be taken of the numerical stability of the provided functions.

1 Introduction

Hawkes processes, or *self-exciting* processes, firstly introduced by Hawkes 1971a; Hawkes 1971b are counting processes which have proven useful in modelling the "arrivals" of some events over time when each arrival increase the probability of subsequent arrivals in its proximity. Typical applications can be found in seismology (Ogata 1988; Ogata and Zhuang 2006; Ogata 2011), crime (G. O. Mohler et al. 2011; G. Mohler 2013; G. Mohler, Carter, and Raje 2018), finance (Azizpour, Giesecke, and Schwenkler 2018; Filimonov and Sornette 2012; Hawkes 2018), disease mapping (Chiang, Liu, and G. Mohler 2022; Garetto, Leonardi, and Torrisi 2021, wildfires (Peng, Frederic Paik Schoenberg, and Woods 2005), and social network analysis (Kobayashi and Lambiotte 2016; Zhou, Zha, and Song 2013).

Hawkes processes, and more in general point processes, are counting processes, assuming value equal to the cumulative number of points recorded in a bounded spatio-temporal region. We start giving the definitions only with respect to time, they can be easily extended to include space and marking variables.

Definition 1.0.1 *A counting process $\{N(t), t \geq 0\}$ is a stochastic process assuming values in the set of non-negative intergers $\mathbb{N}_0$, such that: i) $N(0) = 0$; ii) is a right-continuous step function with unit increments; iii) $N(T) < \infty$ almost surely if $T < \infty$. Also, we define the set of events recorded before time t as the history of the process $\mathcal{H}_t$.*

We remark that the history of the process will be always assumed to be defined only through time t, and therefore, we will keep referring to it as $\mathcal{H}_t$ also in the spatio-temporal case.

before as in time $< t$, or "no later than", as in time $\leq t$? ("up to and including")

Any counting process can be defined by the probability of an arrival at t for any t. This probability is usually referred as conditional intensity:

Definition 1.0.2 *For a counting process $\{N(t), t \geq 0\}$ with history $\mathcal{H}_t$, the conditional intensity function of the process $N(t)$ is:*

$$\lambda(t|\mathcal{H}_t) = \lim_{\Delta t \to 0} \frac{\mathbb{E}[N(t+\Delta t) - N(t)]}{\Delta t}$$

Assuming that the limit exist and $\lambda(t|\mathcal{H}_t) \geq 0$, $\forall t$.

In other words, the conditional intensity is the expected infinitesimal rate at which points occur around time t. This function completely determines a counting process and plays the same role played by the density function for a random variable. For this, we will focus on this quantity for the rest of the paper.

The definition of conditional intensity can be easily extended to include a space location $\mathbf{s}$, and a marking variable m. Calling, $\mathbf{x}_t = (t, \mathbf{s}, m) \in \mathcal{X} = [0, T] \times W \times M$, where $W \subset \mathbb{R}^2$ is the space domain and $M \subset \mathbb{R}$ is the marking variable domain, the conditional intensity $\lambda(\mathbf{x}_t|\mathcal{H}_t)$ is the expected infinitesimal rate at which points occur around time t, space location $\mathbf{s}$, with marking variable around m.

Now, we can define an Hawkes process model through its conditional intensity:

Definition 1.0.3 *An Hawkes process is a counting process with conditional intensity given by:*

$$\lambda(\mathbf{x}_t|\mathcal{H}_t) = \mu(\mathbf{x}_t) + \sum_{h:t_h<t} g(\mathbf{x}_t, \mathbf{x}_{t_h}) \tag{1}$$

The conditional intensity is composed by a part $\mu(\mathbf{x}_t)$ usually referred as background rate, it does not depend on the history but only on the point $\mathbf{x}_t$. The second part represents the contribution to the intensity at the point $\mathbf{x}_t$ of the points in the history. The function $g : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^+$ is known as *excitation* or *triggering* function and measures the influence of observation $\mathbf{x}_{t_h}$ on the point $\mathbf{x}$. The definition implies that the whole history of the process is important to determine the current level of intensity, in this view, Hawkes processes can be seen as a non-Markovian extension of inhomogeneous Poisson processes. Both the background rate and the triggering function depends on a set of parameters $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^m$ which determines the properties of the Hawkes process under study (e.g. number of events per time interval, probability of certain type of events, decay of the *excitement* produce by each observation, type of clustering etc..). Our technique provides a way to have a fully-Bayesian analysis of the parameters $\boldsymbol{\theta}$

A classic problem in retrieving the posterior of the parameters $\boldsymbol{\theta}$ stems from the correlation between them. For Hawkes processes, a typical example is the correlation between the background rate $\mu(\mathbf{x}_t)$ and the parameters regulating the number of events *triggered* by past observations. In fact, it is possible to find alternative combinations of the parameters (high background, low productivity vs low background, high productivity) producing similar number of expected points. With the same rationale, it is often difficult to estimate the spatial distribution of the background rate, indeed, high level of intensity in a cluster of points may be obtained by high productivity or high background rate in the cluster's area. In this context, applying maximum likelihood techniques is not effective due to the presence of many local optima (especially with small sample sizes, Touati, Naylor, and Main 2009; Filimonov and Sornette 2012), and the time needed to compute the conditional-intensity does not scale well increasing the amount of data. In addition, problems of numerical stability usually arises with this approach (Veen and Frederic P Schoenberg 2008). The most common solution to this problem is to leverage on the branching structure of the process and on the introduction of a latent variable which *classifies* the observations as coming from the background or not. This may be done explicitly (Ogata 2011) using a declustering algorithm and fitting the parameters on the appropriate subset of data. Or implicitly, using the EM algorithm (frequentist approach Dempster, Laird, and Rubin 1977; Zipkin et al. 2016) or Markov-Chain Monte Carlo methods (MCMC, Bayesian approach, Rasmussen 2013)

We focus on the Bayesian approach, and propose a technique to approximate the posterior of the parameters using the *Integrated Nested Laplace Approximation* (INLA) method (Rue et al. 2017) through the Inlabru R-package (Bachl et al. 2019). The INLA method is an alternative to MCMC methods (Robert, Casella, and Casella 1999) for Latent Gaussian models (LGMs). The key difference between the approaches is that MCMC methods is simulation-based, while INLA uses a deterministic approximation, which makes it faster and perfectly reproducible. The INLA method was designed to help the iterative process of model building in presence of strongly correlated parameters (e.g. with structure in time/space), usually encountered in applications, for which MCMC methods are impracticable given the higher computational costs. Indeed, has proven its value in various applied fields, for example: seismology (Bayliss

et al. 2020), disease mapping (Goicoa et al. 2016; Riebler et al. 2016; Santermans et al. 2016; Schrödle and Held 2011), genetics (Opitz et al. 2016), public health (Halonen et al. 2015), ecology (Roos et al. 2015), more examples can be found in Bakka et al. 2018; Blangiardo et al. 2013; Gómez-Rubio 2020.

Our goal is to bring the advantages of using the INLA method to the Hawkes process world. The paper is structured as follows: Section 2 illustrates Hawkes process models with special focus on the form of the triggering function; Section 2 explains the likelihood decomposition used to approximate Hawkes process model; Section 3 reports details on the gradient based algorithm used in Inlabru to find the mode of the parameters; Section 4 contains a practical example using the Epidemic Type Aftershock Sequence model (ETAS, Ogata 1988), a very popular model for earthquakes occurrence; Section 5 and 6 are dedicated, respectively to the discussion and conclusions.

2 Hawkes process modelling

Hakwes process models are usually defined in terms of their conditional intensity which takes the functional form illustrated by Equation 1. The Hawkes process intensity is composed by two part, a background rate $\mu(\mathbf{x}_t)$ and an *excitation* or *triggering* function $g(\mathbf{x}_t, \mathbf{x}_{t_h})$. The background rate and the triggering function depends upon a number of parameters $\boldsymbol{\theta}$. Our objective is to provide a technique to determine the posterior distribution of $\boldsymbol{\theta}$ having observed points in $\mathcal{X} = [0, T] \times W \times M$, namely $\mathcal{H}_T = \{\mathbf{x}_{t_1}, ..., \mathbf{x}_{t_n} \in \mathcal{X}\}$. Equation 1 also shows that an Hakwes process can be thought as the sum of $n + 1$ Poisson processes, where n is the number of observations in the history of the process. One Poisson process represents the background rate and has intensity $\mu(\mathbf{x}_t)$, the others n Poisson process are each one generated by an observation $\mathbf{x}_{t_h}$ and have intensity $g(\mathbf{x}_t, \mathbf{x}_{t_h})$. Many algorithms for fitting Hawkes process models are based on this decomposition and makes use of a latent variable assigning the points to one of those $n + 1$ Poisson processes.

Marked spatio-temporal Hawkes process models are usually designed using a Hawkes process for the space-time location of the points and an independent distribution for the marking variable, the conditional intensity is given by:

$$\lambda(\mathbf{x}_t = (t, \mathbf{s}, m)|\mathcal{H}_t) = \left(\mu(\mathbf{x}_t) + \sum_{\mathbf{x}_{t_h} \in \mathcal{H}_t} g(\mathbf{x}_t, \mathbf{x}_{t_h})\right) \pi(m) \tag{2}$$

Given the independence between the process representing the space-time locations and the marking variable's distribution we focus only on the distribution of the space-time locations. The parameters of the marking variable distribution are often estimated independently and based on the observed marks only.

In this paper, we consider a spatially varying background rate which remains constant over time. This is done mainly to limit the number of modes in the likelihood and the correlation between parameters. Furthermore, we are going to consider a background rate parameterised as

$$\mu(\mathbf{x}_t) = \mu u(\mathbf{s}) \tag{3}$$

with $\mu \geq 0$ representing the number of expected background events in the area for a unit time interval, and $u(\mathbf{s})$ represents the spatial variation of the background rate and we assume it is normalized to integrate to one. Different techniques have been employed to estimate $u(\mathbf{s})$. For example, in seismology, it is common practice to estimate it independently from the parameters of the triggering function, smoothing a declustered set of observations (Ogata 2011).

The common approach to model the triggering function is to factorise it in different components representing the effect of the observations $\mathbf{x}_{t_h}$ on the evaluation point $\mathbf{x}_t$ on the different dimensions (i.e. time, space, marking variable). More formally,

$$g(\mathbf{x}_t, \mathbf{x}_{t_h}) = g_m(m_h) g_t(t - t_h) g_{\mathbf{s}}(\mathbf{s} - \mathbf{s}_h) \mathbb{I}(t > t_h) \tag{4}$$

Where, $I(t > t_h)$ is an indicator function assuming value one when the condition holds, and zero otherwise. The function $g_m(m_h)$ is the marking variable triggering function representing the effect of different levels of the marking variable (e.g. large earthquakes have stronger influence); $g_t(t - t_h)$ is the time triggering function determining the time decay of the observed point's effect, and it is usually a decreasing function of $t - t_h$; $g_{\mathbf{s}}(\mathbf{s} - \mathbf{s}_h)$ is the space triggering function which has the same role of the time triggering function but in space and is usually function of the *distance* between points (different distances may be employed). Following this decomposition, also the parameter vector $\boldsymbol{\theta}$ can be decomposed in $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(\mu)}, \boldsymbol{\theta}^{(m)}, \boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{(\mathbf{s})})$, where $\boldsymbol{\theta}^{(\mu)}$ represents the parameters of the background

Table 1: Typical choices of time and space triggering functions

Name	function	parameters
	Time triggering	
Exponential	$\beta e^{-\alpha(t-t_h)}$	$\alpha, \beta >\geq 0$
Power Law	$k\left(1+\frac{t-t_h}{c}\right)^{-p}$	$k \geq 0, c > 0, p > 1$
	Space triggering	
Gaussian	$\det(2\pi\Sigma)^{-1/2} e^{-\frac{1}{2}(\mathbf{s}-\mathbf{s}_h)^T\Sigma^{-1}(\mathbf{s}-\mathbf{s}_h)}$	Σ positive semi-definite
Power Law	$(1+\frac{d(\mathbf{s},\mathbf{s}_h)}{\gamma})^{-q}$	$\gamma > 0, q > 1$

rate, and $\boldsymbol{\theta}^{(m)}$, $\boldsymbol{\theta}^{(t)}$, $\boldsymbol{\theta}^{(\mathbf{s})}$ represent, respectively, the parameters of the magnitude, time and space triggering functions. We call $J_\mu, J_m, J_t, J_\mathbf{s}$ the set of indexes indicating, respectively, the position of the background rate, marking variable triggering function, time triggering function, and space triggering function parameters inside $\boldsymbol{\theta}$, so we can write $\boldsymbol{\theta}_\mu = \{\theta_j : j \in J_\mu\}$. This notation will be particularly useful in Section 3.

Table 1 reports some of the typical choices of the space-time triggering functions. Many modifications of these functions are used in real-data applications. For example, we can imagine a different time or space effect for different levels of the marking variable. In seismology, it is common to consider a magnitude dependent space triggering function representing the fact that earthquakes with large magnitudes affect wider areas. Another modification usually found in applications is to consider the normalized version of the reported functions to ensure they integrate to one over the (respective) domain.

As explained in Laub, Lee, and Taimre 2021, the choice of the triggering function is crucial to the reliability and stability of any estimation procedure for Hawkes processes parameters. For example, many techniques use triggering functions normalized to intergrate to one over an infinite domain. For the approximation illustrated in this paper, we recommend to use functions as linear as possible with respect to the parameters and usually the unnormalised version works best. This is because, as described in details in the next Section, we will linearly approximate the log-intensity which is a function of the linearised triggering function, and therefore, the more linear the function is the better the approximation will be.

3 Hawkes process likelihood approximation

The general point process model log-likelihood is, having observed $\mathcal{H}_T = \{\mathbf{x}_{t_1}, ..., x_{t_n}\}$:

$$\mathcal{L}(\boldsymbol{\theta}|\mathcal{H}_T) = -\Lambda(\mathcal{X}|\mathcal{H}_T) + \sum_{h=1}^{n} \log \lambda(\mathbf{x}_{t_h}|\mathcal{H}_{t_h}) \tag{5}$$

where,

$$\Lambda(\mathcal{X}|\mathcal{H}_T) = \int_{\mathcal{X}} \lambda(\mathbf{x}_t|\mathcal{H}_t) d\mathbf{x}_t \tag{6}$$

is the integrated intensity corresponding to the expected number of points in $\mathcal{X}$. The integrated intensity can be decomposed using the branching structure of Hakwes processes, indeed, we can think to the expected number of points in an area as the expected number of background points plus the expected number of points generated by each observation. Formally,

$$\Lambda(\mathcal{X}|\mathcal{H}) = \Lambda_0(\mathcal{X}) + \sum_{h=1}^{n} \Lambda_h(\mathcal{X}) \tag{7}$$

where,

$$\Lambda_0(\mathcal{X}) = \int_{\mathcal{X}} \mu(\mathbf{x}_t) d\mathbf{x}_t = (T_2 - T_1)\mu \tag{8}$$

is the integrated background rate, the last equation holds only if the background rate follows the definition in Equation 3. The other quantity is given by

$$\Lambda_h(\mathcal{X}) = \int_{\mathcal{X}} g(\mathbf{x}_t, \mathbf{x}_{t_h}) d\mathbf{x}_t = g_m(m_h) \int_{\max(T_1, t_h)}^{T_2} \int_W g_t(t - t_h) g_{\mathbf{s}}(\mathbf{s} - \mathbf{s}_h) dt d\mathbf{s} \tag{9}$$

and is interpreted as the number of expected points generated by the observation $\mathbf{x}_{t_h}$. The last equation holds only if we use Equation 4 to define the triggering function.

The log-likelihood can be decomposed in three parts:

$$\mathcal{L}(\boldsymbol{\theta}) = -\Lambda_0(\mathcal{X}) - \sum_{h=1}^{n} \Lambda_h(\mathcal{X}) + \text{SL}(\mathcal{H}_T) \tag{10}$$

The expected number of background events $\Lambda_0(\mathcal{X})$, the expected number of generated events $\sum_h \Lambda_h(\mathcal{X})$, and the sum of the log-intensities $\text{SL}(\mathcal{H}_T) = \sum_h \log \lambda(\mathbf{x}_h | \mathcal{H}_T)$.

Our technique is based on approximating these three components separately. The approximation is such that the value of the posterior is exact at the mode $\boldsymbol{\theta}^*$, and the degree of accuracy decay as we move from there. Specifically, we perform a linear approximation of $\log \Lambda_0(\mathcal{X})$, $\log \Lambda_h(\mathcal{X})$, and $\log \lambda(\mathbf{x}_h)$, for $h = 1, ..., n$.

The next subsections illustrate the approximation of the different log-likelihood components. The last subsection reports some details on the iterative algorithm used to determine the mode of the posterior distribution around which the approximation is performed. For all of them, we will make explicit the dependence of the log-likelihood components from $\boldsymbol{\theta}$ and omit dependence from the domain $\mathcal{X}$, formally, $\Lambda(\mathcal{X}) = \Lambda(\mathcal{X}, \boldsymbol{\theta}) = \Lambda(\boldsymbol{\theta})$

3.1 Part I - Expected Number of background events

We approximate the integrated background rate using a linear approximation of its logarithm. Namely,

$$\tilde{\Lambda}_0(\boldsymbol{\theta}) = \exp\{\overline{\log \Lambda}_0(\boldsymbol{\theta}, \boldsymbol{\theta}^*)\} \tag{11}$$

where,

$$\overline{\log \Lambda}_0(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \log \Lambda_0(\boldsymbol{\theta}^*) + \frac{1}{\Lambda_0(\boldsymbol{\theta}^*)} \sum_{j=1}^{m} (\theta_j - \theta_j^*) \frac{\partial}{\partial \theta_j} \Lambda_0(\boldsymbol{\theta}) \Big|_{\boldsymbol{\theta} = \boldsymbol{\theta}^*} \tag{12}$$

This approach is particularly convenient if the background rate has the form reported by Equation 3. The only parameter to estimate using this approximation is $\mu \geq 0$. Changing parameter to $\theta_\mu = \log \mu$, we have two huge advantages. First, $\theta_\mu \in (-\infty, \infty)$ is a free-constraint parameter, and second, the logarithm of the expected number of background events is linear in θ_μ, which means that there will be no approximation at this step and this components will be exact for any value of θ_μ.

3.2 Part II - Expected Number of triggered events

We start the approximation of the expected number of triggered events by considering the expected number of events triggered by a single observations $\mathbf{x}_h$. This is given by Equation 9. Considering a partition of the space $\mathcal{X}$, namely $b_1, ..., b_B$ such that $\bigcup_i b_i = \mathcal{X}$ and $b_j \bigcap b_i = \emptyset$, $\forall i \neq j$, we can write:

$$\Lambda_h(\boldsymbol{\theta}) = \sum_{i=1}^{B} \int_{b_i} g(\mathbf{x}, \mathbf{x}_h) d\mathbf{x} = \sum_{i=1}^{B} \Lambda_h(b_i, \boldsymbol{\theta}) \tag{13}$$

We approximate the above quantity linearly approximating the logarithm of the elements of the summation. This increase the computational time and memory required for the algorithm but it provides a much better approximation than consider only one bin. More formally,

$$\tilde{\Lambda}_h(\boldsymbol{\theta}) = \sum_{i=1}^{B} \exp\{\overline{\log \Lambda}_h(b_i, \boldsymbol{\theta}, \boldsymbol{\theta}^*)\} \tag{14}$$

Where $\overline{\log \Lambda}_h(b_i, \boldsymbol{\theta}, \boldsymbol{\theta}^*)$ is the linear approximation with respect to the posterior mode of the expected number of generated events by the observation $\mathbf{x}_h$ in the area b_i and has the same form of Equation 12.

Assuming that we are dealing with a spatio-temporal marked Hawkes process model with triggering function given by Equation 4 and bins partitioning only the time domain such that $b_i = [t_{i-1}, t_i) \times W$ for $i = 1, ..., B$ and $t_i < t_j \forall i < j$ and $t_0 = \max(T_1, t_h)$ and $t_B = T_2$, we have that:

$$\begin{aligned}\Lambda_h(b_i, \boldsymbol{\theta}) &= g_m(\mathbf{x}_{t_h}, \boldsymbol{\theta}^{(m)}) \left(\int_{\max(T_1, t_h)}^{T_2} g_t(t - t_h, \boldsymbol{\theta}^{(t)}) dt \right) \left(\int_W g_{\mathbf{s}}(\mathbf{s} - \mathbf{s}_h, \boldsymbol{\theta}^{(\mathbf{s})}) d\mathbf{s} \right) \\ &= g_m(m_h, \boldsymbol{\theta}^{(m)}) I_t(\boldsymbol{\theta}^{(t)}) I_{\mathbf{s}}(\boldsymbol{\theta}^{(\mathbf{s})}) \end{aligned} \tag{15}$$

where $I_t(\boldsymbol{\theta}^{(t)})$ and $I_{\mathbf{s}}(\boldsymbol{\theta}^{(\mathbf{s})})$ are, respectively, the integral of the time and space triggering function. The derivative of the logarithm of $\Lambda_h(b_i, \boldsymbol{\theta})$ with respect to $\theta_j \in \boldsymbol{\theta}$ is given by

$$\frac{\partial}{\partial \theta_j} \log \Lambda_h(b_i) = \begin{cases} \frac{\partial}{\partial \theta_j} \log g_m(m_h), & \text{if } j \in J_m \\ \frac{\partial}{\partial \theta_j} \log I_t, & \text{if } j \in J_t \\ \frac{\partial}{\partial \theta_j} \log I_{\mathbf{s}}, & \text{if } j \in J_{\mathbf{s}} \end{cases} \tag{16}$$

Therefore, the accuracy of the approximation depends on *how much* linear the functions $g_m(\cdot), I_t(\cdot), I_{\mathbf{s}}(\cdot)$ are with respect the parameters $\boldsymbol{\theta}$. In the case of normalized triggering functions, we have $\Lambda_h(\mathcal{X}) = g_m(m_h)$. This means that, on one hand, we don't need to split the integral in different bins saving computational time and memory; on the other hand, the information on the parameters $\theta_j \in \boldsymbol{\theta}^{(t)} \bigcup \boldsymbol{\theta}^{(\mathbf{s})}$ provided by this likelihood component is lost. Also, the normalized triggering functions tends to be *less linear* than the corresponding unnormalized versions and this is crucial for the approximation of the sum of log-intensities

The divisions in bins is essential for the accuracy of the approximation and the ability to converge of the algorithm. Different binning strategies can be employed, and their performance depends on the form of the triggering function, for example, the time triggering function represents the time-decay of the influence of an observations on the intensity, we expect it to be a monotonic decreasing function of the time difference and, therefore, a convenient strategy would be to consider a denser partition around zero and larger bins far from it where the function flattens. On the other hand, considering too many bins may lead to numerical stability issues and longer computational time and computer memory needed. We suggest to tailor the binning strategy on the problem at hand.

3.3 Part III - Sum of log-intensities

For the sum of log-intensities calculated at the observed points we simply consider the linear approximation of the elements of the summation, namely

$$\tilde{\mathrm{SL}}(\mathcal{H}) = \sum_{h=1}^{n} \overline{\log \lambda}(\mathbf{x}_{t_h}, \boldsymbol{\theta}, \boldsymbol{\theta}^*) \tag{17}$$

where, omitting the dependence from $\mathbf{x}_{t_h}$,

$$\overline{\log \lambda}(\mathbf{x}_h, \boldsymbol{\theta}, \boldsymbol{\theta}^*) = \log \lambda(\boldsymbol{\theta}^*) + \frac{1}{\lambda(\boldsymbol{\theta}^*)} \sum_{j=1}^{m} (\theta_j - \theta_j^*) \frac{\partial}{\partial \theta_j} \lambda(\boldsymbol{\theta}) \Bigg|_{\boldsymbol{\theta} = \boldsymbol{\theta}^*} \tag{18}$$

which is the same as Equation 12.

Assuming to be interested in a spatio-temporal marked Hawkes process model, with background rate specified by Equation 3, considering $u(\mathbf{s})$ known for any $\mathbf{s} \in W$, and triggering function specified by Equation 4, the conditional intensity is given by:

$$\lambda(\mathbf{x}_{t_h}|\mathcal{H}_{t_h}) = \mu u(\mathbf{s}_h) + \sum_{k:\mathbf{x}_k \in \mathcal{H}_{t_h}} g_m(m_k) g_t(t_h - t_k) g_{\mathbf{s}}(\mathbf{s}_h - \mathbf{s}_k) \tag{19}$$

which has derivative with respect to $\boldsymbol{\theta}$ equal to

$$\frac{\partial}{\partial \theta_j}\lambda(\mathbf{x}t_h) = \begin{cases} u(\mathbf{s}_h), & \text{if } \theta_j = \mu \\ \sum_k g_t(t_h - t_k) g_{\mathbf{s}}(\mathbf{s}_h - \mathbf{s}_k) \frac{\partial}{\partial \theta_j} g_m(m_k), & \text{if } j \in J_m \\ \sum_k g_m(m_k) g_{\mathbf{s}}(\mathbf{s}_h - \mathbf{s}_k) \frac{\partial}{\partial \theta_j} g_t(t_h - t_k), & \text{if } j \in J_t \\ \sum_k g_m(m_k) g_t(t - t_k) \frac{\partial}{\partial \theta_j} g_{\mathbf{s}}(\mathbf{s}_h - \mathbf{s}_k), & \text{if } j \in J_{\mathbf{s}} \end{cases} \tag{20}$$

The above expression indicates that accuracy of the approximation depends on the linearity of the different triggering functions components.

3.4 Full approximation and Inlabru implementation

Putting all together, the Hawkes process log-likelihood approximation used by our technique is:

$$\begin{aligned} \tilde{\mathcal{L}}(\boldsymbol{\theta}, \boldsymbol{\theta}^*) &= -\tilde{\Lambda}_0(\boldsymbol{\theta}, \boldsymbol{\theta}^*) - \sum_{h=1}^{n}\sum_{i=1}^{B} \tilde{\Lambda}_h(b_i, \boldsymbol{\theta}, \boldsymbol{\theta}^*) + \tilde{SL}(\mathcal{H}, \boldsymbol{\theta}, \boldsymbol{\theta}^*) \\ &= -\exp\{\overline{\log \Lambda}_0(\boldsymbol{\theta}, \boldsymbol{\theta}^*)\} - \sum_{h=1}^{n}\sum_{i=1}^{B} \exp\{\overline{\log \Lambda}_h(b_i, \boldsymbol{\theta}, \boldsymbol{\theta}^*)\} + \sum_{h=1}^{n} \overline{\log \lambda}(\mathbf{x}_{t_h}, \boldsymbol{\theta}, \boldsymbol{\theta}^*) \end{aligned} \tag{21}$$

The approximation is performed with respect to the mode of the posterior distribution $\boldsymbol{\theta}^*$, which is determined by an iterative algorithm. The algorithm starts from a linearisation point $\boldsymbol{\theta}_0^*$ (which can be provided by the user), finds the mode of the linearised (wrt $\boldsymbol{\theta}_0^*$) posterior using the INLA approximation, namely $\overline{\boldsymbol{\theta}}_1^*$, the value of the linearisation point is updated to $\boldsymbol{\theta}_1^* = \alpha\overline{\boldsymbol{\theta}}_1^*$, where the scaling α is determined by the line search method described here [REF]. This process is repeated until, for each parameter, the difference between two consecutive linearization points, is less than 1% of the posterior standard deviation.

The present model is implemented in Inlabru combining three Poisson models on different dataset. The reference to a Poisson model is merely artificial and used for computational purposes, it has not any specific meaning. We are going to use this construct because it is easy to implement and INLA has the special feature of allowing Poisson models with exposure equal zero (which are improper). A generic Poisson model for counts $c_i, i = 1, ..., n$ observed at locations $\mathbf{x}_i, i = 1, ..., n$ with exposure $E_1, ..., E_n$ with log-intensity $\log \lambda_P(\mathbf{x}) = f(\mathbf{x}, \boldsymbol{\theta})$, in Inlabru has log-likelihood given by:

$$\mathcal{L}_P(\boldsymbol{\theta}) \propto -\sum_{i=1}^{n} \exp\{\overline{f}(\mathbf{x}_i, \boldsymbol{\theta}, \boldsymbol{\theta}^*)\} * E_i + \sum_{i=1}^{n} \overline{f}(\mathbf{x}_i, \boldsymbol{\theta}, \boldsymbol{\theta}^*) * c_i \tag{22}$$

Each Hawkes process log-likelihood component is approximated using one surrogate Poisson model with log-likelihood given by Equation 22 and appropriate choice of counts and exposures data. Table 2 reports the approximation for each log-likelihood components with details on the surrogate Poisson model used to represent it. For example, for the first part (integrated background rate) is represented by a Poisson model with log-intensity $\log \Lambda_0(\mathcal{X})$, this will be automatically linearised by Inlabru. Given that, the integrated background rate is just a quantity, and not a summation, we need only one observation to represent it with counts equal 0 and exposure equal 1. Table 2 shows that to represent

Table 2: Hawkes process log-likelihood components approximation

Name	Objective	Approximation	Surrogate $\log \lambda_P$	Number of data points	Counts and Exposures
Part I	$\Lambda_0(\mathcal{X})$	$\exp \overline{\log \Lambda}_0(\mathcal{X})$	$\log \Lambda_0(\mathcal{X})$	1	$c_i = 0, e_i = 1$
Part II	$\sum_{h=1}^{n} \sum_{i=1}^{B} \Lambda_h(b_i)$	$\sum_{h=1}^{n} \sum_{i=1}^{B} \exp \overline{\log \Lambda}_h(b_i)$	$\log \Lambda_h(b_i)$	$n \times B$	$c_i = 0, e_i = 1$
Part III	$\sum_{h=1}^{n} \log \lambda(\mathbf{x}_h)$	$\sum_{h=1}^{n} \exp \overline{\log \lambda}(\mathbf{x}_h)$	$\log \lambda(\mathbf{x})$	n	$c_i = 1, e_i = 0$

an Hawkes process model having observed n events, we need $n(B+1)+1$ events with B number of bins in the approximation of the second component of the log-likelihood.

Furthermore, Table 2 lists the components that has to be provided by the user, namely the surrogate Poisson models log-intensities. More specifically, the user only needs to create the datasets with counts c_i, exposures e_i, and the information on the events $\mathbf{x}_i$ representing the different log-likelihood components; and, to provide the functions $\log \Lambda_0(\mathcal{X}), \log \Lambda_h(b_i)$, and, $\log \lambda(\mathbf{x})$. The linearisation is automatically performed by Inlabru as well as the retrieving of the parameters' posterior distribution.

4 Real Data Example

We provide a practical example on a temporal marked Hawkes process to illustrate the capabilities of our technique comparing it with a MCMC implementation. Specifically, we implement the temporal version of the Epidemic-Type-Aftershock-Sequence model (ETAS, Ogata 1988), the most popular model to describe the evolution of seismicity, we apply the model to the 2016 Amatrice seismic sequence (Marzocchi, Taroni, and Falcone 2017). The temporal evolution of the number of events is illustrated in 1. We compare the results of our model with results obtained using the R-package bayesianETAS (Ross 2021), which is the only R-package (to the authors' knowledge) which offers an automatic MCMC implementation of the temporal ETAS model.

The ETAS model is the most used Hawkes process to model the evolution of seismicity over time and space (Ogata 1988; Ogata and Zhuang 2006; Ogata 2011). We are going to implement the first version of the model which is a temporal marked Hawkes process model with the event magnitude as marking variable. The conditional intensity of the ETAS model is given by:

$$\lambda_{\mathrm{E}}(t, m|\mathcal{H}_t) = \left(\mu + K \sum_{h:t_h<t} \exp\{\alpha(m_h - M_0)\}(t - t_h + c)^{-p} \right) \pi(m) \tag{23}$$

Where $\pi(m)$ is the magnitude distribution which is estimated independently from the Hawkes process parameters and assumed to follow a form of Gutemberg-Richter (GR) law (Gutenberg and Richter 1956). The temporal evolution of the number of points is regulated by 5 parameters $\mu, K, \alpha, c, \geq 0$ and $p \geq 1$. The parameters $\mu, K and \alpha$ are productivity parameters regulating: the number of background events (μ), the number of triggered events or aftershocks (K), and how the aftershock productivity scales with magnitude (α, the higher the magnitude the more events are generated). The parameters c and p are the parameters of the Omori's law (Omori 1894) and regulates the temporal decay of the aftershock activity. The quantity M_0 is a cut-off magnitude such that $m_h \geq M_0, \forall h$.

The bayesianETAS package implements the ETAS model with a normalized time triggering function. The conditional intensity is given by:

$$\lambda_{\mathrm{bE}}(t, m|\mathcal{H}_t) = \left(\mu + K \sum_{h:t_h<t} \exp\{\alpha(m_h - M_0)\} \frac{c^{p-1}}{p-1} (t - t_h + c)^{-p} \right) \pi(m) \tag{24}$$

They consider the following priors for the parameters

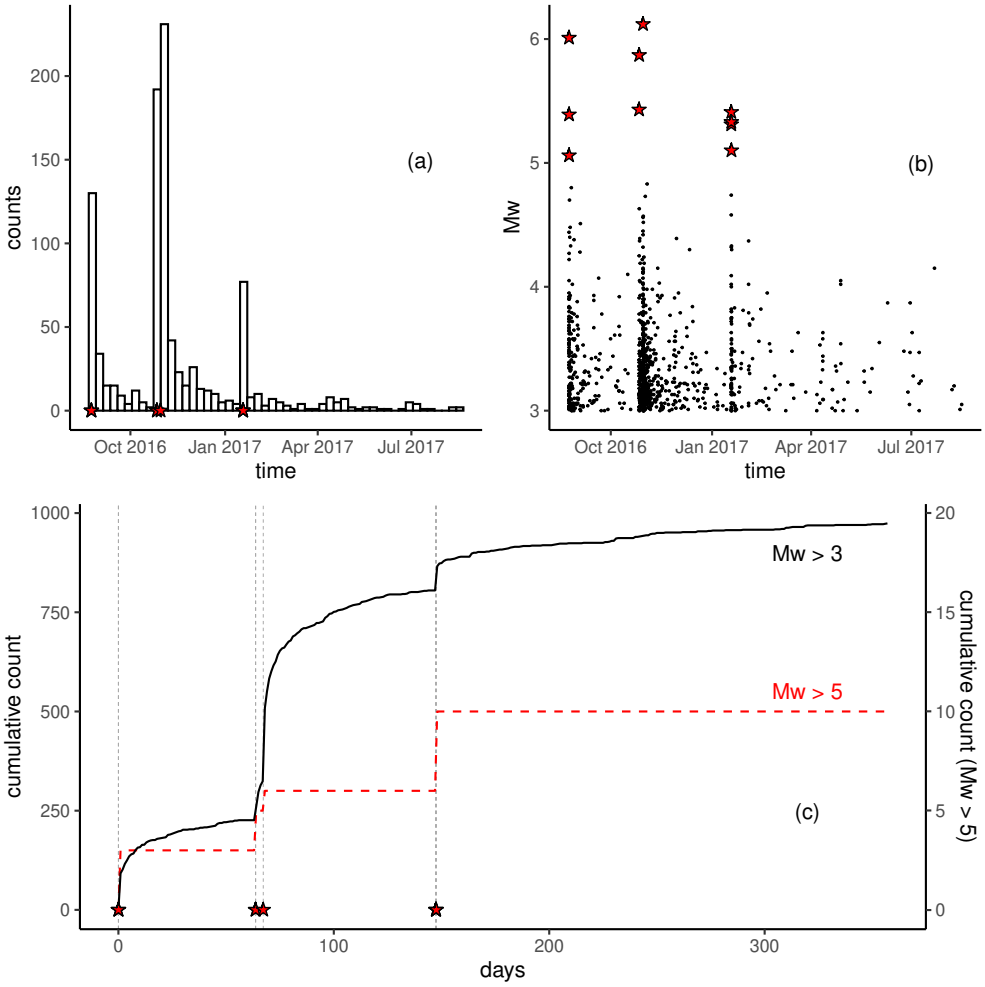

Figure 1: Amatrice sequence comprising 974 events from 24/08/2016 to 15/08/2017, with latitude in $(42.45, 43.08)$ and latitude in $(12.93, 13.54)$. The first event in the catalogue is the magnitude 6.01 which started the sequence. Red stars indicates events with magnitude greater than 5. Panel (a): Histogram reporting the number of events per week; Panel (b): Scatter plot of time versus magnitude; (c) Cumulative number of events as function of the number of days from the first event in the sequence, for event with magnitude greater than 3 (black) and for event with magnitude greater than 5 (red).

$$\begin{aligned} \mu &\sim \text{Gamma}(0.1, 0.1) \\ K, \alpha, c &\sim \text{Unif}(0, 10) \\ p &\sim \text{Unif}(1, 10) \end{aligned} \tag{25}$$

With our technique, it is best to work with a different parametrization than the one used in the bayesianETAS package. Specifically, we will be working with the following conditional intensity

$$\lambda_{\text{bru}}(t, m|\mathcal{H}_t) = \left(\mu_b + K_b \sum_{h:t_h<t} \exp\{\alpha_b(m_h - M_0)\} \left(\frac{t - t_h}{c_b} + 1 \right)^{-p_b} \right) \pi(m) \tag{26}$$

The parameters of the Inlabru implementation have the same constraints as in the bayesianETAS implementation. We are going to consider a different set of priors for the Inlabru parameters. This is done to increase the numerical stability of the algorithm and the fact that uniform priors are very informative in our view (Zhu and Lu 2004). We consider the same log-gaussian priors for all parameters, formally:

$$\mu_b, K_b, \alpha_b, c_b, p_b - 1 \sim \text{logN}(0, 1)$$

Where $X \sim \text{logN}(0, 1) \iff \log X \sim N(0, 1)$.

We want to remark that using different parametrisations and different priors makes a direct comparison of the posterior of the parameters hard, the parameters have the same role (conceptually) but not the same value (numerically). Indeed, the present comparison does not want to compare the two methodologies in terms of ability in retrieving the parameters. Our only goal is to show that our procedure is valuable and produce results similar to other Bayesian methodologies in terms of goodness of fit.

For this reasons, we compare the goodness-of-fit of the models using the Random Time Change Theorem (REF) which is here reported similarly to Laub, Lee, and Taimre 2021 (Theorem 9.1) :

Theorem 4.1 *Say $\mathcal{H} = \{t_1, ..., t_k\}$ is a realisation over time $[0, T]$ from a point process with conditional intensity $\lambda(t|\mathcal{H})$. If $\lambda(t|\mathcal{H})$ is positive over $[0, T]$ and $\Lambda(T) < \infty$ almost surely, then the transformed points $\{\Lambda(t_1), ..., \Lambda(t_k)\}$ form a Poisson process with unit rate.*

Where, in our case,

$$\Lambda(t_i|H) = \int_{M_0}^{\infty} \int_0^{t_i} \lambda(t, m|\mathcal{H}) dt dm \tag{27}$$

In other words, if we calculate the sequence of values $\Lambda(t_1), ..., \Lambda(t_n)$, for observed $t_1, ..., t_n$, using the respective expressions of $\Lambda(t_i)$ for the bayesianETAS implementation and ours, we have to obtain a sequence of points uniformly distributed over the interval $[0, n]$, where n is the number of observed points. For the MCMC method we consider estimates based on 55000 posterior samples with a burn-in of 5000 samples.

Figure 4a compares the sequences $\Lambda_{\text{bE}}(t_1), ...\Lambda_{\text{bE}}(t_n)$, and $\Lambda_{\text{bru}}(t_1), ...\Lambda_{\text{bru}}(t_n)$ with observed cumulative counts $N(t_1), ..., N(t_n)$. Figure 4b shows the cumulative counts as function of $\Lambda(t_h)$ and should look like a straight line if the values are uniformly distributed as expected by the theorem. For both plot we report 95% posterior intervals for the quantity of interest based on 10000 samples from the posterior of the parameters.

5 Discussion and conclusions

We illustrated in this paper a technique to implement Bayesian Hawkes process models based on the INLA algorithm and carried out with the R-package Inlabru. The proposed technique is completely new and differs substantially from other Hawkes process implementation. Specifically, we do not rely on any declustering algorithm assigning the observations to the background rate or the triggered part of the intensity. Also, our algorithm is deterministic ensuring the same numerical results if the analysis is repeated on different machines with the same specifics. Moreover, the user does not have to programme explicitly the algorithm itself, it has only to provide the functions to approximate the three parts of the log-likelihood and set up the different log-likelihood components.

We have seen through the example, that the technique provides similar results (in terms of goodness-of-fit) to the MCMC method implemented in the bayesianETAS package. Also, we have seen that we are not completely free to choose the functional form of the triggering functions, which may cause numerical stability problems, and convergence issues due to multimodality of the posterior distribution. This problems, however, are common in the Hawkes process world and our technique supports all the most widely used triggering functions. They have just to be considered in a convenient form which may differ from the one used in other implementations. Also, we have seen that the choice of the binning strategy to approximate Part II of the log-likelihood is essential to reach convergence. We recommend to try different techniques and to tailor the strategy to the triggering function at hand.

A difference from other algorithms for Hawkes proces models is that we offer a general extendible framework to perform Bayesian analysis of Hawkes processes parameters. Indeed, INLA was designed for models comprising covariates and random effects. This allows us to bring the advantages of the Latent Gaussian model world into the Hawkes process world. For example, the use of covariates may be used to consider link the covariates to the parameters of the model, e.g. earthquakes are usually recorded around faults, therefore, a spatially varying background rate or a productivity parameter (K, α in the ETAS model) which depends on the distance from faults may be a valuable model for this data. Random effect are there to capture the correlation between observations which is not explained by the model. Random effect with a spatial (or temporal) covariance structure are usually employed to model variations in time and space. Due to the complex correlation structure between parameters' values they are usually quite difficult to

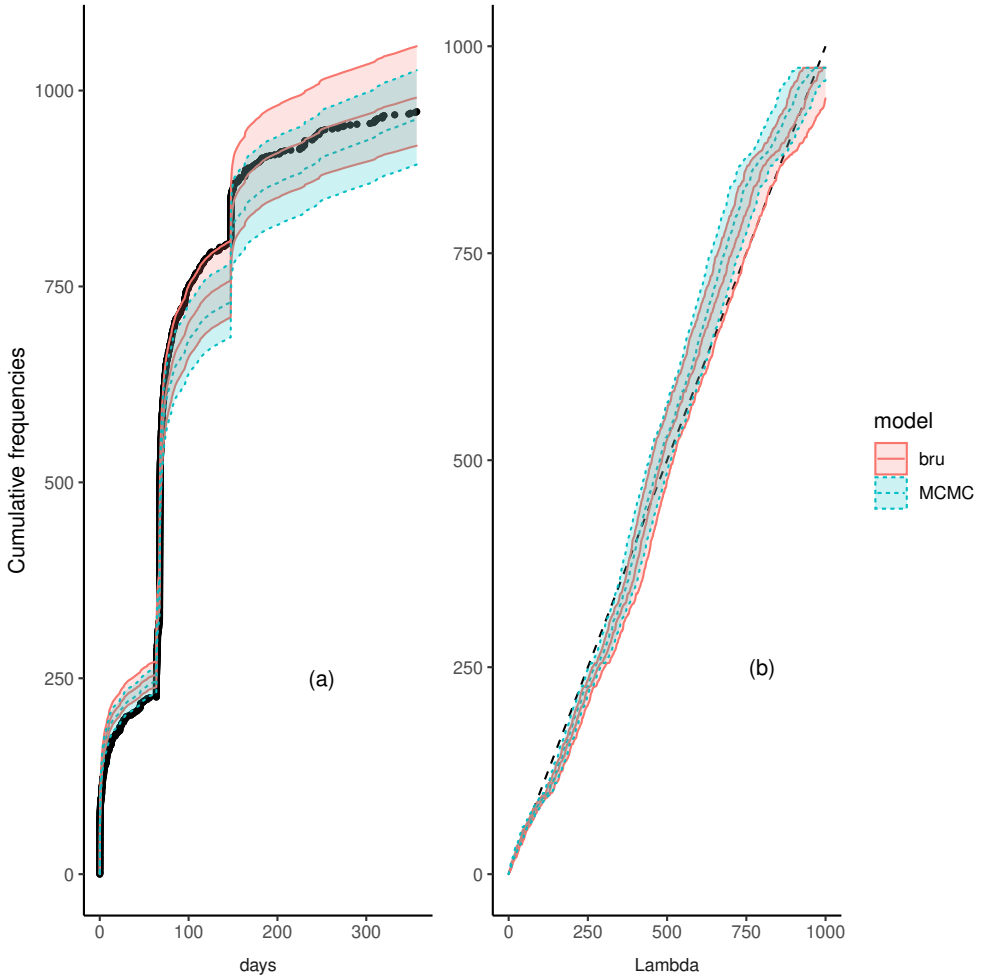

Figure 2: Application of the Random Time Change Theorem. Top row: Histogram of values of $\Lambda(t_h)$ under the two implementation using the posterior median as parameters point estimate; Bottom row: Cumulative counts as function of $\Lambda(t_h)$ under the two implementation. The red dashed line represents Inlabru, and the blue continuous represents bayesianETAS.

inconsistent with the graph!

implement using MCMC methods, which would require extenuating computational times to reach convergence. On the other hand, INLA was designed specifically to support this type of random effects. The structured random effects may be used as alternative to estimate spatially (or temporally) varying parameters in the case of Hawkes process models. For example, we could add spatial variability to a parameter considering funcion of an SPDE effect (Lindgren, Rue, and Lindström 2011), under this model the correlation between the values of the parameter at different locations is a decreasing function of the distance between locations. This means that the parameters will assume values close to each other at close locations and independent values at far locations.

Combinations of covariates and random effects may be used, and we think that providing researchers with the freedom of focusing on the hypothesis incorporated in the model, and not on the optimization routine, is essential, especially in applied contexts. Furthermore, all the models undergo the same optimization routine making them homogeneous under this aspects. In fact, when comparing two models optimized with different routines, it is hard to distinguish if the differences comes from the different models or the different algorithms. Using our technique, researchers may compare models incorporating different hypothesis being sure of no differences at least on the optimization part, and thus, any difference in performance would come only from the model formulation itself.

We believe that our technique is a valid alternative to existing Bayesian methods for Hawkes process models, and we are working towards building a fast, reproducible, easy-to-code, framework for applied research using this class of models.

References

Azizpour, Shahriar, Kay Giesecke, and Gustavo Schwenkler (2018). "Exploring the sources of default clustering". In: *Journal of Financial Economics* 129.1, pp. 154–183.

Bachl, Fabian E et al. (2019). "inlabru: an R package for Bayesian spatial modelling from ecological survey data". In: *Methods in Ecology and Evolution* 10.6, pp. 760–766.

Bakka, Haakon et al. (2018). "Spatial modeling with R-INLA: A review". In: *Wiley Interdisciplinary Reviews: Computational Statistics* 10.6, e1443.

Bayliss, Kirsty et al. (2020). "Data-Driven Optimization of Seismicity Models Using Diverse Data Sets: Generation, Evaluation, and Ranking Using Inlabru". In: *Journal of Geophysical Research: Solid Earth* 125.11, e2020JB020226.

Blangiardo, Marta et al. (2013). "Spatial and spatio-temporal models with R-INLA". In: *Spatial and spatio-temporal epidemiology* 4, pp. 33–49.

Chiang, Wen-Hao, Xueying Liu, and George Mohler (2022). "Hawkes process modeling of COVID-19 with mobility leading indicators and spatial covariates". In: *International journal of forecasting* 38.2, pp. 505–520.

Dempster, Arthur P, Nan M Laird, and Donald B Rubin (1977). "Maximum likelihood from incomplete data via the EM algorithm". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 39.1, pp. 1–22.

Filimonov, Vladimir and Didier Sornette (2012). "Quantifying reflexivity in financial markets: Toward a prediction of flash crashes". In: *Physical Review E* 85.5, p. 056108.

Garetto, Michele, Emilio Leonardi, and Giovanni Luca Torrisi (2021). "A time-modulated Hawkes process to model the spread of COVID-19 and the impact of countermeasures". In: *Annual reviews in control* 51, pp. 551–563.

Goicoa, T et al. (2016). "Age–space–time CAR models in Bayesian disease mapping". In: *Statistics in medicine* 35.14, pp. 2391–2405.

Gómez-Rubio, Virgilio (2020). *Bayesian inference with INLA*. CRC Press.

Gutenberg, Beno and Charles Francis Richter (1956). "Magnitude and energy of earthquakes". In: *Annals of Geophysics* 9.1, pp. 1–15.

Halonen, Jaana I et al. (2015). "Road traffic noise is associated with increased cardiovascular morbidity and mortality and all-cause mortality in London". In: *European heart journal* 36.39, pp. 2653–2661.

Hawkes, Alan G (1971a). "Point spectra of some mutually exciting point processes". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 33.3, pp. 438–443.

– (1971b). "Spectra of some self-exciting and mutually exciting point processes". In: *Biometrika* 58.1, pp. 83–90.

– (2018). "Hawkes processes and their applications to finance: a review". In: *Quantitative Finance* 18.2, pp. 193–198.

Kobayashi, Ryota and Renaud Lambiotte (2016). "Tideh: Time-dependent hawkes process for predicting retweet dynamics". In: *Tenth International AAAI Conference on Web and Social Media*.

Laub, Patrick J, Young Lee, and Thomas Taimre (2021). *The Elements of Hawkes Processes*.

Lindgren, Finn, Håvard Rue, and Johan Lindström (2011). "An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73.4, pp. 423–498.

Marzocchi, Warner, Matteo Taroni, and Giuseppe Falcone (2017). "Earthquake forecasting during the complex Amatrice-Norcia seismic sequence". In: *Science Advances* 3.9, e1701239.

Mohler, George (2013). "Modeling and estimation of multi-source clustering in crime and security data". In: *The Annals of Applied Statistics*, pp. 1525–1539.

Mohler, George, Jeremy Carter, and Rajeev Raje (2018). "Improving social harm indices with a modulated Hawkes process". In: *International Journal of Forecasting* 34.3, pp. 431–439.

Mohler, George O et al. (2011). "Self-exciting point process modeling of crime". In: *Journal of the American Statistical Association* 106.493, pp. 100–108.

Ogata, Yosihiko (1988). "Statistical models for earthquake occurrences and residual analysis for point processes". In: *Journal of the American Statistical association* 83.401, pp. 9–27.

– (2011). "Significant improvements of the space-time ETAS model for forecasting of accurate baseline seismicity". In: *Earth, planets and space* 63.3, pp. 217–229.

Ogata, Yosihiko and Jiancang Zhuang (2006). "Space–time ETAS models and an improved extension". In: *Tectonophysics* 413.1-2, pp. 13–23.

Omori, Fusakichi (1894). "On the after-shocks of earthquakes". In: *J. Coll. Sci., Imp. Univ., Japan* 7, pp. 111–200.

Opitz, Nina et al. (2016). "Extensive tissue-specific transcriptomic plasticity in maize primary roots upon water deficit". In: *Journal of Experimental Botany* 67.4, pp. 1095–1107.

Peng, Roger D, Frederic Paik Schoenberg, and James A Woods (2005). "A space–time conditional intensity model for evaluating a wildfire hazard index". In: *Journal of the American Statistical Association* 100.469, pp. 26–35.

Rasmussen, Jakob Gulddahl (2013). "Bayesian inference for Hawkes processes". In: *Methodology and Computing in Applied Probability* 15.3, pp. 623–642.

Riebler, Andrea et al. (2016). "An intuitive Bayesian spatial model for disease mapping that accounts for scaling". In: *Statistical methods in medical research* 25.4, pp. 1145–1165.
Robert, Christian P, George Casella, and George Casella (1999). *Monte Carlo statistical methods*. Vol. 2. Springer.
Roos, Natalia C et al. (2015). "Modeling sensitive parrotfish (Labridae: Scarini) habitats along the Brazilian coast". In: *Marine Environmental Research* 110, pp. 92–100.
Ross, Gordon J (2021). "Bayesian estimation of the ETAS model for earthquake occurrences". In: *Bulletin of the Seismological Society of America* 111.3, pp. 1473–1480.
Rue, Håvard et al. (2017). "Bayesian computing with INLA: a review". In: *Annual Review of Statistics and Its Application* 4, pp. 395–421.
Santermans, Eva et al. (2016). "Spatiotemporal evolution of Ebola virus disease at sub-national level during the 2014 West Africa epidemic: model scrutiny and data meagreness". In: *PloS one* 11.1, e0147172.
Schrödle, Birgit and Leonhard Held (2011). "A primer on disease mapping and ecological regression using `INLA`". In: *Computational statistics* 26.2, pp. 241–258.
Touati, Sarah, Mark Naylor, and Ian G Main (2009). "Origin and nonuniversality of the earthquake interevent time distribution". In: *Physical Review Letters* 102.16, p. 168501.
Veen, Alejandro and Frederic P Schoenberg (2008). "Estimation of space–time branching process models in seismology using an em–type algorithm". In: *Journal of the American Statistical Association* 103.482, pp. 614–624.
Zhou, Ke, Hongyuan Zha, and Le Song (2013). "Learning social infectivity in sparse low-rank networks using multi-dimensional hawkes processes". In: *Artificial Intelligence and Statistics*. PMLR, pp. 641–649.
Zhu, Mu and Arthur Y Lu (2004). "The counter-intuitive non-informative prior for the Bernoulli family". In: *Journal of Statistics Education* 12.2.
Zipkin, Joseph R et al. (2016). "Point-process models of social network interactions: Parameter estimation and missing data recovery". In: *European journal of applied mathematics* 27.3, pp. 502–529.

6 Appendix: Parameters posterior distribution

Below are shown the marginal posterior distribution of the ETAS parameters calibrated on the Amatrice sequence comprising 927 events from 24/08/2016 to 15/08/2017 with latitude in and longitude in.

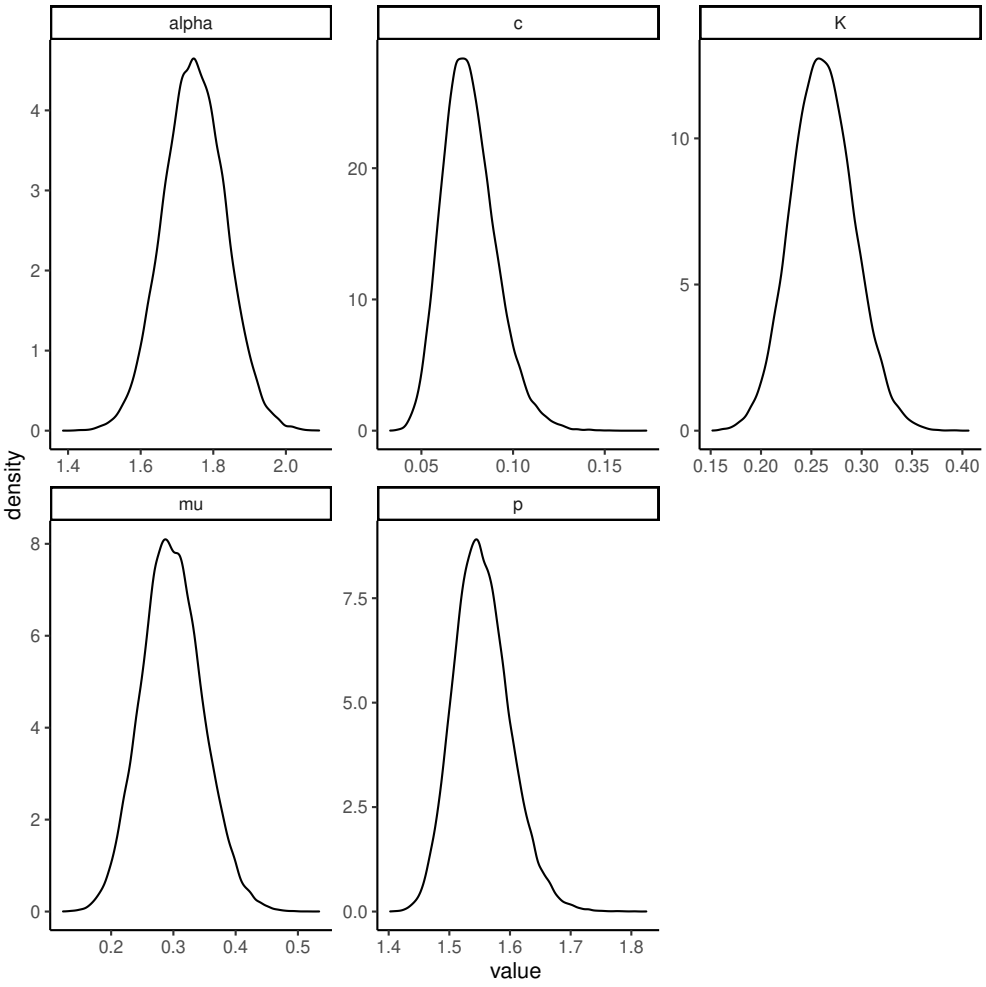

Figure 3: Posterior distribution of the parameters of the bayesianETAS implementation

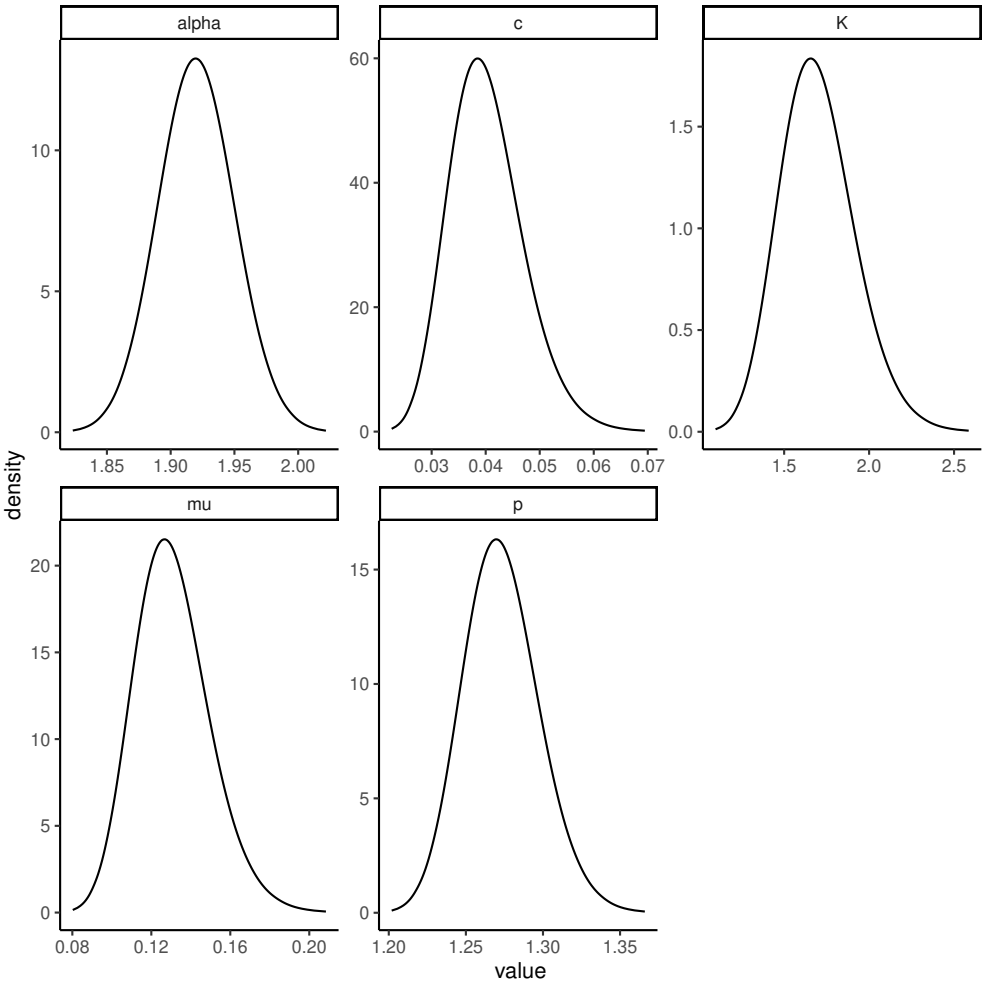

Figure 4: Posterior distribution of the parameters of the Inlabru implementation