Ilaria Spassiani

A new version of the ETAS model for seismic events with correlated magnitudes

Dottorato in Matematica, Roma «La Sapienza» (2015).
<http://www.bdim.eu/item?id=tesi_2015_SpassianiIlaria_1>

SAPIENZA
UNIVERSITÀ DI ROMA

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI
Dottorato in Ricerca in Matematica per le Applicazioni-Ciclo XXVIII

A new version of the ETAS model for seismic events with correlated magnitudes

Supervisor:
Prof. Giovanni Sebastiani

Ph.D. Thesis of:
Ilaria Spassiani

Co-Supervisor:
Prof. Giovanna Nappo

To 'Ngenzone and Alessandro

Contents

Table of notations . . . . **iv**

Introduction and motivation **1**

1 Mathematical background **8**
- 1.1 Stochastic point processes 9
- 1.2 Finite point processes 12
- 1.3 Moment measures and probability generating functionals . . . 15
- 1.4 Marked and cluster processes 18
  - 1.4.1 Marked processes 18
  - 1.4.2 Cluster processes 21
- 1.5 Stationary point processes on $\mathbb{R}$ 23
  - 1.5.1 Palm theory 26
  - 1.5.2 Conditional intensities 27
- 1.6 Point processes for earthquake modeling 29
  - 1.6.1 Poisson processes 29
  - 1.6.2 Self-exciting Hawkes processes 31
  - 1.6.3 Stress-release models 33

2 The Epidemic Type Aftershock Sequence model **35**
- 2.1 Derivation and definition of the model 35
- 2.2 Properties 40
  - 2.2.1 Criticality 41
  - 2.2.2 First and second order moments 44
- 2.3 Catalog simulation, parameters estimation and residual analysis 47
  - 2.3.1 Parameter estimation 48
  - 2.3.2 Simulation 50
  - 2.3.3 Residual analysis 51

3 Experimental analysis **53**
- 3.1 The catalogs analyzed 54
- 3.2 The two types of analysis 61

3.2.1 Algorithm 1: first analysis 61
3.2.2 Algorithm 1.1: choice of the time amplitude δ^* 63
3.2.3 Algorithm 2: second analysis 67
3.3 Results 68
3.4 Conclusions 82

4 The Epidemic Type Aftershock Sequence model with correlated magnitudes 83
4.1 The conditional intensity of the new ETAS model with correlated magnitudes 84
4.2 The Laplace functional 92
4.2.1 The probability generating function obtained from the Laplace functional 98
4.3 Magnitude transition probability density function for triggered events' magnitude: an explicit form 101

5 The analysis of the simulated catalogs 113
5.1 The catalogs simulation 113
5.1.1 A new algorithm for simulation 116
5.2 Results 119
5.3 Conclusions 126

6 Mathematical characterization of the new ETAS model with correlated magnitudes 127
6.1 Probability generating function for earthquake sequences . . . 130
6.2 The case $m_c = m_0$ 137
6.3 The general case $m_c \geq m_0$ 144
6.3.1 Another approximation for small τ 165
6.4 Conclusions 168

Overall conclusions and future work 168

Appendix 173

A A brief review on the probability generating function and the statistical tests for seismic residual analysis 174
A.1 Probability generating function and some of its fundamental properties 174
A.2 The Kolmogorov-Smirnov test 175
A.3 The Runs test 176

B Proof of Lemma 1 177

B.1 Eigenvalues and eigenfunctions of the operators $\mathcal{K}_1$ and $\mathcal{K}_2$ . . 177
B.2 Finiteness of the mean cluster size 183

C Time derivatives of the probability P$\{\tau\}$ of having zero observable events in $[0, \tau]$ **187**

D Algorithms **190**

Bibliography **191**

Table of notations

$\otimes$	convolution operator. It is preferred here to the more usual symbol '$*$', as for in the majority of the literature about mathematical seismology.
a, α	exponent of the productivity law (equation (2.1)); it holds $a = \alpha \ln 10$.
$a(t)$	survivor function of the random variable with density given by the Omori-Utsu law (equation (6.3)); it is defined in (6.15).
$A(t)$	auxiliary function defined in (6.14) as the integral of $a(x)$ between zero and t; if T_{aft} is the random variable distributed according to the Omori-Utsu law (equation (6.3)), it holds $A(t) = \mathbb{E}[\min(T_{aft}, t)]$.
AIC	Akaike's Information Criterion, defined in (2.25), for comparing the goodness-of-fit of seismic models for a fixed dataset.
β, b	exponent of the Gutenberg-Richter law (equation (2.2)); b is the so-called *b-value* and holds that $\beta = b \ln 10$.
$b(t)$	cumulative function of the random variable with density given by the Omori-Utsu law (equation (6.3)); it is defined in (6.7).
c	parameter of the Omori-Utsu law (equation (2.5)).
C_1, C_2	constants entering in the definition of the function $q(m')$; it holds $C_2 = 2C_1$.
$\overline{C}(\cdot)$	parameter defined in (5.3).
$D(z; t, \tau, m_0, \overline{m})$	auxiliary function defined in (6.31) entering in the definition of $y(z; t, \tau, m_0, m')$.
$D_+(z; t, \widetilde{m}, \overline{m})$	auxiliary function defined in (6.32) entering in the definition of $y(z; t, \tau, m_0, m')$ and $s(z; t, m_0, m')$.
δ	parameter defined in (6.63).
Δ	parameter defined in (6.67).
η	branching parameter defined in (2.18).

ETAS	Epidemic Type Aftershock Sequence model.
$f(m', m'')$	auxiliary function defined in (4.36) useful to calculate the transition probability density $p(m''\|m')$.
$f(z, m', m)$	function defined in (6.9) useful to calculate both the probability generating function of the total number of events in $[0, \tau]$ and the one relative to the number of event triggered by a generic shock in the same time interval. The multiplication by this function allows us to add the triggering event if its magnitude is larger than the observability threshold m.
$\Phi(t)$	Omori-Utsu law (equation (2.5)); it can be interpreted as the probability generating function of random times at which first generation shocks independently occur, triggered by an event occurred in $t = 0$.
$F_{inter}(\tau)$	interevent time density.
$\tilde{F}_{(m''\|m')}$	cumulative distribution function of the transition probability $p(m''\|m')$, given in (5.5).
$G[\cdot]$	probability generating functional, defined for a finite point process in (1.6).
$G^{tr}(z; \tau\|s, m')$	probability generating function of the number of observable events triggered in $[0, \tau]$ by a spontaneous shock of magnitude m' occurred in s, defined in (4.22).
$G(z; \tau\|t, m')$	probability generating function of the number of observable events triggered in $[0, \tau]$ by a spontaneous shock of magnitude m' occurred in $-t$, with $t \geq 0$. It holds $G(\cdot; v\|s, m') = G^{tr}(\cdot; v\| - s, m')$ for $s \geq 0$.
H	parameter defined in (6.55).
$I_A(m'')$	magnitude integral from the reference cutoff m_0 to infinity, defined in (6.46), of the product between the Gutenberg-Richter law (2.2), the productivity law (2.1) and the transition probability density function (4.37) of triggered events' magnitude. It is also $\mathcal{K}_2 p(m)$, for $\mathcal{K}_2$ the operator defined in (4.6). We impose that $I_A(m'') = \eta p(m'')$.
$I_B(m'')$	magnitude integral from the completeness value m to infinity, defined in (6.52), of the product between the Gutenberg-Richter law (2.2), the productivity law (2.1) and the transition probability density function (4.37) of triggered events' magnitude. It is also $\mathcal{K}_2 p(m) \mathbb{1}_{[m_c,\infty)}(m)$, for $\mathcal{K}_2$ the operator defined in (4.6). We impose that $I_B(m'') \approx nLp(m'')$.

κ	multiplicative parameter of the productivity law (equation (2.1)).
$K(\cdot)$	kernel, defined in Definition 16.
$\mathcal{K}_1, \mathcal{K}_2$	operators acting on functions (defined in (4.5)) and measures (defined in (4.6)), respectively.
l.c.s.c.	locally compact second countable. A set E is *l.c.s.c.* if it is Hausdorff, each point has a compact neighborhood (locally compactness) and there exists a countable basis such that every open set in E is the union of the open sets of this basis (second countable).
L	parameter entering in the approximate condition that the integral $I_B(m'')$ must satisfy (see page 156).
$L_t(\cdot)$	likelihood function defined for finite point processes in Definition 7.
$L(z;\tau)$	auxiliary function useful to compute the probability generating function $\Omega(z;t)$ and then the probability to have zero events with magnitude larger than m_c in $[0,\tau]$. It is defined as $-\ln(\Omega(z;\tau))/\varpi$ and obtained in (6.5).
$\mathfrak{L}(f)(s)$	Laplace transform of the function f.
$\mathcal{L}(\cdot)$	Laplace functional relative to the total number of events $N(\cdot)$, defined in (4.12).
$\mathcal{L}^{sp}(\cdot), \mathcal{L}^{tr}(\cdot)$	Laplace functionals relative to the total number of spontaneous events $N^{sp}(\cdot)$ and the total number of triggered events $N^{tr}(\cdot)$, respectively (see Section4.2).
$\lambda(t,x,y,m\|\mathcal{H})$	conditional intensity function based on the past history $\mathcal{H}_t = \{(t_i,x_i,y_i,m_i); t_i < t\}$, completely characterizing a space-time-magnitude point process. It is generally defined in (1.15).
Λ_t	random time-change for the residual analysis of the ETAS model, defined in (2.26).
$\bar{\lambda} = \bar{\lambda}(m_c)$	average rate of the whole process of observable events, defined in (6.2).
m_c	threshold value of completeness magnitude.
m_0	threshold value of reference magnitude.
$M_k(\cdot)$	k^{th} moment measure, defined in (1.4) and (1.5) for an extended version.
$M_{[k]}(\cdot)$	k^{th} factorial moment measure, defined in (1.4).
$\hat{M}_\gamma(\cdot)$	kernel density estimator of the empirical magnitude distribution, defined in (3.1).

$(M_E^{lf}, \mathcal{M}_E^{lf})$	canonical space of locally finite measures on $(E, \mathcal{B}(E))$, where E is a *l.c.s.c.* space, $\mathcal{B}(E))$ is its Borel sigma-field, M_E^{lf} is the set of all the locally finite measures on $(E, \mathcal{B}(E))$ and $\mathcal{M}_E^{lf}$ is the relative sigma-field generated by the collection of sets defined in (1.1).
$N(t)$, $N_1(\tau)$	counting measures for a point process: $N(t) = N((0, t])$ is used in Chapter 1 to indicate the number of events in $(0, t]$; the same notation $N(\tau)$ is instead used in Chapter 6 for the total number of events with magnitude bigger than the completeness threshold m_c, occurred in $[0, \tau]$; finally, again in Chapter 6, $N_1(\tau)$ is used for the number of aftershocks in $[0, \tau]$ with magnitude $m \geq m_c$.
$N_i^{tr}(\mathbb{R}^2)$, $N_i^{tr,n}(\mathbb{R}^2)$	number of all the triggered events generated by the spontaneous shock $(S_i = s, M_i = m)$ and number of all the events in the n^{th} generation of the latter spontaneous shock, respectively.
$\overline{N}(t, \tau)$, $\overline{N}_{m_c=m_0}(t, \tau)$, $\overline{N}_-(t)$	three auxiliary functions useful to compute $L(0; \tau)$; it holds $\overline{N}(t, \tau) = \overline{N}_{m_c=m_0}(t, \tau)$.
ϖ	average rate corresponding to the events of the spontaneous component of the process.
$\Omega(z; \tau)$	probability generating function of the total number $N(\tau)$ of observable events in $[0, \tau]$.
p, θ	exponents of the two notations used for Omori-Utsu law, respectively equations (2.5) and (6.3). It holds $\theta = p - 1$.
$p(m)$	Gutenberg-Richter law (equation (2.2)); it is the probability density function of events' magnitude when we do not consider past seismicity.
$p(m''\|m')$	transition probability density function assumed for triggered events' magnitude; it is defined in (4.37).
$p_{tr}(m'')$	conditional probability for an event to have magnitude m'' given the fact that it is triggered; it is defined in (4.26).
$\mathbf{P}\{\tau\}$	probability to have zero events with magnitude larger than m_c in $[0, \tau]$.
PGF	Probability Generating Function.
$\Psi(h, \widetilde{m})$	auxiliary function defined in (6.27) useful to compute $L(z; \tau)$.
$q(m')$, $s(m')$	auxiliary functions entering in the definition of $f(m', m'')$ (equation (4.36)); the function $q(m')$ is defined in (4.35) and holds that $s(m') = 2q(m')$.

$\widetilde{Q}$	parameter defined in (6.68).
Q	parameter defined in (6.51) and computed as the integral of the Gutenberg-Richter law between the completeness value m_c and infinity. It is the probability $\mathbb{P}\{M > m_c\}$ if M is the random variable whose density is the Gutenberg-Richter law (equation (2.2)).
$\varrho(m)$	productivity law (equation (2.1)); it is the contribution of first generation shocks triggered by an event with magnitude m'.
$\widehat{R}(\delta)$	estimate of the autocorrelation function at different integer values of time lag δ, defined in (3.2).
$s(z; t, m_0, m')$, $y(z; t, \tau, m_0, m')$	auxiliary functions used for $L(z; \tau)$ in the case $m_c \geq m_0$. They are respectively defined in (6.35) and (6.30).
$\Theta(s, m)$	auxiliary function, defined in (4.15), for the Laplace functional $\mathcal{L}(\cdot)$.

Introduction and motivation

The study of the earthquake phenomenon represents a very important scientific problem that commits a scientific community more and more extended. Geology and physics are the main disciplines focused on the study of this natural phenomenon. In fact, an earthquake is generally due to a stored elastic strain energy that causes a fracture propagation along the so-called *fault plane*. This sudden rupture in the Earth's crust causes permanent deformation and radiation of seismic waves. Then, ground motion and wave propagation represent two important aspects to study in order to understand the phenomenon. Nevertheless, during the last fifty years, a mathematical approach has been considered to support the above disciplines in the study of earthquakes. Just like in many other fields, as for example engineering and environmental processes, mathematical modeling allows to analyze the seismic process at first in an ideal and easier case and then to interpret the results obtained for the study of a more and more realistic and complex case, to the aim of understanding completely the phenomenon. Obviously, the latter is a very hard aim due to the complexity of the seismic process. More precisely, earthquakes behave like a *Self-Organized Criticality* process (SOC), comparable to a Sand-Pile model [Bak and Tang, 1989; Sornette and Sornette, 1989]. In [Bak and Chen, 1991] we can read ≪the theory of self-organized criticality states that many composite systems naturally evolve to a critical state in which a minor event starts a chain reaction that can affect any number of elements in the system≫. Not only one fault, but several of them interact in the physics underneath the process, adding a component of complexity to the already very complicated phenomenon of earthquakes development. Furthermore, the available data are not always enough or suitable to obtain significant results: even if the first historical earthquakes recorded are dated back to 2100 B.C., there is a lack of description or reliability till the middle of the 18^{th} century. Starting from the first forerunner of modern seismographs, designed by the English geologist *John Milne* in 1870, seismic stations have been technologically gradually improved, but they can still be refined in order to get more precise and complete data.

Despite this very complex scenario, it is of fundamental importance to analyze the seismic process, also in order to obtain results that could avoid human catastrophes and allow to be prepared to this natural phenomenon as well as possible. A mathematical approach, more precisely probabilistic and statistical, is then very useful to this aim. Indeed, it is now of common use the expression "statistical seismology": it indicates the scientific field which regards the earthquake predictions based on the statistical modeling developed to analyze the real data, the results of which are used to evaluate the probability of an earthquake to occur in a certain space-time-magnitude window. In fact, let's recall that one can associate to each shock the following three quantities, obtained by combining the signals registered by several seismographs located in more sites of the Earth's surface:

- **spatial location**, indicated in longitude and latitude, of the *epicenter*. The latter corresponds to the normal projection on the Earth's surface of the inner point, called *hypocenter*, at which the shock has occurred; the distance between the epicenter and the hypocenter is the focal depth of the shock;

- **time instant**, indicated in the data/hour format or in an absolute time starting from a certain year of choice;

- **magnitude**, that measures the "size" of an earthquake. C. F. Richter was the first to introduce the concept of seismic magnitude in 1935, on the basis of some instrumental recordings. The *Richter's local magnitude scale* (M_L) is a base-10 logarithm scale and measures the energy released by the earthquake at the hypocenter in the form of seismic waves. Initially, M_L measured only the energy of earthquakes occurred in California within 600Km from the *Woods-Anderson torsion seismograph*, but then it has been extended to earthquakes of all the distances and all the focal depths.

  The local magnitude is one of the most used scale at the moment, together with the recently introduced *Moment magnitude scale* M_W. It has been defined by Hiroo Kanamori as

  $$M_W = \frac{2}{3} \log M_0 - 10.7,$$

  where M_0 is the seismic moment, defined by the equation $M_0 = \varsigma AD$, where ς is the shear modulus of the faulted rocks, A is the area of the rupture along the fault and D is the average displacement on A [Fowler, 1990]. This magnitude scale is mostly used for strong shocks.

Several other magnitude scales have been proposed, but we mention only the *duration magnitude* M_d, generally used for small earthquakes and defined by

$$M_d = a \log T_s + bD_s - c,$$

where T is the duration of the signal, D is the distance from the source which receives the signal and (a, b, c) are constants.

For completeness of information, let's recall that another measure of the size of an earthquake is given by the *Mercalli's intensity scale*. It measures a shock in terms of observed damage. It is the scale which generally indicates the size of historical earthquakes.

Locations, occurrence times, magnitudes and other earthquakes' characteristics, like for example the depths of the hypocenters, are recorded in the seismic catalogs. The use of the latter could be very important to analyze completely the phenomenon. In fact, it is fundamental to combine a mathematical study with an experimental analysis, in order to have both the theoretical support and the real experimental validation of the results.

In an earthquake sequence, the events are typically divided into spontaneous (or background) and triggered (aftershocks). The background seismicity is the component not triggered by precursory events and is usually connected to the regional tectonic strain rate; on the other hand, the triggered seismicity is the one associated with stress perturbations due to previous shocks [Lombardi et al., 2010]. Both a triggered and a background event may generate aftershocks. In this study, we will use the terms "triggering event" to indicate any mother event that produces its own progeny [Saichev and Sornette, 2005].

In any given catalog one can consider two different threshold magnitude values. The first one is the *completeness magnitude* m_c, that is the smallest value such that all the events with magnitudes exceeding it are surely recorded in the catalog. The second threshold is instead connected to the concept of triggering: the *reference magnitude* m_0 is the minimum value for an event to be able to trigger its own offsprings [Helmstetter and Sornette, 2002b]. Reference magnitude is usually theoretically set less than or equal to the completeness one [Sornette and Werner, 2005]. Indeed, even if the opposite inequality were verified, we could however reduce to the previous case by excluding the data relative to observable events with magnitude less than m_0. In practice, the value of m_c is estimated from the data. Instead, until now, there isn't a formula to estimate m_0; however, the seismologists assess that this value is supposed to be really very small and then, in particular, smaller than m_c.

The instrument used to model seismicity is given by point processes. These stochastic processes consist of punctual configurations in time and/or space. In this thesis, we will discard the spatial locations of the events: we will consider a marked point process on the line in which each element refers to the occurrence time of a shock in the earthquake sequence and the mark is given by the respective magnitude. More precisely, as typically done, we will model the seismic phenomenon as a marked branching process in which, at each generation, each event may give birth to its own aftershock activity independently of the other mother events. As we are going to explain in a moment, we will focus in particular on the events' magnitudes: it seems very natural to suppose that the aftershocks' magnitudes are correlated with the magnitude of the respective mother events. This hypothesis of magnitude correlation will be supported by some statistical analyses and will constitute the basis for the development of a new mathematical model for seismic sequences, with dependent marks.

In particular, the new model we are going to propose is a modified version of the temporal Epidemic Type Aftershock Sequence (ETAS) model, which belongs to the class of linear, stationary, marked Hawkes processes. The ETAS model is based on a specific branching process and actually represents a benchmark in statistical and mathematical seismology. For the first time it has been proposed in its pure temporal version by Ogata [Ogata, 1988], but some years later it has been improved by considering also the spatial location of the events [Ogata, 1998]. The process generates the events along consecutive generations. At any step, each event produces, independently of the others, a random number of aftershocks distributed as a Poisson random variable with rate depending on the generating event's magnitude through the *productivity law*. Occurrence times and magnitudes of each event are independent of each other and are distributed according to respectively the *Omori-Utsu law*, that is a power law, and the *Gutenberg-Richter law*, that is an exponential law (for the above-mentioned three laws see equations (2.2), (2.5) and (2.1)). Since it is expected that the magnitude process obeys the Gutenberg-Richter law, the deviations from this law are interpreted like missing measurements, then the minimum magnitude value after which there's agreement with the exponential decay is interpreted as the completeness threshold. In the classical ETAS model, the above-mentioned magnitudes are also independent of past seismicity. Finally, given an event, its aftershocks are spatially distributed according to a particular model centered at it (see [Ogata, 1998]). As already said, we will consider here only the temporal-magnitude analysis, discarding the spatial component.

This thesis finds its motivation exactly in the property, assumed in the standard ETAS model, according to which the magnitude of each event follows the Gutenberg-Richter law, independently of any past event. Actually, as said before, we find intuitive and more realistic to assume that the distribution of triggered events' magnitude depends on the triggering events' one, in analogy with the ETAS assumption that the number of aftershocks depends on the mother event's magnitude. In fact, in the recent literature these magnitude correlations were found statistically different from zero [Lippiello et al., 2007a,b, 2008; Sarlis et al., 2009, 2010] and independent of the incompleteness of the catalogs [Lippiello et al., 2012], in opposition to the previous assumptions that the above correlations are absent or due only to some problems of the catalogs [Corral, 2004, 2006; Helmstetter et al., 2006].

As anticipated, to the aim of finding further empirical evidences supporting this hypothesis, we will perform two different kinds of (time-magnitude) analysis of four real catalogs: three Italian datasets and a Californian one. The main problem is to determine the mother/daughter causal relations between events and the two analyses differ in this respect. From each catalog, we then obtain the law of triggered events' magnitude by a kernel density method. The results show that the density of triggered events' magnitude varies with the magnitude of their corresponding mother events, in other words it is a transition probability density. As the intuition suggests, the probability of having "high" values of the triggered events' magnitude increases with the mother events' magnitude. In addition, one can see a statistically significant increasing dependence of the aftershocks' magnitude means, again with respect to the triggering events' magnitude.

We will then propose a new version of the temporal ETAS model, in which the magnitudes are correlated by means of a class of transition probability density function for the triggered events' magnitudes, which includes also the case of independent magnitudes. The transition densities class is chosen in accordance with the results of the above experimental analyses and imposing other conditions. The most important one being that, averaging over all the mother events' magnitudes, we obtain again the well-validated Gutenberg-Richter law. This property is fundamental for our study and ensures the validity of this law at any event's generation when ignoring past seismicity. The explicit form of the class is given in (4.2).

The above class of transition probability densities is then tested by a simulation. More precisely, we will perform the same two types of analysis as before for some synthetic catalogs. We illustrate only the study concerning two of them, being the results for the other simulated datasets very similar. The two synthetic catalogs considered are obtained with two different approaches. The first one is simulated by the classical algorithms for the temporal ETAS. The second one is instead simulated with a new program in which the magnitudes of the triggered events are simulated according to the explicit form (4.2) of the new transition probability density functions. The results obtained for these two catalogs still support our hypothesis of correlated magnitudes.

The distribution of the time delay between two consecutive shocks, i.e., the interevent time, plays a very important role in the assessment of seismic hazard and scientific attention has been focused on it [Bak et al., 2002; Corral, 2003, 2004; Davidsen and Goltz, 2004; Molchan, 2005]. Following the Saichev and Sornette's approach for the classical ETAS model [Saichev and Sornette, 2007, 2013], we then study this random variable for our new ETAS model with correlated magnitudes. Due to the characteristics of the branching process involved and the homogeneity of the background component, both the classical and the new ETAS model are stationary and all the interevent times have the same distribution. Because of the incompleteness of the catalogs, it is interesting to study the interevent time between observable events, i.e., events with magnitudes bigger than the completeness threshold m_c. Then, in order to find the density $F_{inter}(\tau)$ of this variable, we will consider the probability generating function (PGF), that is a very important tool for the analysis of mathematical models in seismology [Özel and İnal, 2008; Saichev and Sornette, 2004, 2006a,b; Saichev et al., 2005]. In particular, we will consider the PGF of the total number of observable events in the time interval $[0, \tau]$ evaluated in zero, that is the probability of having zero observable events in the above time interval (void probability). More precisely, we derive an approximation of the probability $\mathbf{P}\{\tau\}$ of zero events with magnitude larger than m_c in $[0, \tau]$, for small τ. Recalling that, by the Palm equation [Cox and Isham, 2000], the density $F_{inter}(\tau)$ is obtained by scaling the double time derivative of $\mathbf{P}\{\tau\}$, we will then interpret the results as an approximation of the density $F_{inter}(\tau)$. The results we are going to derive are a generalization of the ones obtained in [Saichev and Sornette, 2007] for the classical time-magnitude ETAS model. Indeed, on the one hand, when the transition probability reduces to the Gutenberg-Richter law or when $m_c = m_0$, our results coincides with those by Saichev and Sornette. On the other hand, in our general framework, when $m_c > m_0$, the above approximations depend explicitly on the parameters involved in the explicit form of the transition probabilities.

This thesis is precisely structured as follows. Chapter 1 is focused on definitions, properties and examples of the point processes. We discuss also some mathematical concepts and techniques useful to study earthquake sequences, like the *Palm theory*. Some main classes of models very used in seismology are also presented here: the *Poisson process*, the *Hawkes processes* and the *stress-release model*. The Hawkes processes in particular have a relevant role, since they represent the class to which belongs the ETAS model. Chapter 2 is focused exactly on the latter model, discussing the non-explosion issue, the moment measures and giving a panoramic of its statistical analyses. Chapter 3 is devoted to our two statistical analyses of four real catalogs. In Chapter 4 we introduce and study our new ETAS model with correlated magnitudes. We show the non-explosion of the process, we give the Laplace functional equation and derive the explicit form for the class of the transition densities. Chapter 5 is then devoted to the experimental validation of our new model by analyzing simulated catalogs. Finally, in Chapter 6 we derive the approximations, for small τ, of the probability $\mathbf{P}\{\tau\}$ of zero events with magnitude larger than m_c in $[0, \tau]$, and of the interevent time density $F_{inter}(\tau)$.

The thesis ends with four appendices, containing a brief review on PGF and two classical statistical tests, some technical proofs and a public link to the codes of all the programs used in the thesis.

Chapter 1

Mathematical background

The theoretical study of earthquake sequences is based on a robust mathematical background that allows to model and review the phenomenon through an analytical approach. Probability and statistics theories give a strong support in this sense. It is since the first use of the terms “Statistical Seismology”, in 1928 by Kishinouye and Kawasumi [Kishinouye and Kawasumi, 1928], that the theoretical study of earthquakes has become a very important research field, committing a lot of scientists worldwide.

Just like in many scientific disciplines, such as for example biology or informatics, the mathematical tool for the modeling of seismic sequences is represented by the *point processes*. This chapter is devoted to definitions, properties, results and examples of the latter. The *Palm theory* will be also discussed, in order to explain the link between point processes and the probability density function of the interevent time between consecutive shocks. Let’s recall that this random variable has a very important meaning for seismic hazard scopes and is asymptotically analyzed in this study.

Finally, a particular attention is given to those processes typically used in seismology. Two of them have a particular relevance:

- *Poisson process* represents a fundamental basis for the mathematical modeling of earthquakes;

- *Hawkes processes* represent the general class to which belongs the ETAS model. The latter is a benchmark in statistical seismology; it is the starting point of the theoretical study in this thesis, thus all Chapter 2 will be devoted to it.

For the results in this chapter one can refer to [Brémaud, 2013; Daley and Vere-Jones, 2003, 2008].

1.1 Stochastic point processes

Point processes are stochastic processes whose realizations are punctual configurations in space and/or time. The points represent the events corresponding to some observable phenomenon; they may be labeled with respect to some of their features, like for example locations and/or occurrence times.

In order to define rigorously the point processes, let's consider a measurable space $(E, \mathcal{B}(E))$, where E is a locally compact second countable space (*l.c.s.c.*) and $\mathcal{B}(E)$ is the Borel sigma-field on it. Let's recall that a space has a *l.c.s.c.* topology if it is a Hausdorff (or separate) space, each point has a neighbourhood with a compact closure, and there exists a countable family of open sets $\{B_n\}_{n\geq 1}$, the countable basis, such that every open set B is the union of the open sets of this basis. A locally compact second countable space is in particular a complete, separable, metrizable, topological space. As an example, we can choose $(E, \mathcal{B}(E)) = (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, for some integer $k > 0$.

A measure on $(E, \mathcal{B}(E))$ is said to be *locally finite* if it assigns a finite mass to the relatively compact sets in E. Now, let's indicate with M_E^{lf} the set of all the locally finite measures on $(E, \mathcal{B}(E))$ and with $\mathcal{M}_E^{lf}$ the relative sigma-field generated by the collection of sets

$$\{\nu \in M_E^{lf} \ s.t. \ \nu(A) \in C\}, \ A \in \mathcal{B}(E), \ C \in \mathcal{B}(\mathbb{R}^+ \cup +\infty). \tag{1.1}$$

Then, the measurable space $(M_E^{lf}, \mathcal{M}_E^{lf})$ is the *canonical space of (locally finite) measures* on $(E, \mathcal{B}(E))$, and a measure $\nu \in M_E^{lf}$ is called a *point measure* taking values in $(\mathbb{N} \cup +\infty)$. One of the simplest examples of point measure is the *Dirac measure* $\delta_\omega(A) = \mathbb{1}_A(\omega)$, where $\mathbb{1}_A(\omega)$ is the indicator function having value one when $\omega \in A$, zero otherwise.

Definition 1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a given probability space. The *random measure* (measurable mapping)

$$N : (\Omega, \mathcal{F}) \to (M_E^{lf}, \mathcal{M}_E^{lf}),$$

defined on $(E, \mathcal{B}(E))$, is a *point process* if $N(\omega)$ is a point measure for all $\omega \in \Omega$.

Remark 1. An alternative definition of point process may be given in terms of a sequence of finite or countably finite random variables $\{y_i \ s.t. \ i \in \mathbb{N}_+\}$, with values in E, as stated in the following proposition.

Proposition 1. *Let $\{y_i\}$ be a sequence of E-valued random elements defined on a certain probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose that there exists an event $F_0 \in \mathcal{F}$ such that $\mathbb{P}\{F_0\} = 0$ and $\omega \notin F_0$ implies that for any bounded set*

$A \in \mathcal{B}(E)$, only a finite number of elements of $\{y_i(\omega)\}$ lies within A. Define $N(\cdot)$ to be the zero measure on F_0 and otherwise set

$$N(A) = \#\{y_i \in A\}, \quad A \in \mathcal{B}(E).$$

Then, $N(\cdot)$ is a point process.

Here we use the terms "bounded set" to indicate that it is relatively compact. The quantity $N(A)$ is a random variable if N is a random measure. In fact, it consists of the mapping $\omega \to N(\omega)(A)$, obtained by composing the measurable mappings

$$\omega \to N(\omega), \quad \nu \to \nu(A).$$

The measurability of this random variable need not to be verified for all the sets $A \in \mathcal{B}(E)$: it is enough to consider a smaller class of sets generating this sigma-field, as stated in the theorem below.

Theorem 1. *i) For $N : \Omega \to M_E^{lf}$ to be measurable it suffices that $N(I) : \Omega \to (\mathbb{R}^+ \cup +\infty)$ is a random variable for all $I \in H_0$, where H_0 is a collection of relatively compact subsets of E generating $\mathcal{B}(E)$, that are closed under finite intersection and such that there exists a sequence $\{E_n\}_{n\geq 1}$ such that either $E_n \uparrow E$, or the E_n's form a partition of E.*

ii) If $\mathfrak{I}$ is the collection of subsets of M_E^{lf} of the form $\{\nu | \nu(I) \in A\}$, where $I \in H_0$ and $A \in \mathcal{B}(\mathbb{R}^+ \cup +\infty)$, then $\mathcal{M}_E^{lf} = \sigma(\mathfrak{I})$.

Remark 2. A locally finite point process N can be rewritten also as $\sum_{n\in\mathbb{N}} \delta(X_n)$, for a sequence of random variables $\{X_n\}_{n\geq 1}$ with values in $E \cup \{a\}$, for a given $a \notin E$, endowed with the sigma field generated by $\mathcal{B}(E)$ and $\{a\}$. The point a plays the role of infinity.

There may be situations in which one considers $N(\{\omega\})$ only *sigma-finite*, for $\omega \in \Omega$. Then, we have to consider the set of the measures of all kinds M_E, where E can now be taken as an arbitrary measurable space endowed with the sigma field $\mathcal{E}$.

Definition 2. Let's consider a random measure on $(E, \mathcal{E})$ $N : (\Omega \times \mathcal{E}) \to (\mathbb{R}^+ \cup +\infty)$. If there exists a sequence $\{K_n\}_{n\geq 1}$ of measurable sets increasing to E and such that $N(\omega, K_n) < \infty$ for all $\omega \in \Omega$ and $n \geq 1$, then the random measure N is called *sigma-finite*.

Now, recalling that the singletons are sets consisting of a single element, we give the following definition.

Definition 3. If $\mathcal{E}$ contains all the singletons of E, then a point measure ν on $(E, \mathcal{E})$ is called a *simple measure* if

$$\nu(\{e\}) \in \{0; 1\}, \quad \forall e \in E.$$

Consequently, a point process N on $(E, \mathcal{E})$ is a *simple process* if $N(\{\omega\})$ is a simple measure, for all $\omega \in \Omega$.

The examples of point processes are typically specified through their *finite-dimensional distributions (fidi)*. In order to define the latter, let's consider again the probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ and the random measure N on $(E, \mathcal{E})$. The probability $P_N := P \circ N^{-1}$, defined on $(M_E^{lf}, \mathcal{M}_E^{lf})$, is the *distribution* of N. Now, the *fidi* distributions of a point process N are specified by consistent joint distributions

$$P_k(A_1, \dots, A_k; n_1, \dots, n_k) = \mathbb{P}\{N(A_1) = n_1, \dots, N(A_k) = n_k\},$$

indicating the number of points falling in finite collection of bounded Borel sets $A_\ell \in \mathcal{B}(E)$, for $\ell = 1, \dots, k$ and for all $k \in \mathbb{N}_+$. Here the conditions for consistency can be expressed by

$$\sum_{r=0}^{n} P_k(A_1, \dots, A_k; n-r, r, n_3, \dots, n_k) = P_{k-1}(A_1 \cup A_2, A_3, \dots, A_k; n, n_3, \dots, n_k)$$

and

$$\lim_{A_k \downarrow \varnothing} P_1(A_k; 0) = 1,$$

for all the sequences $\{A_k\}_{k \in \mathbb{N}_+}$ of bounded Borel sets.

If we suppose that E is *l.c.s.c.* and we consider again the collection H_0 of relatively compact subsets of E, generating $\mathcal{B}(E)$ and closed under finite intersection, then if $\mathfrak{I}$ is the collection of subsets of M_E^{lf} of the form

$$\{\nu \in M_E^{lf} \ s.t. \ \nu(I_\ell) \in A_\ell, \quad \ell = 1, \dots, k\},$$

where $I_\ell \in H_0$ and $A_\ell \in \mathcal{B}(E)$, then $\mathfrak{I}$ is stable under finite intersection. By the well-known Dynkin's Lemma, it follows that if two probability measures agree on $\mathfrak{I}$, so they do on $\sigma(\mathfrak{I})$, that is equal to $\mathcal{M}_E^{lf}$ by Theorem 1. Then, the following theorem holds.

Theorem 2. *Let E be l.c.s.c. and consider the random measure N on $(E, \mathcal{E})$. Its distribution is completely characterized by the distributions of the vectors $(N(I_1), \dots, N(I_k))$, for all $(I_1, \dots, I_k) \in H_0$ and $k \in \mathbb{N}_+$, where H_0 is defined as above.*

As a consequence, the *fidi* distribution restricted to $\mathfrak{I}$ as in Theorem 2, completely characterizes a locally finite point process. This is true also for sigma-finite point processes on a *l.c.s.c.* space, since this is true for the restrictions of this point process to the sequence $\{K_n\}_{n\geq 1}$ of measurable sets increasing to E and such that $N(\omega, K_n) < \infty$, for all $\omega \in \Omega$ and $n \geq 1$, as in Definition 2.

In many applications, the point processes considered are defined on the real line and are *stationary*. The latter means that the process remains invariant under translation, that is

$$P_k(A_1, \dots, A_k; n_1, \dots, n_k) = P_k(A_1 + s, \dots, A_k + s; n_1, \dots, n_k), \quad \forall s \in \mathbb{R},$$

where $\Big\{A_\ell + s = \{t + s,\ t \in A_\ell\}\Big\}_{\ell=1,\dots,k}$.

Let's conclude this section specifying that, for simplicity of notations, in what follows we will use the terms "a point process on E" excluding the dependence on the relative sigma field, too.

1.2 Finite point processes

Let's consider a measurable space $(E, \mathcal{E})$ and a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let's consider also the *finite configuration space* $M^{fc}(E)$, consisting of the collection of the configurations in E. A *configuration* $\mathbf{x} = (x_1, \dots, x_n)$ with cardinal $n \in \mathbb{N}$ is a finite, unordered sequence of points in E, with repetitions allowed. This means that multiple points $(x_i = x_j,\ i \neq j)$ may belong to the sequence. In the cases of time locations, this possibility is instead excluded. Let $\mathcal{M}^{fc}(E)$ be the relative sigma field, generated by the mappings from $\mathbf{x}$ to the cardinality of $\mathbf{x} \cap A$, for all the measurable sets $A \in \mathcal{E}$.

Definition 4. A *finite point process* N^{fin} on E is the following random measure:

$$N^{fin} : (\Omega, \mathcal{F}) \to (M^{fc}(E), \mathcal{M}^{fc}(E)).$$

For such a process we can define on $\mathbb{N}$ the *count distribution*

$$p_n = \mathbb{P}\{N(E) = n\}, \quad n \in \mathbb{N},$$

determining the total number of points in the process, with $\sum_{n=0}^{\infty} p_n = 1$. Furthermore, if the total number of events is $n \in \mathbb{N}$, we can also define the probability distribution Π_n on $E^{(n)} = \underbrace{E \times \cdots \times E}_{n-times}$, determining the joint distribution of the positions of the points belonging to the process. Since we consider unordered sequences, the latter probability distribution should be symmetric. If it isn't already symmetric, we can obtain this property by introducing

$$\Pi_n^{sym}(A_1 \times \cdots \times A_n) = \frac{1}{n!} \sum_{perm} \Pi_n(A_{i_1} \times \cdots \times A_{i_n}),$$

that is the symmetrized form for any partition $(A_1, \cdots, A_n)$ of E, where the sum is taken over the $n!$ permutations $(i_1, \ldots, i_n)$ of the integers $(1, \cdots, n)$. It is important to consider also non-probability measures which, as we are going to see, can be interpreted in terms of counting measures and are useful in terms of conditions for a process to be simple. Let's define then the *Janossy measures* as

$$J_n(A_1 \times \cdots \times A_n) = p_n \sum_{perm} \Pi_n(A_{i_1} \times \cdots \times A_{i_n}).$$

If the derivatives exist, we can introduce the densities of $J_n(\cdot)$. For example, if $E = \mathbb{R}^d$, we can define the *Janossy densities* as

$$j_n(x_1, \ldots, x_n) dx_1 \cdots dx_n = \mathbb{P} \left\{ \begin{array}{c} \text{The process has exactly } n \text{ points,} \\ \text{one in each of the } n \text{ distinct} \\ \text{infinitesimal locations } dx_1, \ldots, dx_n \end{array} \right\}.$$

The Janossy measure may be interpreted in terms of the count distribution $\{p_n\}_{n \in \mathbb{N}}$ by setting

$$J_0(E^{(0)}) = p_0, \quad J_n(E^{(n)}) = p_n \sum_{perm} \Pi_n(E^{(n)}) = p_n n!, \quad \text{for } n \geq 1.$$

Then, the normalization condition $\sum_{n=0}^{\infty} p_n = 1$ for the count distribution becomes

$$\sum_{n=0}^{\infty} \frac{J_n(E^{(n)})}{n!} = 1.$$

Now, if we consider a finite partition $(A_1, \ldots, A_k)$ of E, the probability of having respectively $(n_1, \ldots, n_k)$ points in the k elements of the partition, with $\sum_{\ell=1}^{k} n_\ell = n$, is easily obtained as

$$\mathbb{P}\{N(A_1) = n_1, \ldots, N(A_k) = n_k\} = \frac{J_n(A_1^{(n_1)} \times \cdots \times A_k^{(n_k)})}{n_1! \cdots n_k!}. \tag{1.2}$$

Furthermore, if the $A_1, \dots, A_k$ are disjoint and $\overline{A}$ is the complement of their union, the above Janossy measure becomes

$$J_n(A_1^{(n_1)} \times \cdots \times A_k^{(n_k)}) = \sum_{\ell=0}^{\infty} \frac{J_{n+\ell}(A_1^{(n_1)} \times \cdots \times A_k^{(n_k)} \times \overline{A}^{(\ell)})}{\ell!}.$$

As anticipated before, the definition of the Janossy measures allows to obtain a necessary and sufficient condition for a point process to be simple. In fact, the case in which $J_n(\cdot)$ assigns a non-zero mass to at least one of the diagonal sets $\{x_i = x_j\}$, is equivalent to a strictly positive probability of having two coincident points at the same location. We can then state the following proposition.

Proposition 2. *i) A necessary and sufficient condition for a point process to be simple is that the associated Janossy measure $J_n(\cdot)$ assigns zero mass to all the diagonal sets $\{x_i = x_j\}$, for all $n = 1, 2, \dots$.*

ii) If $E = \mathbb{R}^d$, the point process is simple if the Janossy measures have densities $j_n(\cdot)$, for all $n = 1, 2, \dots$, with respect to the (nd)-dimensional Lebesgue measure.

It is important to say that the Janossy densities play a relevant role in the likelihood analysis of finite point processes. To discuss this issue, let's define at first the local Janossy measures and densities.

Definition 5. Given any bounded Borel set A, we define the *local Janossy measures* localized to A as

$$J_n(dx_1 \times \cdot \times dx_n; A) = \mathbb{P}\left\{\begin{array}{c}\text{The process has exactly } n \\ \text{points in } A \text{, one in each of the } n \\ \text{distinct locations } dx_1, \dots, dx_n\end{array}\right\},$$

for $x_i \in A$, $i = 1, \dots, n$. If the above measures have densities, the latter are called the *local Janossy densities*.

The local Janossy measures allow to define also the property of regularity.

Definition 6. A point process N defined on $E = \mathbb{R}^d$ is regular on a given bounded set $A \subseteq \mathcal{B}(\mathbb{R}^d)$ if, for all $n \geq 1$, the local Janossy measures $J_n(dx_1 \times \cdots \times dx_n; A)$ are absolutely continuous on $A^{(n)}$ with respect to the Lebesgue measure in $E^{(n)}$. The point process N is *regular* if it is regular on all the bounded sets $A \in \mathcal{B}(\mathbb{R}^d)$.

By recalling the above Proposition 2, we can deduce that a regular process is simple.

Now, recalling that the likelihood is a function of the parameters in the joint density and that the latter, in the case of finite processes, is simply the probability of the points to occur exactly at locations x_i in a given bounded Borel set A, we can give the following definition.

Definition 7. The *likelihood* of a realization $(x_1, \ldots, x_n)$ of a regular point process N, defined on a bounded set $A \in \mathcal{B}(\mathbb{R}^d)$, is the local Janossy density $j_n(x_1, \ldots, x_n; A)$, with $n = N(A)$.

1.3 Moment measures and probability generating functionals

Let's consider a point process N defined on the *l.c.s.c.* space E equipped with the Borel sigma field $\mathcal{B}(E)$. If the process has finite mean $\mathbb{E}\big[N(A)\big]$, for any bounded $A \in \mathcal{B}(E)$, then we can define the *first (order) moment measure*, or *mean measure*, of the N as

$$M(A) = M_1(A) = \mathbb{E}\big[N(A)\big].$$

This measure is finite additive since the point process is a random measure on $(E, \mathcal{B}(E))$. Furthermore, if we consider an increasing monotone sequence of bounded Borel sets $\{A_n\}$, for $n = 1, 2, \ldots$, converging to A as n tends to ∞, we have also that $\lim_{n\to\infty} M(A_n) = M(A)$; then, $M(A)$ is also sigma additive. When $M(\cdot)$ exists, we can also define the random integrals $\int_E f(x)N(dx)$, for f measurable function with bounded support on E. Recalling Remark 2, we can say that the above integral is the same mathematical object of $\sum_{n\in\mathbb{N}} f(X_n)$, where the sum extends only to those indices n such that $X_n \in E$ without points "at infinity": the value $f(a)$, for a as in Remark 2, is not defined. If $f = \mathbb{1}_A$, for any bounded $A \in \mathcal{B}(E)$, it follows that

$$M(A) = \mathbb{E}\left[\int_A f(x)N(dx)\right].$$

Linear combination and monotone limits allow to obtain the result of the *Campbell's theorem* concerning the above expected value:

$$\mathbb{E}\left[\int_E f(x)N(dx)\right] = \int_E f(x)M(dx). \tag{1.3}$$

This equation has been firstly studied by Campbell in his work in 1909 [Campbell, 1909] and can be generalized in the *Campbell measures*, which represent a very important object in the theory of point processes.

The concept of first order moment measure can be extended to a generic k^{th} (order) moment measure. To this aim, we have to consider the measure $N^{(k)}$, defined on the rectangles $A_1 \times \cdots \times A_k$, for bounded $\{A_\ell\}_{\ell=1,\dots,k} \in \mathcal{B}(E)$, given by the k-fold product of N with itself. More precisely, $N^{(k)}$ is the point process consisting of all the k-tuples of points from the original realization, with repetitions allowed and distinguishing the order. The k^{th} *(order) moment measure* is the expected value of such a process, that is

$$M_k(A_1 \times \cdots \times A_k) = \mathbb{E}\Big[\prod_{\ell=1}^{k} N(A_\ell)\Big]. \tag{1.4}$$

Just like the mean measure, the k^{th} moment measure is finite additive in the above k-dimensional rectangle. The Campbell's theorem can be generalized for this latter measure as follows.

Proposition 3. *If the k^{th} moment measure exists, then the random integrals $\int_E f(x)N(dx)$, for f measurable function with bounded support on E, has finite k^{th} moment, given by*

$$\mathbb{E}\bigg[\Big(\int_E f(x)N(dx)\Big)^k\bigg] = \int_{E^{(k)}} f(x_1)\cdots f(x_k)M_k(dx_1 \times \cdots \times dx_k).$$

A relation that connects first and second order measures, very often considered in the theory of point process, is the *covariance measure*:

$$C_2(A \times B) = M_2(A \times B) - M_1(A)M_1(B),$$

for disjoint bounded sets $A, B \in \mathcal{B}(E)$.

The measure M_k can be extended to arbitrary rectangle sets $A_1^{(k_1)} \times \cdots \times A_r^{(k_r)}$, for $\sum_{i=1}^r k_i = k$ and $k_i \geq 1$, $i = 1, \dots, r$, where the $\{A_i\}_{i=1,\dots,r}$ are disjoint bounded sets of $\mathcal{B}(E)$. Then, we obtain

$$M_k(A_1^{(k_1)} \times \cdots \times A_r^{(k_r)}) = \mathbb{E}\Big[\big(N(A_1)\big)^{k_1} \cdots \big(N(A_r)\big)^{k_r}\Big]. \tag{1.5}$$

The point process $N^{(k)}$ for which we compute the above expectation to define M_k, may contain multiple points. In order to avoid this case, we can define the k^{th} *factorial moment measure* $M_{[k]}$ as

$$M_{[1]}(A) = M_1(A) \qquad \text{and}$$
$$M_{[k]}(A_1^{(k_1)} \times \cdots \times A_r^{(k_r)}) = \mathbb{E}\Big[\big(N(A_1)\big)^{[k_1]} \cdots \big(N(A_r)\big)^{[k_r]}\Big], \quad k > 1,$$

for $\{A_i\}$ and k_i, $i = 1, \dots, r$, as in (1.5). In this case, the underlying process consists of all the k-tuples of distinct points from the original realization, distinguishing the order but not allowing repetitions. This means that a double point is labeled as two distinct points with the same coordinates.

In the case of finite point processes, recalling (1.2), we can write $M_{[k]}$ in terms of the Janossy measures. If $\mathbb{E}\left[\left(N(E)\right)^{[k]}\right] < \infty$ and $\{A_i\}_{i=1,\dots,r}$ is a partition of E, we have that

$$M_{[k]}(A_1^{(k_1)} \times \cdots \times A_r^{(k_r)}) = \sum_{u_i \geq k_i, i=1,\dots,r} \frac{J_{u_1+\cdots+u_r}(A_1^{(u_1)} \times \cdots \times A_r^{(u_r)})}{\prod_{i=1}^r (u_i - k_i)!}.$$

Finally, let's discuss about the *probability generating functionals.* At first, let's focus again on finite point processes.

Definition 8. Let $\mathcal{U} : E \to \mathbb{C}$ be the class of bounded, complex-valued, Borel, measurable functions $\zeta(\cdot)$ such that $|\zeta(x)| \leq 1$ for all $x \in E$. The probability generating functional for a finite point process, defined for the function $\zeta(\cdot)$ is

$$G[\zeta] = \mathbb{E}\left[\prod_{i=1}^{N} \zeta(x_i)\right], \tag{1.6}$$

where the above product becomes zero if $N > 0$ and $\zeta(x_i) = 0$ for some i, and becomes one if $N = 0$.

Let's observe that, since $\zeta(\cdot)$ is a bounded, complex-valued, Borel, measurable function, the random product $\prod_{i=1}^{N} \zeta(x_i)$ is well defined for a realization of a finite point process. Furthermore, since $|\zeta(x)| \leq 1$, the expected value of the above product exists and is finite. The probability generating functional may be reduced to the probability generating function by considering a measurable partition $A_1, \dots, A_k$ of E and by setting

$$\zeta(x) = \sum_{\ell=1}^{k} z_\ell \mathbb{1}_{A_\ell}(x),$$

where $\mathbb{1}_A(\cdot)$ is the indicator function of A and where $|z_\ell| \leq 1$, for $\ell = 1, \dots, k$. Then, one can easily obtain that

$$G\left[\sum_{\ell=1}^{k} z_\ell \mathbb{1}_{A_\ell}(\cdot)\right] = \mathbb{E}\left[\prod_{\ell=1}^{k} z_\ell^{N(A_\ell)}\right].$$

In the case of a general point process, the probability generating functional is defined as above, provided that the functions considered are now defined in the space $\mathfrak{H}(E)$, consisting of the non-negative, measurable functions $h(\cdot)$ bounded by unity and such that $h(x) = 1$ outside some bounded set. In this case, the following proposition holds.

Proposition 4. *Let $G[\cdot]$ be the probability generating functional of a point process with finite k^{th} order moment measure, for $k \in \mathbb{N}$. Then, for $(1-h) \in \mathfrak{H}(E)$ and $\vartheta \in (0,1)$,*

$$G[1 - \vartheta h] = 1 + \sum_{\ell=1}^{k} \frac{(-\vartheta)^\ell}{\ell!} \int_{E^{(\ell)}} h(x_1) \cdots h(x_\ell) M_{[\ell]}(dx_1 \times \cdots \times dx_\ell) + o(\vartheta^k).$$

1.4 Marked and cluster processes

In this section, we are going to present two classes of widely used processes, both in theoretical and applied studies. The first class consists of the marked point processes, very useful when one wants to look at a point process as part of a more complex model. In this case, the point process itself represents the component of the model which carries all the information about time occurrence or spatial locations of the elements, having themselves a stochastic evolution. The second class consists instead of the cluster point processes, which are very important to model the locations of objects in one or three-dimensional space. One of their main applications is seismology. In this case, they model the locations of earthquake's epicenters and may naturally describe the branching evolution of seismic generations. These two classes of processes are connected in fact, if one considers the point process associated to the events' (time or space) locations $\{x_i\}$, not necessary simple, then this process may have the following two interpretations:

- the underlying process (or ground process, as we are going to explain in a moment) of a marked point process, or
- the cluster process in which the cluster centers are the locations of the events and the cluster elements are the pairs (x_i, κ_{ij}), where the $\{\kappa_{ij}\}$ represent the marks associated to the elements located in $\{x_i\}$.

1.4.1 Marked processes

Let's consider a *l.c.s.c.* space E and its Borel sigma field $\mathcal{B}(E)$. The first important component of a marked point process is the *ground process* N_g, with values in E, indicating the process of the events' locations which can be either their times of occurrence or their spatial positions. Let's consider also another measurable space $(K, \mathcal{B}(K))$, where K is again *l.c.s.c.* and $\mathcal{K}$ is the relative sigma field.

Definition 9. A *marked point process (MPP)*, with locations in E and marks in K, is a locally finite point process N consisting of the pairs $(x_n, \kappa_n)_{n\in\mathbb{N}}$ in $(E \times K)$, with the additional property that the ground process $N_g(\cdot)$ is itself a point process: $N_g(A) = N(A \times K) < \infty$ for all the bounded sets $A \in \mathcal{B}(E)$.

The ground process N_g of the marked point process N is also indicated as the *marginal process of locations*.

Definition 10. A *multivariate (or multitype) point process* is a marked point process with mark space give by the finite set $\{1, 2, \ldots, k\}$, for some $k \in \mathbb{N}$.

Thanks to the fact that the mark space is finite, each component process $N_\ell(\cdot) = N(\cdot \times \ell)$ in the multivariate point process is locally finite and the ground process can be written as

$$N_g(\cdot) = N(\cdot \times \{1, 2, \ldots, k\}) = \sum_{\ell=1}^{k} N_\ell(\cdot).$$

An equivalent way of defining a marked point process may be of considering a not necessary simple point process on $(E, \mathcal{B}(E))$, equipped with the sequences of K-valued random variables representing the marks.

As already specified, the MPP are widely used in applications. This depends also on the fact that the form of the marks and the dependence relation between the ground and the mark processes may be of many types. One important example is given by the process in which the ground component is defined on the real line and the elements' marks consist of some of the features inherited by the past history of the process, up to the time of each event considered.

The two properties of simpleness and stationarity of general point processes can be stated for the special case of MPP as follows.

Definition 11. i) The marked point process is simple if so is the ground process.

ii) The marked point process defined on $\mathbb{R}^d$ is stationary if its the probability structure remains invariant under translations in $\mathbb{R}^d$.

Finally, we can give the two other definitions characterizing two important kinds of dependence concerning the mark component of MPPs.

Definition 12. Let's consider a marked point process $N = \{x_n, \kappa_n\}_{n\in\mathbb{N}}$ on $(E \times K)$.

i) The MPP has *independent marks* if, given the ground process, the marks are mutually independent random variables with distribution depending only on the relative locations.

ii) When $E = \mathbb{R}$, the MPP is said to have *unpredictable marks* if the distribution of the i^{th} mark, associated to the i^{th} location, does not depend on the pair $\{x_j, \kappa_j\}$, for $x_j < x_i$.

In the case of MPP with independent marks, we can state the following theorem, which concerns the structure of such a process.

Theorem 3. *Let's consider a marked point process N with independent marks.*

i) The probability structure of N is completely defined by the distribution of the ground process N_g and the mark kernel $\{M_K(K|x) : K \in \mathcal{B}(K)), x \in E\}$, which represents the distribution of the marks conditioned to the location x.

ii) Let's consider the space $\mathfrak{H}(E \times K)$ of the measurable functions $h(x, \overline{\kappa})$ between zero and one such that $h(x, \overline{\kappa}) = 1$ for all $\overline{\kappa} \in K$ and $x \notin A$, for some bounded set A. The probability generating functional for N is

$$G[h] = G_g[h_{M_K}], \quad h \in \mathfrak{H}(E \times K),$$

where $G_g[\cdot]$ is the probability generating functionals of $N_g[\cdot]$ and $h_{M_K}(x) = \int_K h(x, \overline{\kappa}) M_K(d\overline{\kappa}|x)$.

iii) The moment measure M_r of order r for N exists if and only if the corresponding moment measure M_r^g for the ground process N_g exists. In this case,

$$M_r(dx_1 \times \cdots \times dx_r \times d\kappa_1 \times \cdots \times d\kappa_r) = M_r^g(dx_1 \times \cdots \times dx_r) \prod_{i=1}^{r} M_K(d\kappa_i|x_i).$$

Similar representations hold for factorial moments and covariance measure.

1.4.2 Cluster processes

The modeling of a phenomenon through a cluster process allows to distinguish two different components. The first one is relative to the cluster centres: they identify the clusters themselves, are often unobserved and are modeled with a given process N_c, whose generic realization are the points $\{y_i\}$ in a given *l.c.s.c.* space Y. The second component is instead relative to the elements within each cluster; it is modeled by a countable family of component point processes $N(\cdot|y_i)$, labeled with the centres of the clusters. Together, the two components of above constitute the observed process. One of the applications in which this class of processes is particularly used is again seismology. Since the earthquake sequences are often described as branching processes, evolving in consecutive generations, it is natural to consider the ancestors, or triggering events, as cluster centres, and the relative triggered shocks as the elements of the clusters. In this case, the elements representing the centres may be considered as clusters' elements, too. In fact, there isn't a particular characteristic that allows to distinguish between these two kinds of elements.

In order to give a rigorous definition of the cluster process, let's consider the two *l.c.s.c.* spaces E and Y.

Definition 13. A *cluster process* N is a process on E such that, for any bounded set $A \in \mathcal{B}(E)$,

$$N(A) = \int_Y N(A|y)N_c(dy) = \sum_{y_i \in N_c(\cdot)} N(A|y_i) < \infty, \quad \text{almost surely (a.s.)} \tag{1.7}$$

where $N_c(\cdot)$ is the centre process on the *l.c.s.c.* space Y and $\{N(\cdot|y) : y \in Y\}$ is the measurable family of the component processes.

Let's observe that the superposition of the clusters must be almost surely locally finite, but this condition is not necessary for the individual clusters. It could be also useful to define an *independent cluster process*, in which the component processes are required to be mutually independent and then they are supposed to come from an *independent measurable family.* Furthermore, the case in which the cluster process and the component processes are defined on the same space, one has to impose that the translated distributions $N(A-y|y)$ are identically distributed. This is the natural candidate for a stationary version of the cluster process. Instead, we will say that the cluster process has *stationary components* if the process relative to the cluster centres is stationary and the distribution of the cluster elements depends only on the relative positions between the elements themselves and the cluster centres.

The issue of existence of a cluster process is quite difficult to solve. In fact, if the dimension of the space state increases, so does the number of clusters of points contributing to a given set. However, the conditions for the existence of the process may be obtained formally when one considers independent clusters, as stated in the following theorem.

Theorem 4. *An independent cluster process exists if and only if*

$$\int_Y p_A(y) N_c(dy) = \sum_{y_i \in N_c(\cdot)} p_A(y_i) < \infty, \quad \Pi_c \ a.s.,$$

for any bounded set $A \in \mathcal{B}(E)$, where $p_A(y) = \mathbb{P}\{N(A|y) > 0\}$, for $y \in Y$, and Π_c is the probability measure of the process of cluster centres.

Remark 3. When the cluster process has stationary components, that is the cluster centres process is stationary and the distribution of the cluster events depends only on their distances from the relative centres, a sufficient condition for the existence of the cluster process is the finiteness of the mean cluster size. In the case of the Poisson processes, this condition is not necessary.

Finally, let's discuss about the moments and the probability generating functional. If we take the conditional expectations on the cluster centres in the first and the last members of equation (1.7), we obtain

$$\mathbb{E}\big[N(A)|N_c\big] = \sum_{y_i \in N_c(\cdot)} M_1(A|y_i) = \int_Y M_1(A|y) N_c(dy), \tag{1.8}$$

where $M_1(\cdot|y)$ is the first moment measure of the process relative to the elements of the cluster centred at y, if the latter exists. Since the processes of the cluster elements form a measurable family, $M_1(A|y)$ (if exists) defines a measurable kernel, that is a measure in A for every y and a measurable function in y for each given Borel set $A \in \mathcal{B}(E)$. Then, by taking the expectations of (1.8) with respect to the centres, we get

$$\mathbb{E}\big[N(A)\big] = \int_Y M_1(A|y) M^c(dy),$$

for any bounded set $A \in \mathcal{B}(E)$, where $M^c(\cdot) = \mathbb{E}[N_c(\cdot)]$ is the first moment measure for the centres cluster process. It follows that the first moment measure exists if and only if the latter integral is finite.

For what matter the second factorial moment measure, given the process of the centres and the independence of the clusters, we obtain

$$\begin{aligned} M_{[2]}(A \times B) = &\int_Y M_{[2]}(A \times B|y) M^c(dy) \\ &+ \int_{Y \times Y} M_1(A|y_1) M_1(A|y_2) M^c_{[2]}(dy_1 \times dy_2), \end{aligned} \tag{1.9}$$

where $M^c_{[2]}(\cdot)$ is the second factorial moment measure for the cluster centres. Equation (1.9) includes both the possibilities that the two different points considered, falling in the product space $A \times B$, for $A, B \in \mathcal{B}(E)$, may belong to the same cluster or not. Again, if the component measures exist and the integrals in (1.9) are finite, then the second factorial moment measure exists. Concluding, whenever the cluster process exists, if one assumes independent clusters and consider $h \in \mathfrak{H}(E)$, where we recall that the latter is the space of non-negative, measurable functions bounded by unity and equal to one outside some bounded set, then the probability generating functional for the cluster process is

$$G[h] = \mathbb{E}\big[G[h|N_c]\big] = G_c[G_m[h|\cdot]],$$

where, for $h \in \mathfrak{H}(E)$, $G_m[h|y]$ is the probability generating functional of $N(\cdot|y)$. The conditional probability generating functionals relative to N given N_c is instead

$$G[h|N_c] = \prod_{y_i \in N_c} G_m[h|y_i].$$

1.5 Stationary point processes on $\mathbb{R}$

The point processes defined on the line are widely used in applied studies, thanks to their simplicity and applicability. They consist of points representing the events' occurrence times of a given phenomenon. In particular, this is the case of the shocks in an earthquake sequence.

The characterization of such processes can be made in terms of four different but connected well defined random variables. The first characterization is through the counting measures $N(\cdot)$, that indeed count the number of events of the process falling in a given set. More precisely, we can define the counting measure

$$N(A) = \#\{t_i \in A\},$$

for any Borel subset A of the real line. This random variable is non-negative and takes integer values, also infinite. In order to exclude the possibility of many points occurring very close, $N(\cdot)$ is required also to be finite for any bounded set. Furthermore, given an infinite sequence of mutually disjoint sets $(A_1, A_2, \dots)$, such that $A = \bigcup_{i=1}^{\infty} A_i$, it holds

$$N(A) = N\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} N(A_i).$$

This first characterization is strictly connected to the second one. In fact, if $A = (0, t]$, we can write

$$N(t) = \begin{cases} N((0,t]), & \text{if } t > 0 \\ 0, & \text{if } t = 0 \\ -N((t,0]), & \text{otherwise.} \end{cases}$$

Then, the point process on the line is characterized through the non-decreasing, right-continuous, integer-valued step function $N(t)$. Let's specify that the notation $N(t)$ is used for the number of events in the interval $(0, t]$; instead, we will use $N(\{t\})$ to indicate the number of events occurred in the exact time t.

In order to present the third way of describing a point process on $\mathbb{R}$, let's start with the case of the positive half-line. We can consider the sequence of increasing times

$$t_i := \inf\{t > 0 |\ N(t) \geq i\}, \quad i = 1, 2, \ldots.$$

Since it obviously follows that

$$\{N(t) \geq i\} \quad \Leftrightarrow \quad \{t \geq t_i\},$$

it is equivalent to specify the sequence $\{t_i\}_{i=1,2,\ldots}$ or the above step function $N(t)$, when $N((-\infty, 0]) = 0$. If we extend this result to the hole line, the sequence of increasing times becomes the doubly infinite sequence

$$t_r = \begin{cases} \inf\{t > 0 |\ N((0,t]) \geq r\}, & \text{if } r > 0 \\ -\inf\{t > 0 |\ N((-t,0]) \geq -r+1\}, & \text{if } r \leq 0, \end{cases} \tag{1.10}$$

with the property that $t_0 \leq 0 < t_1$ and $t_r \leq t_{r+1}$ for all r.

The last characterization can be obtained from (1.10) by setting

$$\tau_r = t_r - t_{r-1}, \quad r = 1, 2, \ldots.$$

This sequence of intervals (*interevent times*) and one of the times $\{t_r\}$, usually t_0, completely describe the process on the real line.

The properties of stationarity, regularity and simplicity in the case of point processes on $\mathbb{R}$ can be stated as follows.

Definition 14. Let N be a point process on $\mathbb{R}$. For all $t \in \mathbb{R}$, we say that:

- the process is *simple (or crude) stationary* when the distribution of $N((t, t+s])$ depends only on the length s and not on the location t;

- the process is *stationary* when, for all bounded Borel subsets $\{A_\ell\}_{\ell=1,\dots,k} \in \mathbb{R}$ and $k = 1, 2, \dots$, the joint distribution of

$$\{N(A_1 + t), \dots, N(A_k + t)\}$$

  is independent of t;

- the process is *regular* when

$$\mathbb{P}\{N((t, t + \varepsilon]) > 1\} = o(\varepsilon);$$

- the process is *simple* when

$$\mathbb{P}\Big\{N(\{t\}) \in \{0; 1\}\Big\} = 1.$$

Let's observe that crude stationary is weaker than stationary and regularity guarantees simplicity. By recalling the characterization of the point process on the line through the interevent times, we can give this further definition.

Definition 15. A point process is *interval stationary* if, for every $t \in \mathbb{R}$ and all the integers $i_1, \dots, i_t$, the joint distribution of $\{\tau_{i_1+k}, \dots, \tau_{i_t+k}\}$ doesn't depend on $k \in \mathbb{Z}$.

The moment measures in the case of point processes on the line are defined as for the general case. In particular, the function

$$M(t) = \mathbb{E}\big[N(t)\big]$$

is non-negative and satisfies the Cauchy's functional equation

$$M(t + s) = M(t) + M(s), \quad \text{for } t, s \in [0, \infty).$$

By defining the *mean density* as $m_d = M(1) = \mathbb{E}\big[N((0, 1])\big]$, it can be proved that $M(t) = mt$, for $t \in [0, \infty)$. Furthermore, if the second moment measure $M_2(t)$ exists and is finite and the process is ergodic, meaning that

$$\mathbb{P}\left\{\lim_{t\to\infty} \frac{N(t)}{t} = m_d\right\} = 1,$$

we have that $M_2(t) \sim (m_d t)^2$ as t tends to infinity. This is because

$$\lim_{t\to\infty} var(N(t)/t) = 0$$

for an ergodic process N with a finite second moment.

In order to measure the rate of occurrence of a stationary point process, we can state the following *Khinchin's existence theorem.*

Theorem 5. *If N is a stationary point process, then the limit*

$$\lambda = \lim_{t \downarrow 0} \frac{\mathbb{P}\{N(t) > 0\}}{t} \tag{1.11}$$

exists, but may be infinite.

The above parameter is the so-called *intensity* of the point process. When it is finite, we can write

$$\mathbb{P}\{N((t, t+\varepsilon]) > 0\} = \lambda\varepsilon + o(\varepsilon),$$

for ε decreasing to zero.

1.5.1 Palm theory

The *Palm theory* provides a link between counting and interval properties. In this section, we will focus on the *Palm-Khinchin* equations which allow to connect the survivor function of the interevent times and the probability of having zero events in a certain time interval. This will be very useful in the last Chapter of this thesis where, as will be explained, we will discuss about the explicit form of the density of the random variable associated to the time between successive shocks, when considering a modeling of the earthquake phenomena through a new version of the ETAS model.

Let's start then by considering a stationary point process N defined on the real line and with finite intensity. For such a process, the following proposition holds.

Proposition 5. *Given a stationary point process of finite intensity λ, the limit*

$$Q_j(t) = \lim_{\varepsilon \downarrow 0} \mathbb{P}\{N((0, t]) \leq j | N((-\varepsilon, 0]) > 0\} \tag{1.12}$$

exists for $\varepsilon > 0$ and $j \in \mathbb{N}$. It is also right-continuous, non-increasing in t with $Q_j(0) = 1$.

It then follows that, for $j = 1, 2, \dots$ on $(0, \infty)$,

$$R_j(t) = 1 - Q_{j-1}(t) = \lim_{\varepsilon \downarrow 0} \mathbb{P}\{N((0, t]) \geq j | N((-\varepsilon, 0]) > 0\}$$

are distributions, provided that $\lim_{t \to \infty} R_j(t)$ is not less than one. These distributions may be interpreted in terms of the interevent times. In fact, conditioned on the occurrence of one event in zero, the number of events in $(0,t]$ is bigger than j if the time of the j^{th} event, and consequently the interevent times till $\tau_j = t_j - t_{j-i}$, has occurred before t. This means that

$$R_j(t) = \lim_{\varepsilon \downarrow 0} \mathbb{P}\Big\{ \sum_{i=1}^{j} \tau_i \geq t | t_0 = 0, t_1 > 0, N((-\varepsilon, 0]) > 0 \Big\}.$$

By induction on j, we can also prove the existence of

$$q_j(t) = \lim_{\varepsilon \downarrow 0} \mathbb{P}\{N((0,t]) = j | N((-\varepsilon, 0]) > 0\},$$

which follows directly from (1.12) for $j = 0$.

Now, if the process is also regular and recalling Proposition 5, we have that

$$\begin{aligned}
\mathcal{P}_j(t+\varepsilon) =& \sum_{i=0}^{j} \mathbb{P}\{N((0,t]) \leq (j-i) | N((-\varepsilon, 0]) = i\} \\
=& \mathbb{P}\{N((0,t]) \leq j\} - \mathbb{P}\{N((0,t]) \leq j, N((-\varepsilon, 0]) > 0\} \\
&+ \mathbb{P}\{N((0,t]) \leq (j-1), N((-\varepsilon, 0]) = 1\} + o(\varepsilon),
\end{aligned}$$

where

$$\mathcal{P}_j(t) = \mathbb{P}\{N((0,t]) \leq j\}.$$

It follows that

$$\mathcal{P}_j(t+\varepsilon) - \mathcal{P}_j(t) = -\lambda \varepsilon q_j(t) + o(\varepsilon).$$

Then, by considering the right-hand derivative operator D^+, we obtain that

$$D^+ \mathcal{P}_j(t) = -\lambda q_j(t).$$

If we consider $j = 0$ and $q_j(t)$ continuous, so that its derivative is everywhere defined, the latter formula becomes

$$\frac{d\mathcal{P}_0(t)}{dt} = -\lambda q_0(t), \tag{1.13}$$

that is exactly the equation we will use in Chapter 6 in order to get the explicit form of the interevent time density.

1.5.2 Conditional intensities

In Section 1.5, we have proposed four equivalent description of point processes on the real line. Actually, time-like evolutionary processes are naturally described by successive conditioning. This means that at each stage, we condition on the past history up to the current event considered. To be precise, let's consider a regular point process N defined on the positive half-line $\mathbb{R}_+$. The *conditional intensity function* $\lambda(t|\mathcal{H}_t)$, where $\mathcal{H}_t = \{t_i | t_i < t\}$ is the history of the events' occurrence times up to t, is the analogous of (1.11) but with the conditioning on past history:

$$\lambda(t|\mathcal{H}_t) = \lim_{\varepsilon \downarrow 0} \frac{\mathbb{P}\{N((t, t+\varepsilon]) > 0 | \mathcal{H}_t\}}{\varepsilon} \tag{1.14}$$

The conditional intensity function completely characterizes the process [Liptser and Shiryaev, 1978]. This is the characterization we are going to use for the model proposed in this thesis.

Remark 4. The conditional intensity function for a space-time-magnitude process is equivalently defined as

$$\lambda(t, x, y, m|\mathcal{H}_t) = \lim_{\varepsilon \downarrow 0} \frac{\mathbb{P}\{N((t, t+\varepsilon], (x, x+\varepsilon], (y, y+\varepsilon](m, m+\varepsilon]) > 0 | \mathcal{H}_t\}}{\varepsilon}, \tag{1.15}$$

where now $N((t, t+\varepsilon], (x, x+\varepsilon], (y, y+\varepsilon](m, m+\varepsilon])$ is the counting measure in the product space $(t, t+\varepsilon] \times (x, x+\varepsilon] \times (y, y+\varepsilon] \times (m, m+\varepsilon]$ and $\mathcal{H}_t = \{(t_i, x_i, y_i, m_i); t_i < t\}$.

Now, given a sequence $\{t_i\}_{i=0,1,\ldots}$ of events' occurrence times of a point process, we can give an alternative definition of the conditional intensities through the hazard functions. The latter are defined as

$$h_n(t|t_1, \ldots, t_{n-1}) = \frac{p_n(t|t_1, \ldots, t_{n-1})}{S_n(t|t_1, \ldots, t_{n-1})} = \frac{p_n(t|t_1, \ldots, t_{n-1})}{1 - \int_{t_n}^{t} p_n(s|t_1, \ldots, t_{n-1}) ds},$$

for a sequence $\{t_i\}$ such that $0 < t_1 < \cdots < t_n < \ldots$ and where $p_n(\cdot|\cdot)$ and $S_n(\cdot|\cdot)$ are the conditional densities and the survivor functions of the process, respectively. Then we can rewrite the conditional intensity piecewise as

$$\lambda(t|\mathcal{H}_t) = \begin{cases} h_1(t), & \text{if } 0 < t \leq t_1 \\ h_n(t|t_1, \ldots, t_{n-1}), & \text{if } t_{n-1} < t \leq t_n,\ n = 2, 3, \ldots. \end{cases}$$

The above definition may be extended to the whole line by considering $\cdots < t_{-n} < \cdots < t_1 < \cdots < t_n < \ldots$ and an infinite past history:

$$\lambda(t|\mathcal{H}_t) = h_n(t|\mathcal{H}_t) = \lim_{k \to \infty} \frac{p_n(t|t_{-k}, \ldots, t_{n-1})}{S_n(t|t_{-k}, \ldots, t_{n-1})},$$

provided that the above limit exists.

1.6 Point processes for earthquake modeling

As already explained, a seismic sequence can be mathematically described by a point process. In this field, the points of the process are typically thought to be the occurrence times of the shocks, but one can also consider marked processes to label each event with its spatial location or magnitude, too. Due to the complete characterization of these processes by the conditional intensity, strictly connected to likelihood functions, one can also perform the statistical analysis to estimate the parameters and select the model. In this section, we will present some examples of point processes widely used for earthquake sequence. In general, let's say that we can distinguish between two kinds of models: *inhibitory* and *exciting*. In the first type, the stress is gradually accumulated till a certain time, then it is strongly released; instead, the second type describes a situation in which each shock may potentially produce its own progeny and the total process consists of a cascade of events.

1.6.1 Poisson processes

The Poisson process represents an archetype of point process, really very used in many applications, among which the seismic one, where it is often used as the null model in the hypothesis test.

Let's consider a *l.c.s.c.* space E and its Borel sigma field $\mathcal{B}(E)$. A point process N on E is a *Poisson process* with a locally finite Borel measure $\Lambda(\cdot)$ when, for every finite sequence of disjoint bounded sets $\{A_i\}_{i=1,\dots,n}$, $n = 1, 2, \dots$, it holds

ii) the random variables $N(A_1), \dots, N(A_n)$, for $n = 1, 2, \dots$, are mutually independent;

i)
$$\mathbb{P}\{N(A_i) = k_i, i = 1, \dots, n\} = \prod_{i=1}^{n} \frac{[\Lambda(A_i)]^{k_i}}{k_i!} e^{-\Lambda(A_i)}, \tag{1.16}$$

where $N(A_i)$ is the generic counting measure for the number of events in A_i.

The measure $\Lambda(\cdot)$ is the *parameter measure*. When $E = \mathbb{R}$, if $\Lambda(A) = \lambda\ell(A)$, where $\ell(\cdot)$ is the Lebesgue measure, the process is said to be *homogeneous*; if instead $\Lambda(A) = \int_A \lambda(t)dt$, it is *non-homogeneous*. In both the cases, the constant or function λ is said the *intensity (mean density, mean rate)* of the process. Furthermore, in the case of a Poisson process, simple stationarity ensures stationarity. Then, in this case, we will simply use the terms "*stationary process*" when the distribution of its number of points in a given interval, depends only on the length of the latter.

The stationary homogeneous Poisson process on $\mathbb{R}$ is the typical process used for the modeling of background events; instead, non-homogeneous stationary Poisson processes on the line are the natural candidates to model the aftershocks sequences.

Let's focus now on the stationary Poisson processes defined on $\mathbb{R}$. Actually, a stationary Poisson process is sometimes referred to as a completely random process, that is a random distribution of points on the line. To be precise, we can state the following theorem.

Theorem 6. *A stationary point process, with a finite and non-zero number of point in any finite interval of the line, is a Poisson process if and only if the numbers of points in disjoint intervals are independent random variables (complete independence) and*

$$\mathbb{P}\{N((0,t]) > 1\} = o(t), \quad \textit{for } t \downarrow 0. \tag{1.17}$$

A process satisfying equation (1.17) is said *orderly*. When this property is assumed for a point process, we can give a further characterization of the Poisson process through the *void probability function* $\mathbb{P}\{N(\cdot) = 0\}$.

Theorem 7. *An orderly point process N on $\mathbb{R}$ is a stationary Poisson process if and only if*

$$\mathbb{P}\{N(A) = 0\} = e^{-\lambda \ell(A)},$$

for all sets A which can be represented as the union of a finite number of finite intervals.

The same result holds for point processes on $\mathbb{R}^d$, by considering a non-atomic measure on this space as mean measure. Furthermore, we can generalize the above results for simple point processes on a complete separable metric space E endowed with its borel sigma field $\mathcal{B}(E)$. In fact, the following theorem holds.

Theorem 8. *The finite-dimensional distribution of a simple point process N on $(E, \mathcal{B}(E))$ is characterized by its void probability function.*

Now, in the stationary Poisson process if $A = (0,t]$, the void probability $\mathbb{P}\{N((0,t]) = 0\} = e^{-\lambda t}$ may be interpreted also like the probability that the first point to the right of the origin occurs after t. This allows to introduce the *backward and forward recurrence times*, that are the length of the intervals between an arbitrary given point and the first events occurred before and after it, respectively. In general, it is easy to see that the time delay between two consecutive events in a Poisson process is exponentially distributed, and the interval containing the origin is Erlang distributed, i.e., it is the sum of two i.i.d. (independent identically distributed) exponential random variables.

Among the several extensions of a Poisson process, we will briefly present only some examples, listed below (we use the terminology in [Daley and Vere-Jones, 2003]).

Mixed Poisson process. If we set $A_i = (a_i, b_i]$ and $\Lambda(A_i) = \lambda(b_i - a_i)$ in equation (1.16), the latter can be viewed as functions of the real random variable λ. By averaging them with respect to a given distribution for λ, we get the *fidi* distribution of a new point process, that is the mixed Poisson.

Cox process. It is also called *doubly stochastic Poisson process* and is obtained by randomizing the parameter measure in a Poisson process. More precisely, given a locally finite random measure ν on $\mathbb{R}^d$, the point process N is a Cox process directed by ν if, conditional on this random measure, $N(\cdot|\nu)$ is a Poisson process on $\mathbb{R}^d$ with parameter measure ν. If the intensity is random but constant in time, the Cox process is the mixed Poisson process.

Compound Poisson process. If $Y_1, Y_2, \ldots$ are i.i.d. non-negative integer-valued random variables, independent of a Poisson process N_c with mean λ, then the compound Poisson process is defined by the counting variable

$$N((0,t]) = \sum_{i=1}^{N_c((0,t])} Y_i.$$

It is the case of a marked point process with a Poisson ground process.

1.6.2 Self-exciting Hawkes processes

The *Hawkes processes* represent a very relevant class of models widely analyzed in the literature, due to their widely use in many theoretical and applied studies. In particular, they are very important in seismology: one of the most used model in this field is the ETAS model, described in Chapter 2. It is just is a special case of the marked Hawkes process on $\mathbb{R}$, presented below in this subsection. The Hawkes processes were introduced by Hawkes in 1971 and are an example of exciting processes [Hawkes, 1971a,b; Hawkes and Oakes, 1974].

The *self-exciting Hawkes process* is a branching process constituted by two components. The first one consists of all the "immigrants", that are the points without existing ancestors; the second one contains instead all the "offsprings", that are the elements generated by previous points. The immigrants are modeled through a stationary homogeneous Poisson process with constant rate λ_{im}. Then, any point of the process, occurred at a certain time t, may produce its own progeny according to a non-homogeneous, stationary, finite Poisson process, with a mean density $\lambda_{of}(\cdot)$ that is function of the distance between the current event considered and the relative ancestor. All the finite point processes associated to the progenies are mutually independent and are independent of the immigrant component; they can also be regarded as elements of clusters whose centre are the relative ancestors. The mean rate is assumed to have total mass less than one, guaranteeing that the branching process is subcritical and of finite total size. Furthermore, since the immigrant component is a stationary Poisson process, the condition of mean cluster size finite is sufficient for the existence of the total process.

Since one can think of a cluster centre as an infected point from the outside and of the cluster elements as the elements infected by the centre point, the mean density $\lambda_{of}(dt)$ may be considered as a measure of infectivity at t, given an infected element at the origin. Now, if we can write $\lambda_{of}(dt) = \lambda_{of}(t)dt$, then we can express the conditional intensity of the Hawkes process as the function

$$\lambda(t|\mathcal{H}) = \lambda_{im} + \int_{-\infty}^{t} \lambda_{of}(t-u)N(du),$$

where the integral is made over all the elements occurred at $t_i < t$, which contribute to the risk of infection at t.

For the generalization to the multivariate case, one has to consider a point process constituted by J different types of points. The immigrants enter in the process from outside, are of type ℓ and form a Poisson process with rate $\lambda_{im}^{(\ell)}$, with $\ell = 1, \ldots, J$. Then, for every $\ell, k = 1, \ldots, J$, there is a Poisson process of elements of type k generated by an ancestor of type ℓ at time t. All these Poisson processes have conditional parameter measure $\lambda_{of}^{(lk)}(\cdot|t)$ such that $\lambda_{of}^{(lk)}(s|t) = \lambda_{of}^{(lk)}(s-t)$ and are mutually independent. Furthermore, the eigenvalue of largest modulus of the matrix $\lambda_{of}^{(lk)}(\mathbb{R})$ is less than one, guaranteeing to have an a.s. finite number of progeny for any element. The total process obtained is called the *Hawkes mutually exciting point process*. The conditional intensity for the ℓ^{th} Poisson process is

$$\lambda_\ell(t|\mathcal{H}) = \lambda_{im}^{(\ell)} + \sum_{k=1}^{J} \int_{-\infty}^{t} \lambda_{of}^{kl}(t-u)N_k(du);$$

the conditional intensity of the whole process is then obtained by superposition.

An important extension of the Hawkes process is its marked version. For simplicity, we will treat here only the case of unpredictable marks. Now, if we consider for example the self-exciting process, we can extend it to a marked point process through the interpretation of the marks κ_i as the "types" of the elements in the process in a multi-type branching model. In this case, the immigrants enter in the system as a compound Poisson process with constant rate λ_{im} and a fixed marked distribution $Z(\cdot)$; then, each element may produce its own offspring according to a Poisson process with mark rate $\lambda_{of}(\cdot|\kappa)$, depending only on the mark of the ancestor and on the distance between the ancestor and the offspring considered. A typical case is the one in which the above mark rate is magnitude separable, that is $\lambda_{of}(\cdot|\kappa) = \lambda_{of}(\cdot)\Psi(\kappa)$. Finally, the marks relative to the offsprings have the same distribution $Z(\cdot)$ of the immigrants and are i.i.d. random variables. In the case of a magnitude separable process, the imposition of having i.i.d. marks implies that the ground process of the MPP is an ordinary Hawkes process with immigration rate λ_{im} and infectivity measure $\lambda_{of}(dt) \cdot \mathbb{E}[\Psi(\kappa)]$. The condition $\mathbb{E}[\Psi(\kappa)] \int_0^\infty \lambda_{of}(t)dt < 1$ guarantees also to have an a.s. finite total number of progeny.

1.6.3 Stress-release models

The stress-release models are based on the elastic rebound theory, proposed for the first time by the seismologist Harry Fielding Reid in his study of the 1906 San Francisco earthquake. They belong to the more general class of *self-correcting processes*, and represent an example of inhibitory type models [Ogata and Vere-Jones, 1984; Vere-Jones, 1978; Vere-Jones and Ogata, 1984]. The conditional intensity is in this case governed by a Markov process, generally partially observed, that is the stress level of the region under consideration. More precisely, the probability of occurrence increases with the stress level and abruptly decreases after that an event has occurred. If we indicate with $S(t)$ the stress level, that is an unobserved jump-type Markov chain, we can write

$$\lambda(t|\mathcal{H}) = \psi(S(t)),$$

where $\mathcal{H}_t = \{(t_i, m_i); t_i < t\}$ is the past history, $\psi(\cdot)$ is an increasing function and

$$S(t) = S(0) + at - \sum_{\{i|t_i<t\}} 10^{1.5(m_i-m_0)},$$

that is the level of stress accumulated till time t. Starting from an initial unknown stress level $S(0) > 0$, the above expression for $S(t)$ says that the stress itself linearly increases with time, with an unknown rate $a > 0$, but it depends also on the stress accumulated till t by considering all the previous shocks occurred at times t_i and with magnitudes m_i. It follows that the conditional intensity is fully determined by the initial value $S(0)$, the parameters of the model and the observations (t_i, m_i).

A typical form of the above function $\psi(\cdot)$ is

$$\psi(t) = e^{a+bt},$$

which characterizes the *stress-release model*. Its multivariate version is the so-called *linked stress-release*, consisting of the model that considers the interaction of stresses among a finite number of different regions. If we label these components with the index $i = 1, 2, \ldots, I$ and we consider the stress level of the generic i^{th} region $S_i(t)$, then we can write, for each i,

$$S_i(t) = S_i(0) + a_i t - \sum_{j=1}^{I} \theta_{ij} X_j(t),$$

where $S_i(0)$ is the initial stress for the i^{th} region, a_i is its rate, θ_{ij} is its proportion of stress drop transferred to the j^{th} region and

$$X_j(t) = \sum_{\{i|t_i<t, r_i=j\}} 10^{1.5(m_i-m_0)},$$

where r_i is the region where the i^{th} events has occurred.

Chapter 2

The Epidemic Type Aftershock Sequence model

The *Epidemic Type Aftershock Sequence* (ETAS) is a well-known model in statistical seismology, widely used by the scientists to analyze the earthquake phenomenon from a probabilistic point of view [Ogata, 1988, 1989, 1998, 1999]. As already specified, it belongs to the more general class of the self-exciting Hawkes processes, presented in the previous chapter. More precisely, it is a linear marked self-exciting Hawkes process. The ETAS model is based on a specific branching process in which each event, belonging to any given generation, may produce its own offspring independently of the other shocks. The first version proposed by Ogata was pure temporal, but some years later he improved the model by considering the spatial locations, too. This chapter is focused on the derivation and the definition of the ETAS model; some of its properties and connected studies are also discussed.

2.1 Derivation and definition of the model

The ETAS model derives from the *Omori law*, which describes the time decay of the aftershocks in a sequence:

$$\Phi_{first}(t) = \frac{\mathrm{K}}{(t+c)}, \quad t > 0,$$

where K and c are constants and t is the elapsed time from the triggering event occurred in $t = 0$. This law has been found by Omori to fit well the aftershocks of the 1981 Nobi earthquake of magnitude 8 [Omori, 1894]. Nevertheless, several years later, Utsu [Utsu, 1957] found that the fit for the decay of the first order aftershocks was better expressed by

$$\Phi_{first}(t) = \frac{\mathrm{K}}{(t+c)^p}, \quad t > 0.$$

This is the so-called *Omori-Utsu law* or *modified Omori law.* The value of the parameter p has been estimated for more that 200 aftershocks sequences, ranging from about 0.6 to 2.5, with a median value at 1.1.

Actually, the above law does not model in a good way the cases in which the strong aftershocks themselves produce their own daughters. That is, the sequence has secondary aftershocks, too. In fact, as already specified, the Omori-Utsu law describes the decay only of the first generation aftershocks. For this reason, Utsu and Ogata [Ogata, 1983; Utsu, 1970] proposed to use the superposition of Omori-Utsu law:

$$\bar{\Phi}(t) = \sum_{i=0}^{N_{aft}} \mathbb{1}_{[0,t)}(t_i) \frac{\mathrm{K}_i}{(t - t_i + c_i)^{p_i}}, \quad t > 0,$$

where: $\mathbb{1}_A(x)$ is the indicator function having value one when $x \in A$, zero otherwise; t_0 is the occurrence time of the first triggering event; N_{aft} is the number of its strong triggering aftershocks, occurred at times $(t_1, \cdots, t_{N_{aft}})$; K_i, c_i and p_i, for $i = 0, 1, \cdots, N_{aft}$, are constants.

The next step was made by Ogata [Ogata, 1988, 1998], who understood that not only the first order aftershocks, but all the events in a sequence may produce their own offsprings. Then, he introduced self-similarity in the modeling of earthquakes and proposed a weighted superposition of the modified Omori functions. More precisely, he modeled the aftershock activity through a non-stationary Poisson process, in which the occurrence rate of the events generated by the generic i^{th} triggering shock, occurred at t_i and with magnitude m_i, was given by

$$\Phi_i(t) = \frac{\kappa e^{a(m_i - m_0)}}{(t - t_i + c)^p}, \quad t > 0,$$

where the parameters (κ, a, c, p) are common to all the shocks in the sequence and m_0 is the reference magnitude. Each weight is a function of the magnitude of the triggering event considered. It is the so-called *productivity law*, indicated from now on as

$$\varrho(m') = \kappa 10^{\alpha(m'-m_0)} = \kappa e^{a(m'-m_0)}, \quad m \geq m_0, \tag{2.1}$$

where $a = \alpha \ln 10$ represents the contribution of first generation shocks triggered by an event with generic magnitude m'. This law tells that the number of the aftershocks generated by an event depends on its magnitude. Assuming also that this number is proportional to the area S in which the aftershocks occur, one can deduce the proportionality between area and magnitude. This dependence is expressed in the following formula, proposed by Utsu and Seki [Utsu and Seki, 1954]:

$$\log S = 1.02m' + const,$$

where m' is the magnitude of the generic triggering event considered.

If, on one side, it is supposed that each aftershock gives birth to a non-homogeneous, stationary Poisson process, independently of the other aftershocks, on the other side, the background component is assumed to be modeled by a homogeneous Poisson process, again independent of the other processes, with a rate constant in time. It will be indicated here with ϖ. Actually, some recent works have shown that the background rate may depend on some variables, like for example the time [Lombardi et al., 2010], but for simplicity we will not consider this case.

The spatial component has been included in the ETAS model by Ogata in 1998 [Ogata, 1998]. In this case, the background rate is considered dependent on the spatial locations. More precisely, the spatial component relative to the aftershocks is a function of the mother event's magnitude m' and of the difference $(x - x', y - y')$, between the spatial locations of the triggered and the triggering events, respectively. Many examples has been proposed to describe the latter function. In general, they can all be rewritten in the common standard form

$$f(x - x', y - y'|m') = c(m')f\left(\frac{(x - x')^2 + (y - y')^2}{g(m')}\right),$$

where $g(\cdot)$ is a certain magnitude function and $c(m')$ is the normalization constant. However, we can say that the two most used forms are the following two:

- the Gaussian form with an exponential $g(\cdot)$, as for example

$$\frac{1}{2\pi d e^{a(m'-m_0)}} \exp\left\{-\frac{1}{2}\frac{(x - x')^2 + (y - y')^2}{d e^{a(m'-m_0)}}\right\},$$

  where (x', y') and $\sqrt{d e^{a(m'-m_0)}}$ are the location and scale parameters, respectively (see [Zhuang et al., 2002]). Let's notice that a is the usual parameter of the productivity law, instead d is a parameter associated only to the spatial function;

- the power law form, as for example

$$\frac{c_{dq}}{\left\{(x-x')^2+(y-y')^2+[de^{\bar{\gamma}(m'-m_0)}]^2\right\}^q},$$

where c_{dq} is the normalization constant and $(d, q, \bar{\gamma})$ are the spatial parameters (see [Marzocchi and Lombardi, 2009]). Let's notice that in the latter case, the exponential part relative to the magnitudes has a parameter different from the one in the productivity law.

The last ingredient is given by the law describing the magnitudes. In the ETAS, they are modeled with an exponential distribution, both in the case of background and triggered components. More precisely, the experimental law modeling the magnitudes of all the events in a catalog is the well-known *Gutenberg-Richter law*:

$$p(m') = b \ln 10 \cdot 10^{-b(m'-m_0)} = \beta e^{-\beta(m'-m_0)}, \quad m' \geq m_0, \tag{2.2}$$

with $\beta = b \ln 10$ and b is the so-called *b-value*. The latter is a measure of the relative frequency of small and large earthquakes and is typically estimated near 1. In a log scale, the b-value is the slope of the decreasing line representing the Gutenberg-Richter law. Then, the lower is this value, the higher is the probability of having events with higher magnitudes. Let's notice that, according to the above law, the magnitudes are independent between themselves and of the characteristics of past seismicity [Gutenberg and Richter, 1944].

Now, recalling that a point process is completely characterized by its conditional intensity, as explained in Chapter 1 (see Subsection 1.5.2), we are ready to define the general **space-time-magnitude ETAS model**. It is characterized by the following intensity, based on the history of occurrence $\mathcal{H}_t = \{(t_i, x_i, y_i, m_i); t_i < t\}$:

$$\lambda(t, x, y, m|\mathcal{H}_t) = \varpi(t, x, y, m) + \sum_{\{i|t_i<t\}} h(t, x, y, m; t_i, x_i, y_i, m_i), \tag{2.3}$$

where:

- $\varpi(t, x, y, m)$ is the intensity function of the Poisson process modeling the background component of the sequence. In the case of the space-time-magnitude ETAS model it supposed to be homogeneous in time and magnitude separable, where the law for the magnitudes of this component is the Gutenberg-Richter one. This means that we can write $\varpi(t, x, y, m) = \varpi(x, y)p(m)$.

- $h(t, x, y, m; t_i, x_i, y_i, m_i)$ is the response function from the i^{th} triggered event when $t_i < t$. More precisely, once occurred the i^{th} triggered event (t_i, x_i, y_i, m_i), it is the intensity function of the Poisson process modeling its own aftershocks. This function is stationary and separable in space and time. Let's recall that here stationarity means that the spatial and the temporal functions depend only on the difference between the current event and its mother. Then, we can write

$$h(t, x, y, m; t_i, x_i, y_i, m_i) = \varrho(m_i)\Phi(t - t_i)f(x - x_i, y - y_i|m_i)p(m|m_i), \tag{2.4}$$

where

$$\Phi(t) = \frac{p-1}{c}\left(1 + \frac{t}{c}\right)^{-p}, \quad t > 0, \tag{2.5}$$

is the parametrization of the Omori-Utsu law in order to be a probability density function for the aftershocks decay. This parametrization has already been used for the study of supercritical and subcritical regimes in the ETAS model [Helmstetter and Sornette, 2002a; Saichev and Sornette, 2007; Sornette and Sornette, 1999]. Furthermore, $\varrho(m)$ is the productivity law (equation (2.1)) and the functions $f(x, y|m)$ and $p(m|m_i)$ are the probability densities of spatial locations and magnitudes.

The **space-time ETAS model** is derived from above by considering the further hypothesis of independent magnitudes. Then, we can write the magnitude transition probability density function as $p(m|m_i) = p(m)$, where $p(m)$ is the Gutenberg-Richter law (equation (2.2)), and the above conditional intensity function (2.3), based on the history of occurrence $\mathcal{H}_t = \{(t_i, x_i, y_i, m_i); t_i < t\}$ (see also [Zhuang et al., 2002]), as

$$\begin{aligned}\lambda(t, x, y, m|\mathcal{H}_t) =& p(m)\Bigg[\varpi(x, y) + \sum_{\{i|t_i<t\}} \kappa e^{a(m_i - m_0)} \\ &\cdot \frac{p-1}{c}\left(1 + \frac{t - t_i}{c}\right)^{-p} f(x - x_i, y - y_i|m_i)\Bigg].\end{aligned} \tag{2.6}$$

This is the conditional intensity function characterizing the space-time ETAS model. Since the magnitudes are independent and the law $p(m)$ is the same both for background and triggered events, very often in the literature the Gutenberg-Richter law does not appear in the previous formula.

Finally, if the background rate is constant and the function $f(x - x_i, y - y_i|m_i)$ is not considered, one obtains the conditional intensity characterizing the **pure temporal ETAS model**:

$$\lambda(t, m|\mathcal{H}_t) = p(m)\left[\varpi + \sum_{\{i|t_i<t\}} \kappa e^{a(m_i-m_0)}\frac{p-1}{c}\left(1+\frac{t-t_i}{c}\right)^{-p}\right], \tag{2.7}$$

where now $\mathcal{H}_t = \{(t_i, m_i); t_i < t\}$. Again, very often in the literature the Gutenberg-Richter law does not appear in the previous formula and the conditional intensity for the temporal ETAS is indicated with

$$\lambda(t|\mathcal{H}_t) = \frac{\lambda(t, m|\mathcal{H}_t)}{p(m)}. \tag{2.8}$$

The new model we are going to propose in this thesis is a variation of the one characterized by the intensity in (2.7). As we are going to see in Chapter 4, the variation consists in the fact that the Gutenberg-Richter law is supposed to be valid only for the background events, while the magnitude probability density function in the case of the triggered events is a transition probability function depending on the magnitude of the mother events. Then, the new conditional intensity coincides with the general conditional intensity $\lambda(t, x, y, m|\mathcal{H}_t)$ in (2.3), when discarding the spatial locations, that is

$$\lambda(t, m|\mathcal{H}_t) = \varpi p(m) + \sum_{\{i|t_i<t\}} \varrho(m_i)\Phi(t-t_i)p(m|m_i), \tag{2.9}$$

with a particular choice of the transition function $p(m|m_i)$ (see (4.2)).

A similar variation of the temporal ETAS model is the *self-similar ETAS model* [Saichev and Sornette, 2005; Vere-Jones, 2005], which is characterized by a conditional intensity depending on the difference between the magnitude of the triggered event m', and the magnitude of the relative triggering shock m, i.e.,

$$\varrho(m')p(m|m') = \lambda(m-m').$$

The model we propose in Chapter 4 does not fall into the class of *self-similar ETAS models*, and has the fundamental characteristics that the Gutenberg-Richter law remains invariant with respect to the transition probability density considered for the triggered events' magnitude. This condition is instead not valid for the above self-similar ETAS model.

As we will see later (see Subsection 2.3.1), the parameters defining an ETAS model are typically estimated using maximum-likelihood algorithms.

2.2 Properties

This section is devoted to the properties of the general space-time-magnitude ETAS model, defined by the conditional intensity (2.3). In what follows, we are going to discuss the criticality issue and the moment properties.

2.2.1 Criticality

The concept of point process stability concerns the convergence to equilibrium: the problem is to see if, starting from a given process, not necessarily stationary, there exists a stationary point process to which the initial process converges. The condition for stability is fundamental in data analysis in order to assess the reasonability of estimated model parameters. More precisely, it refers to a situation in which the process is non-explosive. Several works exists in the literature studying the stability several types of Hawkes processes (see for example [Brémaud, 1996; Brémaud et al., 2002]), but we will focus here on the specific case of the ETAS model. In the ETAS model, this condition is strictly connected to the sub-criticality of the process. In fact, in terms of seismic predictions, the use of parameters specifying a supercritical process would lead to the overestimation of the earthquake risk in the medium and long-term [Zhuang and Ogata, 2006; Zhuang et al., 2012].

To the aim of finding the explicit conditions for the stability of the process, we have to study an eigenvalue problem [Saichev and Sornette, 2005; Zhuang, 2002]. Since the spontaneous sources are modeled through a stationary Poisson process with i.i.d. marks, given by the magnitudes, the number of the background shocks is finite in a given bounded time interval. Then, in order to guarantee the non-explosion of the process, we have to check that the average number of events generated by a triggering shock with a given magnitude m is finite and then the relative process dies out in a finite time.

Let's start then by considering the average number $\mathbb{E}[N_{\bar{v}}]$ of events with magnitude above the reference threshold m_0, triggered by a spontaneous shock occurred in $\bar{v'} = (t', x', y', m')$. It is easy to see that

$$\mathbb{E}[N_{\bar{v}'}] = \int_{\mathfrak{S}} h(\bar{v};\bar{v}')\mathbb{1}_{(t',+\infty)}(t)\mathbb{E}[N_{\bar{v}}]d\bar{v} + \int_{\mathfrak{S}} h(\bar{v};\bar{v}')\mathbb{1}_{(t',+\infty)}(t)d\bar{v}, \qquad (2.10)$$

where $h(\bar{v};\bar{v}')$ is the intensity (2.4) of the process of events triggered by a the shock identified by $\bar{v}'$ and

$$\mathfrak{S} = \mathbb{R}^2 \times (-\infty,+\infty) \times [m_0,\infty).$$

The presence of the indicator function in the previous integrals is due to the fact that the intensity $h(\bar{v};\bar{v}')$ is based on the history of the event in $\bar{v}$ up to its occurrence time t.

One can interpret the above integral equation (2.10) by considering the following. If $V_0(\bar{v}) = \varpi(\bar{v})$ is the intensity function of the background events' generation Z_0, then the intensity relative to the first generation Z_1 is given by

$$V_1(\bar{v}) = \int_{\mathfrak{S}} h(\bar{v};\bar{v}')\mathbb{1}_{(-\infty,t)}(t')V_0(\bar{v}')d\bar{v}'.$$

It easily follows that the intensity of the j^{th} generation Z_j is

$$V_j(\bar{v}) = \int_{\mathfrak{S}} h(\bar{v};\bar{v}')\mathbb{1}_{(-\infty,t)}(t')V_{j-1}(\bar{v}')d\bar{v}' = \int_{\mathfrak{S}} h^{[j]}(\bar{v};\bar{v}')\mathbb{1}_{(-\infty,t)}(t')V_0(\bar{v}')d\bar{v}', \tag{2.11}$$

where

$$h^{[1]}(\bar{v};\bar{v}') = h(\bar{v};\bar{v}') \quad \text{and} \quad h^{[j]}(\bar{v};\bar{v}') = \int_{\mathfrak{S}} h^{[j-1]}(\bar{v};\bar{v}'')h(\bar{v}'';\bar{v}')\mathbb{1}_{(t',t)}(t'')d\bar{v}''. \tag{2.12}$$

Now, in order to search for the stationary solutions of equation (2.10), let's consider $\ell_1(\bar{v}')$ and $\ell_2(\bar{v})$ the left and right eigenfunctions of $h(\bar{v};\bar{v}')$, respectively, corresponding to the maximum eigenvalue η:

$$\eta\ell_1(\bar{v}') = \int_{\mathfrak{S}} \ell_1(\bar{v})h(\bar{v};\bar{v}')\mathbb{1}_{(t',+\infty)}(t)d\bar{v} \tag{2.13}$$

and

$$\eta\ell_2(\bar{v}) = \int_{\mathfrak{S}} h(\bar{v};\bar{v}')\mathbb{1}_{(-\infty,t)}(t')\ell_2(\bar{v}')d\bar{v}', \tag{2.14}$$

satisfying

$$\int_{\mathfrak{S}} \ell_1(\bar{v})\ell_2(\bar{v})d\bar{v} = 1.$$

Let's notice that $\Lambda(\bar{v};\bar{v}') = \ell_1(\bar{v}')\ell_2(\bar{v})$ is the projection operator of the intensity function corresponding to η, that is

$$\begin{aligned}\int_{\mathfrak{S}} \Lambda(\bar{v};\bar{v}'')h(\bar{v}'';\bar{v}')\mathbb{1}_{(t',+\infty)}(t'')d\bar{v}'' &= \int_{\mathfrak{S}} h(\bar{v};\bar{v}'')\mathbb{1}_{(-\infty,t)}(t'')\Lambda(\bar{v}'';\bar{v}')d\bar{v}'' \\ &= \eta\Lambda(\bar{v};\bar{v}').\end{aligned}$$

Then, we have that

$$\lim_{j\to\infty} \frac{h^{[j]}(\bar{v};\bar{v}')}{\eta^j} = \Lambda(\bar{v};\bar{v}').$$

Consequently, it holds

$$\lim_{j\to\infty} \frac{V_j(\bar{v})}{\eta^j} = \int_{\mathfrak{S}} \Lambda(\bar{v};\bar{v}')V_0(\bar{v}')d\bar{v}', \tag{2.15}$$

from which one has that:

- if $\eta < 1$, then $\lim_{j\to\infty} V_j(\bar{v}) = 0$,
- if $\eta = 1$, then $\lim_{j\to\infty} V_j(\bar{v}) = const$,
- if instead $\eta > 1$, then $\lim_{j\to\infty} V_j(\bar{v}) = \infty$.

We can deduce that η is the *critical parameter*: when it is less than one the process is stable, otherwise it is explosive. By looking at the result of the limit in (2.15), which can be rewritten as

$$\eta^j \ell_2(\bar{v}) \int_{\mathfrak{S}} \ell_1(\bar{v}')V_0(\bar{v}')d\bar{v}' = \eta^j \ell_2(\bar{v}) \cdot \text{constant},$$

it follows that the eigenfunction $\ell_2(\bar{v})$ is proportional to the asymptotic intensity of the population at the j^{th} generation, when j tends to infinity. On the other side, since it holds

$$\begin{aligned}
\lim_{j\to\infty} \sum_{k=j}^{\infty} \int_{\mathfrak{S}} h^{[k]}(\bar{v};\bar{v}')d\bar{v} &= \lim_{j\to\infty} \sum_{k=j}^{\infty} \eta^k \int_{\mathfrak{S}} \Lambda(\bar{v};\bar{v}')d\bar{v} \\
&= \lim_{j\to\infty} \frac{\eta^j}{1-\eta} \ell_1(\bar{v}') \int_{\mathfrak{S}} \ell_2(\bar{v})d\bar{v} \\
&= \lim_{j\to\infty} \frac{\eta^j}{1-\eta} \ell_1(\bar{v}') \cdot \text{constant},
\end{aligned}$$

one can deduce that the eigenfunction $\ell_1(\bar{v}')$ can be interpreted as the asymptotic ability of a triggering event, identified by $\bar{v}'$, in producing its directly and indirectly aftershocks.

Now, recalling that (2.4) is the conditional intensity function of the general space-time-magnitude ETAS model, the two equations (2.13) and (2.14) become

$$\begin{aligned}
\eta \ell_1(t',x',y',m') &= \int_{\mathfrak{S}} \ell_1(t,x,y,m)\varrho(m')\Phi(t-t') \\
&\quad \cdot f(x-x',y-y'|m')p(m|m')dtdxdydm, \\
\eta \ell_2(t,x,y,m) &= \int_{\mathfrak{S}} \varrho(m')\Phi(t-t')f(x-x',y-y'|m') \\
&\quad \cdot p(m|m')\ell_2(t',x',y',m')dt'dx'dy'dm'.
\end{aligned}$$

By integrating the above equations with respect to space and time, considering the eigenfunctions as functions only of the magnitudes and recalling that $p(m|m') = p(m)$, we have

$$\eta \ell_1(m') = \varrho(m') \int_{\mathfrak{M}} \ell_1(m)p(m)dm \tag{2.16}$$

and

$$\eta\ell_2(m) = p(m)\int_{\mathfrak{M}} \varrho(m')\ell_2(m')dm'. \tag{2.17}$$

Let's notice that the condition for the temporal integral to be finite is $p > 1$, as one can easily verify. Now, from (2.16) and (2.17) it follows that

$$\ell_1(m') = A_1\varrho(m')$$

and

$$\ell_2(m) = A_2 p(m),$$

where A_1 and A_2 are constants. By substituting the latter equalities again in (2.16) and (2.17) and considering the reference magnitude threshold m_0, we get the critical parameter as

$$\begin{aligned}
\eta &= \int_{m_0}^{\infty} p(m')\varrho(m')dm' \\
&= \int_{m_0}^{\infty} \beta e^{-\beta(m'-m_0)}\kappa e^{a(m'-m_0)}dm' \\
&= \beta\kappa\int_{m_0}^{\infty} e^{-(\beta-a)(m'-m_0)}dm' \\
&= \frac{\beta\kappa}{\beta - a},
\end{aligned} \tag{2.18}$$

with $\beta > a$. Equation (2.18) has been obtained using formulae (2.1) and (2.2). Let's notice that the condition $\eta < 1$ implies that $p > 1$ and $\beta > a$. Concluding, the conditions for the non-explosion of the process are

$$p > 1, \qquad \beta > a, \qquad \kappa < \frac{\beta - a}{\beta}. \tag{2.19}$$

As we are going to see in Chapter 4, the new version of the ETAS model we will propose has the same conditions for the non-explosion as in the classical case treated here.

2.2.2 First and second order moments

In Chapter 1, we have defined the first moment measure (or mean measure) and the second order measure of a Poisson process $N(\cdot)$ as

$$\begin{aligned}
M_1(\bar{v})d\bar{v} &= \mathbb{E}\big[N(dt \times dx \times dy \times dm)\big] \quad \text{and} \\
M_2(\bar{v};\bar{v}')d\bar{v}d\bar{v}' &= \mathbb{E}\big[N(dt \times dx \times dy \times dm)N(dt' \times dx' \times dy' \times dm')\big].
\end{aligned}$$

Now, recalling that in the case of the ETAS model the number $N(\cdot)$ of events in a certain space-time-magnitude interval $(dt \times dx \times dy \times dm)$ is a Poisson random variable, with rate $\lambda(t,x,y,m|\mathcal{H}_t)$ defined in (2.3), it follows that

$$\mathbb{E}\big[N(dt \times dx \times dy \times dm)|\mathcal{H}_t\big] = \lambda(\bar{v}|\mathcal{H}_t)d\bar{v},$$

where we have set again $\bar{v} = (t,x,y,m)$. By using the properties of the conditional expected value, equation (2.3) and the Campbell's theorem (see 1.3 in Chapter 1), we consequently obtain that the first moment measure becomes

$$\begin{aligned} M_1(\bar{v}) &= \frac{1}{d\bar{v}}\mathbb{E}\Big[\mathbb{E}\big[N(dt \times dx \times dy \times dm)|\mathcal{H}_t\big]\Big] \\ &= \mathbb{E}\big[\lambda(\bar{v}|\mathcal{H}_t)\big] \\ &= \mathbb{E}\Bigg[\varpi(\bar{v}) + \sum_{\{i|t_i<t\}} h(\bar{v};\bar{v}_i)\Bigg] \\ &= \varpi(\bar{v}) + \int_{\mathfrak{S}} h(\bar{v};\bar{v}'')\mathbb{1}_{(-\infty,t)}(t'')M_1(\bar{v}'')d\bar{v}'', \end{aligned} \tag{2.20}$$

where we recall that $\mathfrak{S}$ is the space-time-magnitude domain defined in (2.2.1) and $\bar{v}'' = (t'',x'',y'',m'')$. Instead, for the second order measure we get

$$\begin{aligned} M_2(\bar{v};\bar{v}') =& \frac{1}{d\bar{v}d\bar{v}'}\mathbb{E}\Big[\mathbb{E}\big[N(dt \times dx \times dy \times dm)N(dt' \times dx' \times dy' \times dm')|\mathcal{H}_{t'}\big]\Big] \\ =& \frac{1}{d\bar{v}d\bar{v}'}\mathbb{E}\Big[N(dt \times dx \times dy \times dm)\mathbb{E}\big[N(dt' \times dx' \times dy' \times dm')|\mathcal{H}_{t'}\big]\Big] \\ =& \frac{1}{d\bar{v}d\bar{v}'}\mathbb{E}\big[N(dt \times dx \times dy \times dm)\lambda(\bar{v}'|\mathcal{H}_{t'})d\bar{v}'\big] \\ =& \frac{1}{d\bar{v}d\bar{v}'}\mathbb{E}\Bigg[N(dt \times dx \times dy \times dm)\bigg(\varpi(\bar{v}') + \sum_{\{i|t_i'<t'\}} h(\bar{v}';\bar{v}_i')\bigg)d\bar{v}'\Bigg] \\ =& \varpi(\bar{v}')M_1(\bar{v}) + \frac{1}{d\bar{v}d\bar{v}'}\mathbb{E}\Bigg[N(dt \times dx \times dy \times dm)\Bigg]d\bar{v}' \\ & \cdot \int_{\mathfrak{S}} h(\bar{v}';\bar{v}'')\mathbb{1}_{(-\infty,t')}(t'')N(dt'' \times dx'' \times dy'' \times dm'') \\ =& \varpi(\bar{v}')M_1(\bar{v}) + h(\bar{v}';\bar{v})M_1(\bar{v}) \\ & + \int_{\mathfrak{S}\cap(\bar{v}''\neq\bar{v})} h(\bar{v}';\bar{v}'')\mathbb{1}_{(-\infty,t')}(t'')M_2(\bar{v}'';\bar{v})d\bar{v}''. \end{aligned} \tag{2.21}$$

Before concluding this subsection, it is interesting to look at the formula (2.20), concerning the mean measure, in the case of a time independent background rate and a process crude stationary in time. By the Neumann series expansion [Taylor and Lay, 1980], equation (2.20) becomes

$$M_1(\bar{v}) = \varpi(x,y,m) + \sum_{\ell=1}^{\infty} \int_{\mathfrak{S}} h^{[\ell]}(t-t'',x,y,m;x'',y'',m'')\varpi(x'',y'',m'')d\bar{v}'',$$

where

$$h^{[1]}(t-t'',x,y,m;x'',y'',m'') = h(t-t'',x,y,m;x'',y'',m'')$$

and

$$\begin{aligned} h^{[\ell]}(t-t'',x,y,m;x'',y'',m'') = \int_{\mathfrak{S}} & h^{[\ell-1]}(t-t''',x,y,m;x''',y''',m''') \\ & \cdot h(t'''-t'',x''',y''',m''';x'',y'',m'')d\bar{v}'''. \end{aligned}$$

Since there is no dependence on t, we can write $M_1(\bar{v}) = M_1(x,y,m)$. Let's notice that the above two formulas are the same in (2.12), with the exception of the temporal difference in the first argument of $h(\cdot;\cdot)$.

Finally, if we furthermore consider a magnitude separable process and recall equation (2.4) for $h(t,x,y,m;t_i,x_i,y_i,m_i)$, we obtain that equation (2.20) for the mean measure becomes

$$\begin{aligned} M_1(x,y,m) =& \varpi(x,y)p(m) + \int_{\mathfrak{S}} \varrho(m'')\Phi(t-t'')f(x-x'',y-y''|m'') \\ & \cdot p(m|m'')M_1(x'',y'',m'')d\bar{v}''. \end{aligned} \tag{2.22}$$

By taking the expectations on both sides with respect to space and time, we get the relationship between the total magnitude distribution $j_{tot}(\cdot)$, the distribution of the background events' magnitude $p(\cdot)$ and the one of the aftershocks' magnitude $p(\cdot|\cdot)$:

$$j_{tot}(m) = \frac{\overline{\varpi}p(m)}{\overline{m}} + \int_{\mathfrak{M}} \varrho(m'')p(m|m'')j_{tot}(m'')dm'', \tag{2.23}$$

where

$$\overline{\varpi} = \int_{\mathbb{R}^2} \varpi(x,y)dxdy \quad \text{and} \quad \overline{m} = \int_{\mathfrak{M}\times\mathbb{R}^2} M_1(x,y,m)dxdydm.$$

By integrating again with respect to the magnitude, we get

$$\overline{m} = \overline{\varpi} + \overline{m}\int_{\mathfrak{M}} \varrho(m'')j_{tot}(m'')dm''.$$

In the case of the space-time ETAS model, the total magnitude distribution is the Gutenberg-Richter law and the minimum magnitude is the reference one, indicated with m_0. We then obtain

$$\int_{m_0}^{\infty} \varrho(m')p(m')dm' = 1 - \frac{\overline{\varpi}}{\overline{m}}. \tag{2.24}$$

The right member of the above equation is called the *branching ratio*. By definition, it is the average number of first generation triggered events per triggering shock, but can be defined also as the proportion of triggered events with respect to all the shocks.

Remark 5. In the case of the ETAS model, the branching ratio coincides with the critical parameter η, as one can easily deduce by looking at the first equality in (2.18).

2.3 Catalog simulation, parameters estimation and residual analysis

In this section, we will discuss about the theoretical bases and the algorithms concerning parameter estimation, simulation and residual analysis of the ETAS model. They represent very useful tools for the data analysis of earthquake sequences. All the algorithms we are going to present are specific for the temporal ETAS model or the more general space-time ETAS. The most of them have been firstly proposed by Yoshihiko Ogata between the '80s and the '90s; later, they have been refined, completed and improved. As specified later, the Ogata's codes belong to the "STATSEI" software [Ogata, 2006], available online at `http://www.ism.ac.jp/~ogata/Ssg/ssg_softwaresE.html`.

Before illustrating the algorithms and their underlying theories, it is important to define the so-called *learning (or precursory) period*. Due to the long-living nature of the aftershock activity, it can be supposed that the seismicity of a given period may be influenced by some shocks occurred in a certain temporal interval preceding the above given period. In order to take into account this effect, in the data analysis one may consider a target interval, where performing the analysis itself, and a precursory period, which includes the events affecting the seismicity occurred in the target interval. The results of the analysis will obviously concern only the shocks belonging to the latter interval. The choice of the length of the learning period is not easy, due to the absence of a given law which can help in this sense. It may depend for example on the number of events in the catalog, or on the proportion of cluster events with respect to the total shocks. However, the precursory period influences only the analysis concerning the triggered component and then, for example, the choice of a long learning interval improves the estimation of the parameters concerning the triggered component, but makes worse the estimation of the background rate.

2.3.1 Parameter estimation

The most used method for the estimation of the parameters in the ETAS model is based on the maximum-likelihood technique. This is true both for the temporal and the spatiotemporal settings. However, this section focuses only on the temporal case. In fact, since the estimation in the spatiotemporal case is a very complex issue, the relative algorithms proposed are till now not so efficient.

The Ogata's FORTRAN estimation program for the parameters of the temporal ETAS, available online at `http://www.ism.ac.jp/~ogata/Ssg/ssg_softwaresE.html`, is [**etas.f**]. It allows to choose between two versions: an approximated one and an exact one. The difference is in the elapsed time when processing relative bigger catalogs: in the first version it is proportional to the number of events N, in the second one it is proportional to N^2. In input the program requires obviously the catalog for which estimating the ETAS parameters. Furthermore, one has to specify the starting times for the learning and the target periods, the thresholds magnitudes and the initial set of parameters. Starting from the latter, for each step of the parameter vector, the program implements repeated computations of the likelihood function and of its gradient. The squared sums of the gradients are also calculated. The likelihood is decreasing and converges to a finite value; the iteration procedure stops when the squared sum of the gradients approaches zero.

Given the total number of events N_{tot} in the precursory and target periods, and the times t_{in}, t_{end}, respectively corresponding to the initial and ending times of the target period in which the estimation is computed, the likelihood function $L_{t_{in},t_{end}}(\zeta)$ for the estimation of the vector of parameters ζ is:

$$L_{t_{in},t_{end}}(\zeta) = \prod_{i=i_{in}}^{N_{tot}} \{\lambda(t_i|\mathcal{H}_{t_i})\} e^{-\int_{t_{in}}^{t_{end}} \lambda(t|\mathcal{H}_t)dt},$$

where $\lambda(t|\mathcal{H}_t)$ is the intensity of the temporal ETAS model, given by equation (2.8), based on the history $\mathcal{H}_t$. Furthermore, i_{in} is the index of the first event occurred in the target period. Now, by computing the logarithm of the likelihood function for simplicity of computation, we obtain the formula

$$
\begin{aligned}
&\ln\left[L_{t_{in},t_{end}}(\zeta)\right]\\
&=\sum_{i=i_{in}}^{N_{tot}} \ln\left[\lambda(t_i|\mathcal{H}_{t_i})\right] - \int_{t_{in}}^{t_{end}} \left[\varpi + \sum_{\{i|t_i<t\}} \kappa e^{a(m_i-m_0)} \frac{p-1}{c}\left(1+\frac{t-t_i}{c}\right)^{-p}\right] dt\\
&=\sum_{i=i_{in}}^{N_{tot}} \ln\left[\lambda(t_i|\mathcal{H}_{t_i})\right] - \varpi(t_{end}-t_{in}) - \sum_{i=1}^{i_{in}-1}\int_{t_{in}}^{t_{end}} \kappa e^{a(m_i-m_0)} \frac{(p-1)c^{p-1}}{(t-t_i+c)^p} dt\\
&\quad - \sum_{i=i_{in}}^{N_{tot}}\int_{t_i}^{t_{end}} \kappa e^{a(m_i-m_0)} \frac{(p-1)c^{p-1}}{(t-t_i+c)^p} dt\\
&=\sum_{i=i_{in}}^{N_{tot}} \ln\left[\lambda(t_i|\mathcal{H}_{t_i})\right] - \varpi(t_{end}-t_{in})\\
&\quad + \sum_{i=1}^{i_{in}-1} \kappa e^{a(m_i-m_0)} c^{p-1}\left[(t_{end}-t_i+c)^{1-p} - (t_{in}-t_i+c)^{1-p}\right]\\
&\quad + \sum_{i=i_{in}}^{N_{tot}} \kappa e^{a(m_i-m_0)} c^{p-1}\left[(t_{end}-t_i+c)^{1-p} - c^{1-p}\right]\\
&=\sum_{i=i_{in}}^{N_{tot}} \ln\left[\lambda(t_i, m_i|\mathcal{H}_{t_i})\right] - \varpi(t_{end}-t_{in}) + \sum_{i=1}^{N_{tot}} \kappa e^{a(m_i-m_0)} c^{p-1}(t_{end}-t_i+c)^{1-p}\\
&\quad - \sum_{i=1}^{i_{in}-1} \kappa e^{a(m_i-m_0)} c^{p-1}(t_{in}-t_i+c)^{1-p} - \sum_{i=i_{in}}^{N_{tot}} \kappa e^{a(m_i-m_0)}.
\end{aligned}
$$

Once obtained the maximum likelihood estimates, one may ask how to choose the best model when proposing several competitive ones. The tool used for comparing the goodness-of-fit of the models for a fixed dataset is the *Akaike's Information Criterion (AIC)* [Akaike, 1974]:

$$AIC = -2(maximum\ of\ log-likelihood) + 2(number\ of\ parameters). \tag{2.25}$$

The model with the smaller AIC value is considered the one that shows the better fit to the data. It is useful to say that, if one compares the models H_0 and H_1 with k_0 and k_1 parameters, respectively, then the log-likelihood ratio statistic is

$$-2\ln\left(\frac{L_0}{L_1}\right) = AIC(H_0) - AIC(H_1) + 2|k_0 - k_1|,$$

where L_0, L_1 are the likelihoods of the first and the second models, respectively. When H_0 is a particular case included in H_1, then, under the null hypothesis H_0, the statistic $-2\ln\left(\frac{L_0}{L_1}\right)$ is expected to follow a Chi-squared distribution with $k = k_0 + k_1$ degrees of freedom.

Concluding, recently other methods have been proposed for the estimation of the ETAS parameters, as for example the simulated annealing [Lombardi, 2015]. This could be very interesting in the spatiotemporal setting, in order to propose easier algorithms with a higher efficiency.

2.3.2 Simulation

The simulation procedure is fundamental to construct synthetic catalogs, with which one can obtain results not influenced by any kind of "real effect" underlying the real process. The algorithm for the temporal ETAS has been proposed by Ogata on the basis of the *thinning method* for the simulation of point processes [Lewis and Shedler, 1979; Musmeci and Vere-Jones, 1992; Ogata, 1981, 1998]. His FORTRAN program is [**etasim.f**], available online at `http://www.ism.ac.jp/~ogata/Ssg/ssg_softwaresE.html`. This program allows to choose the option of the pure temporal simulation or the spatiotemporal one. It requires in input the set of parameters appearing in the conditional intensity function characterizing the ETAS model. Let's notice that in the Ogata's algorithm considered, the combination of parameters could lead to explosive simulated data, since the program does not control whether the non-explosion conditions $\beta > a$, $p > 1$ and $\kappa < \frac{\beta - a}{\beta}$ are verified.

The magnitudes can be either simulated by the Gutenberg-Richter law, or can be taken from a given catalog to be specified in input. In the first case, one has to provide the b-value and the number of events to be simulated; furthermore, the Gutenberg-Richter law is used independently for all the events in the catalog, without distinguishing between spontaneous and triggered shocks. In the second case, the program simulates instead the same number of events with magnitudes bigger that or equal to the completeness threshold of the input catalog; the simulation starts after a precursory period which depends on the same history of the input catalog itself. For the computation of the ETAS intensity, one can also specify a value for the reference magnitude different from the threshold of completeness. In the case of the spatiotemporal simulation, one has to give in input also a file for the identification of the background. It consists of a grid of the area under consideration, in which each cell is labeled with latitude, longitude and number of events contained.

The simulation is organized as follows.

- Based on the choice done, the magnitudes are taken from the real catalog given in input, or are simulated according to the exponential distribution given by the Gutenberg-Richter law. More precisely, in the latter case the magnitudes are obtained by considering the usual method for the inversion of the cumulative distribution function.

- A uniform random number between zero and one is generated to simulate the occurrence time, to which the event is assigned with a certain probability. The time simulation is based on the temporal ETAS intensity (2.8).

- Once the event has been assigned to a simulated time, it is labeled as a background or a triggered shock with a certain probability. This probability is obtained again by considering only the temporal intensity of the model.

- If the event under consideration is labeled as spontaneous, it is spatially collocated based on the rates of the cells in the grid given in input.

- If the event under consideration is labeled as triggered, its potential mother shock has to be assigned to it in order to be spatially located. This is done by considering all the contributes of the possible mother shocks and individuating the one which gives the highest probability for the triggered event considered. Again the intensity used in this case is the temporal one. Once found the triggering event, the program assigns a spatial location to the epicenter of the aftershock considered. To this aim, it generates a random angle between $0°C$ and $360°C$ and simulates the radius again by the usual method for the inversion of the cumulative distribution function.

The simulation is done in condition of isotropy, that is without preferential directions. Obviously, when the simulation is pure temporal, all the steps relative to the spatial location of the epicenters are not performed.

We want to add that, in the literature, there are several works about the simulation of the Hawkes processes, class to which belongs the ETAS model, too [Brémaud et al., 2002; Møller and Rasmussen, 2005, 2006]. Nevertheless, since the Ogata'a program is specific for the ETAS, this is the program used in this thesis when we want to get the simulation of the catalogs based on the temporal ETAS rate.

2.3.3 Residual analysis

The minimum AIC procedure allows to find the model which provides the best fit for the earthquake process data, among some competitive models considered. Nevertheless, it remains the possibility that one or some of the major features of the real data are not reproduced by the estimated model, selected with the Akaike's criterion. Then, further models with a similar AIC value must be taken into account. To this aim, the *residual analysis* has been proposed to amplify the features of the data deviating from the model. It is a pure temporal analysis based on the following random time-change:

$$\Lambda_t = \int_0^t \left[\varpi + \sum_{\{i|t_i<s\}} \frac{\kappa e^{a(m_i-m_0)}}{(s-t_i+c)^p} \right] ds, \tag{2.26}$$

where the integrand is the time-magnitude ETAS conditional intensity with the non-normalized Omori law given in (2.8). This integral is a monotonically increasing function, due to the non-negativity of the integrand, and consists of a one-to-one transformation from $\{t_i\}$ to $\{\Lambda_{t_i}\}$. This latter sequence of transformed times is called *residual process* and is a stationary Poisson process with rate one [Papangelou, 1972]. It follows that, if the conditional intensity obtained with the estimated parameters is a good approximation of the real one, then the relative transformed times should be a stationary Poisson process. Each property of the residual process which deviates from that expected from a stationary Poisson process, corresponds to a feature of the real catalog not reproduced by the model considered. For the residual analysis one can consider any graphical test for complete randomness, or stationary Poisson.

The Ogata's FORTRAN program is [**retas.f**], available on line at `http://www.ism.ac.jp/~ogata/Ssg/ssg_softwaresE.html`. This program computes only the above random time-change. Instead, the FORTRAN program [**RESIDUALS_ETAS.f**], written by the seismologist Anna Maria Lombardi of the INGV, performs also two graphical tests: the Kolmogorov-Smirnov and the Runs ones. The first one verifies if the interevent times are distributed according to an exponential law with parameter one. The second test verifies instead if the interevent times are independent. The typical acceptance threshold is 5%. Then, if the probabilities obtained for the considered model with the above two tests are bigger than this threshold, the model is accepted. The program [**RESIDUALS_ETAS.f**] requires in input the catalog for which performing the residual analysis, the set of parameters, the target period and the starting time of the learning one and the magnitude thresholds. The output consists of two files: the first one contains the data of the input catalog with also the transformed times; the second one contains the total number of data, the number of events expected by the model and the results of the tests. This program is the one used for the experimental analysis of this thesis. The analysis will be performed for real catalogs in Chapter 3 and for simulated catalogs in Chapter 5.

Chapter 3

Experimental analysis

As explained in Chapter 2, in the classical ETAS model the Gutenberg-Richter law is assumed to be valid both for the background events and for the triggered ones and is independent of the characteristics of past seismicity. More precisely, the magnitude of each shock is independent of the magnitude of the corresponding triggering event [Zhuang et al., 2002]. However, by considering events in causal relation, for example mother/daughter, it seems natural to assume that the magnitude of a daughter event is conditionally dependent on the magnitude of the corresponding mother event. In order to look for experimental evidences supporting this hypothesis, we perform two different types of analysis of four seismic catalogs. As explained below, in each analysis we use the kernel density estimation method to obtain the distribution of the triggered events' magnitudes, in order to assess its variation with the magnitude of the corresponding mother events. The main problem for this analyses is assessing if an event is spontaneous or triggered and, in the latter case, who is the relative triggering shock. The difference between the two types of analysis stays in the assessment of the mother/-daughter relation. The aim is to find some kind of relation that would model the real combined behavior of the magnitudes of these historically connected events. It is reasonable and natural to expect that the distribution density of triggered events' magnitude increases (decreases) with the mother events' magnitude increase (decrease). Moreover, if this is true, the expected value of the triggered events' magnitude should also be increasing in the same way.

As very often done in practice, we assume here that the reference magnitude is equal to the completeness one. Then, in this chapter, we will refer only to the latter threshold value.

3.1 The catalogs analyzed

We analyze three Italian catalogs and a Californian one. They differ in spatial end/or temporal extension and in the presence or not of a strong shock. The main information of the four catalogs are the following.

1) The first catalog includes events occurred from April the 16^{th}, 2005 till January the 25^{th}, 2012 in the region including the whole Italy (latitude from 35 to 48, longitude from 6 to 19). The ZMAP estimated completeness magnitude is 2.7. However, the latter is considered too high by the seismologists of the *Istituto Nazionale di Geofisica e Vulcanologia* (INGV), who suggested us to use 2.5. For the seismicity map, see Fig. 3.1.

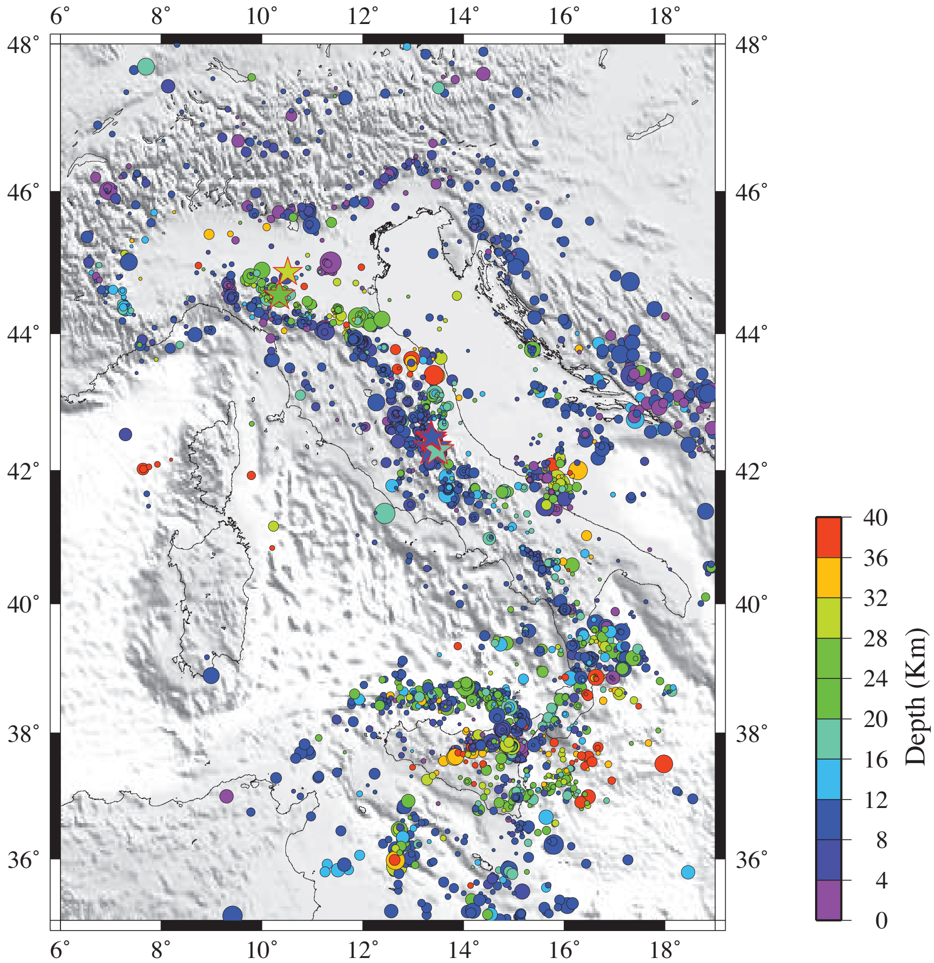


Figure 3.1: Seismicity map of the first catalog concerning the whole Italy from April the 16^{th}, 2005 till January the 25^{th}, 2012. The map has been obtained by the Generic Mapping Tools (GMT).

2) The second catalog includes events occurred in the portion of Abruzzo region (Italy) corresponding to the square from latitude +41.866 to +42.866 and from longitude +12.8944 to +13.8944. This subregion includes L'Aquila. The temporal interval is the same as for catalog one. The estimated value for the completeness magnitude is 1.8. This catalog includes the strong shock of magnitude 6.1 occurred in L'Aquila on April the 6^{th}, 2009. The seismicity map is shown in Fig. 3.2.

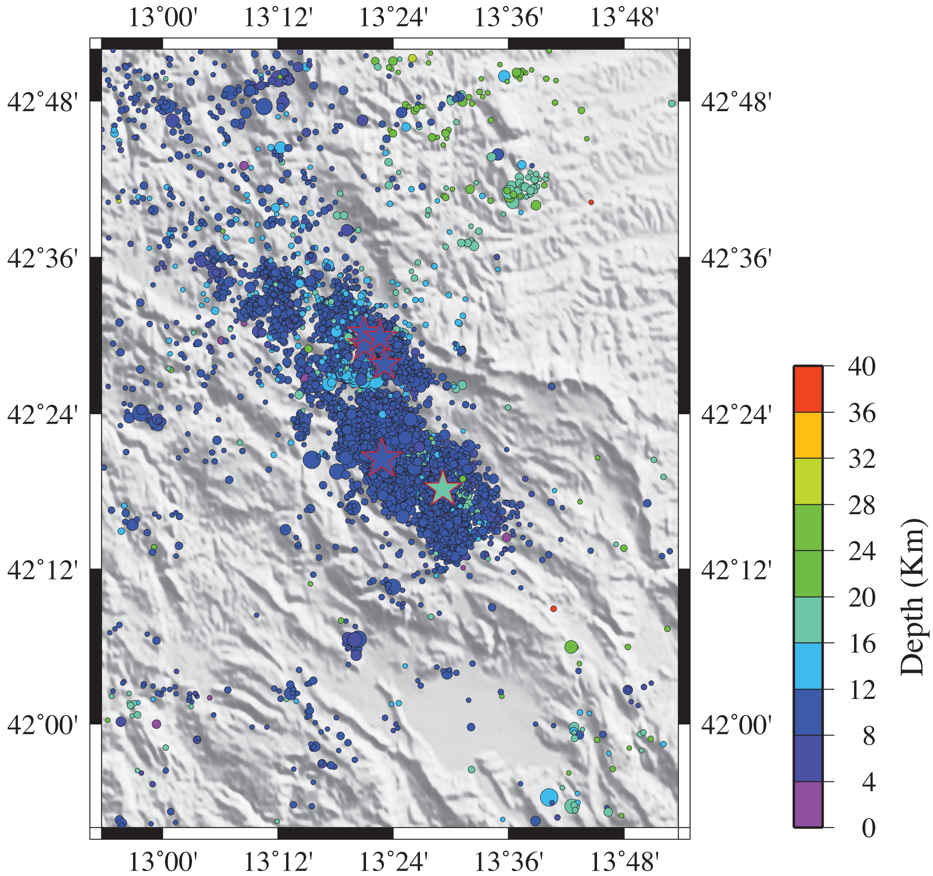


Figure 3.2: Seismicity map of the second catalog concerning the square centered at L'Aquila, from April the 16^{th}, 2005 till January the 25^{th}, 2012. The map has been obtained by the Generic Mapping Tools (GMT).

3) The third catalog differs from the previous one only for the temporal interval, which now goes from April the 16^{th}, 2005 till April the 5^{th}, 2009. The value of completeness magnitude is estimated equal to 1.5. This catalog doesn't include strong shocks. For the seismicity map, one can see Fig. 3.3.

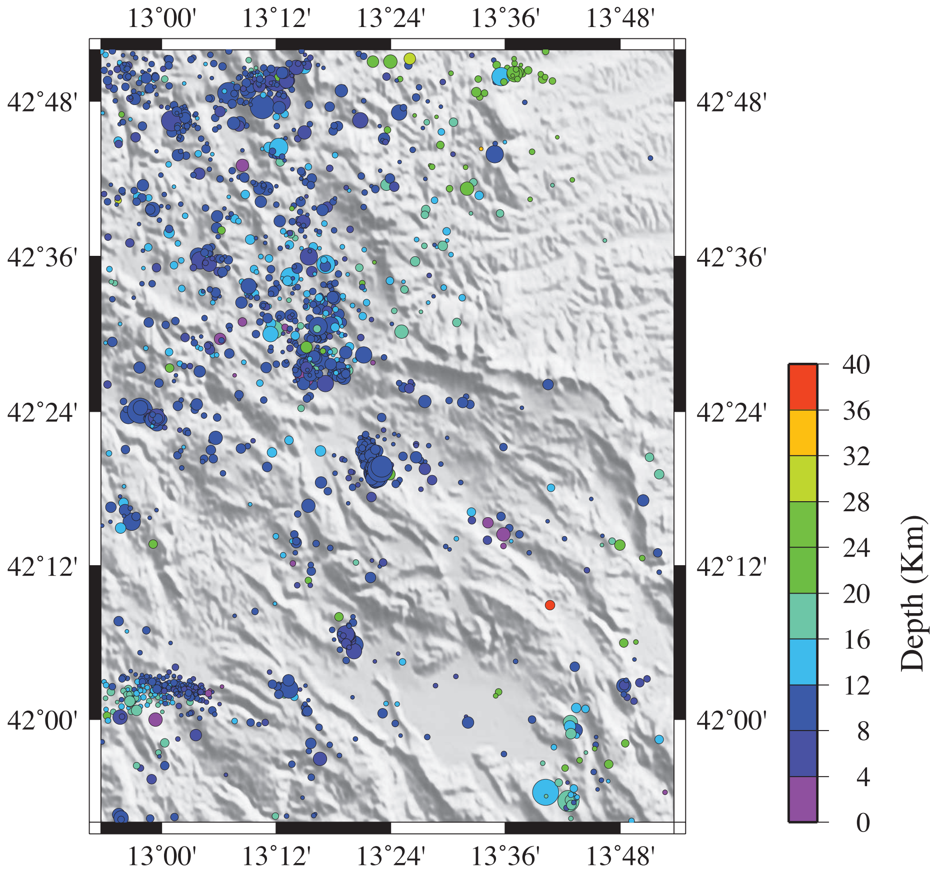


Figure 3.3: Seismicity map of the third catalog concerning the square centered at L'Aquila, from April the 16^{th}, 2005 till April the 5^{th}, 2009. The map has been obtained by the Generic Mapping Tools (GMT).

4) The fourth catalog includes events occurred from January the 1^{st}, 1984 till December the 31^{th}, 1991, in the portion of the Southern California corresponding to the square from latitude +33.75 to +34.75 and from longitude -117.5 to -116.5. This catalog is a portion of the waveform earthquake catalog relocated by Hauksson *et al.* in 2011 [Hauksson et al., 2012; Lin et al., 2007]. It has been given to us by the seismologists of the INGV and it contains events with magnitudes from 2 on. This is the value estimated by the seismologists for the completeness magnitude. To be thorough, we have estimated again the completeness magnitude of this catalog with the ZMAP program, obtaining again 2 (see Fig. 3.5). This ensures us that the completeness value is not bigger than this number. Even if it were smaller, the portion of catalog used here would however be complete. Fig. 3.4 contains the seismicity map for this catalog.

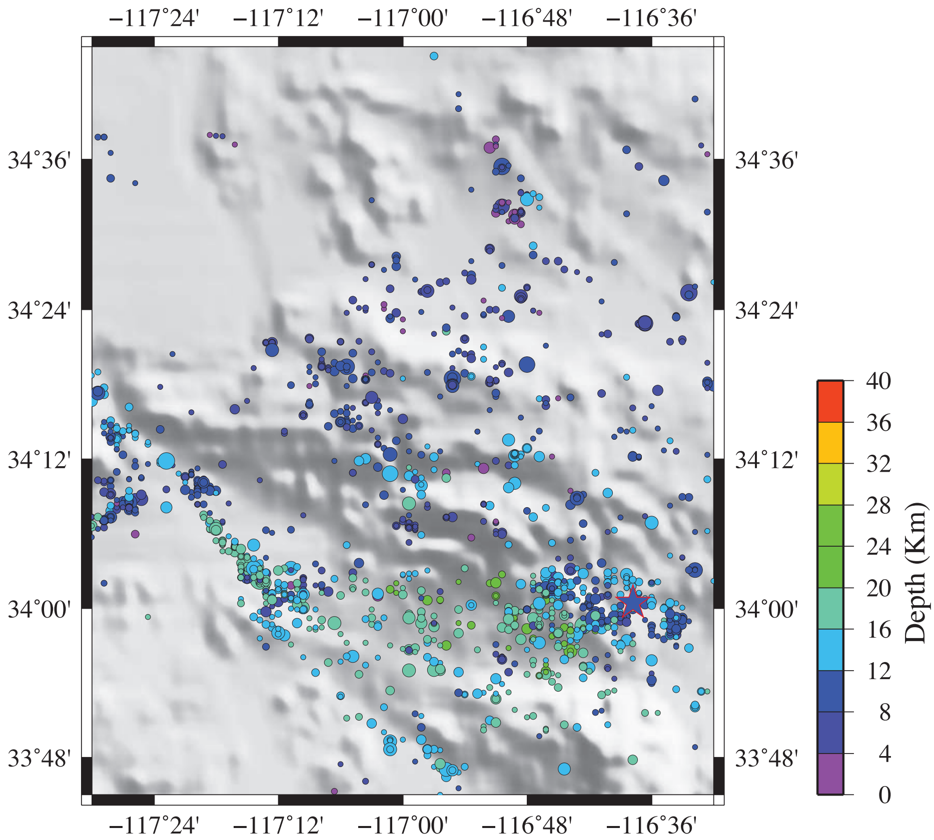


Figure 3.4: Seismicity map of the fourth catalog concerning the Southern California, from January the 1^{st}, 1984 till December the 31^{th}, 1991. The map has been obtained by the Generic Mapping Tools (GMT).

We have chosen the above temporal periods for the Italian catalogs since, when we have started the study, only the data available online for these periods had had the final revision.

In the first catalog there is the presence of events that are very distant from each other, as can be seen from the mean values of the distances between the events in causal relation, shown below in Section 3.3. Because of this, we expect that in this case the above-mentioned hypothesis of magnitude correlation is absent, or not so evident. On the other hand, we expect that this dependence becomes far more evident for catalogs not including events that are spatially "too" distant to each other. This is the reason for including in the study catalogs two, three and four. More precisely, we consider the third catalog to investigate the influence of a strong shock and the fourth one to verify the validity of our hypothesis also for a catalog relative to a region far away from the others and with a different seismicity.

In all the previous cases the maximum depth considered is equal to 40 km: all the events with deeper depths are excluded. This is due to the fact that, for the latter events, one could have problems with the complete record of the data. In fact, the deeper the hypocenter is, the lower is the precision of the seismographs. Actually, in the examined catalogs either there are no events deeper than 40 km or there are only a few of them, and then their exclusion doesn't affect the analysis.

The values of the completeness magnitude have been computed with the ZMAP software [Wiemer, 2001]. More precisely, we have used the Shi and Bolt uncertainty [Marzocchi and Sandri, 2003; Shi and Bolt, 1982], according to whom the error of the b-value of the Gutenberg-Richter law is estimated as $\hat{\sigma}_{\hat{b}} = 2.30\hat{b}^2\sqrt{\frac{\sum_{i=1}^{N_{ev}}(m_i-\hat{\mu}_m)^2}{N_{ev}(N_{ev}-1)}}$, where N_{ev} is the number of events and $\hat{\mu}_m$ is the sampling average of the magnitudes m_i, $i = 1, \ldots, N_{ev}$, which are supposed to be identically distributed. The plots obtained by estimating the completeness magnitude for the four catalogs considered are shown in Fig. 3.5.

Furthermore, the magnitudes versus occurrence times plots of Fig. 3.6 show the general composition of the events in each of the four catalogs considered. Let's focus on the second plot from the top of the above figure, relative to L'Aquila till 2012. If we zoom a rectangle including the days just after the strong shock, corresponding to the peak, we get Fig. 3.7, where we can see that there are no events with a magnitude lower than about 2.5: even if the ZMAP estimated completeness magnitude is 1.8 for this catalog, this value is probably too low in the days just after the shock of magnitude 6.1. Then, in these days there are there are probably several events with low magnitude (less than 2.5) and they have not been recorded.

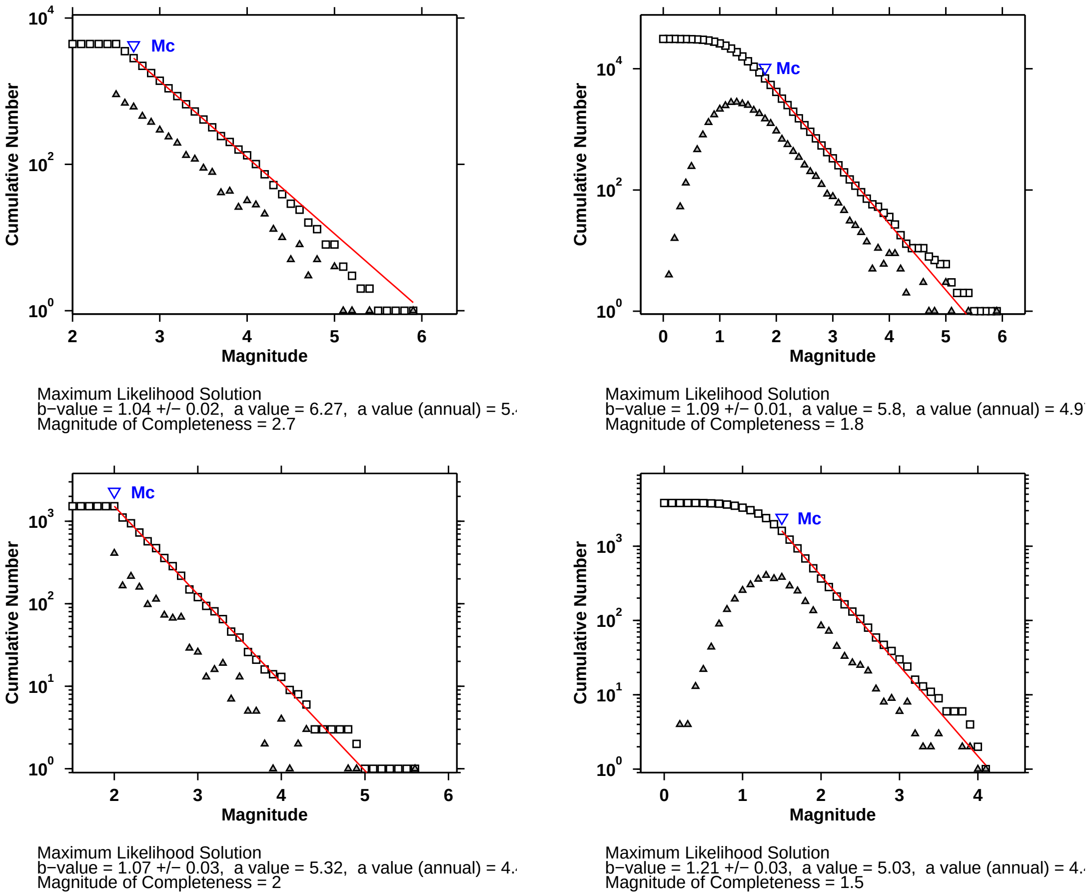


Figure 3.5: Frequency-magnitude distributions obtained with the ZMAP program for the four catalogs considered. Starting with the upper-left plot and proceeding clockwise, the plots concern the first, second, third and fourth catalogs, respectively relative to the whole Italy, L'Aquila till 2012, L'Aquila till April the 5^{th}, 2009 and the Southern California. The minimum values at which there's a deviation from the Gutenberg-Richter law are the completeness magnitudes.

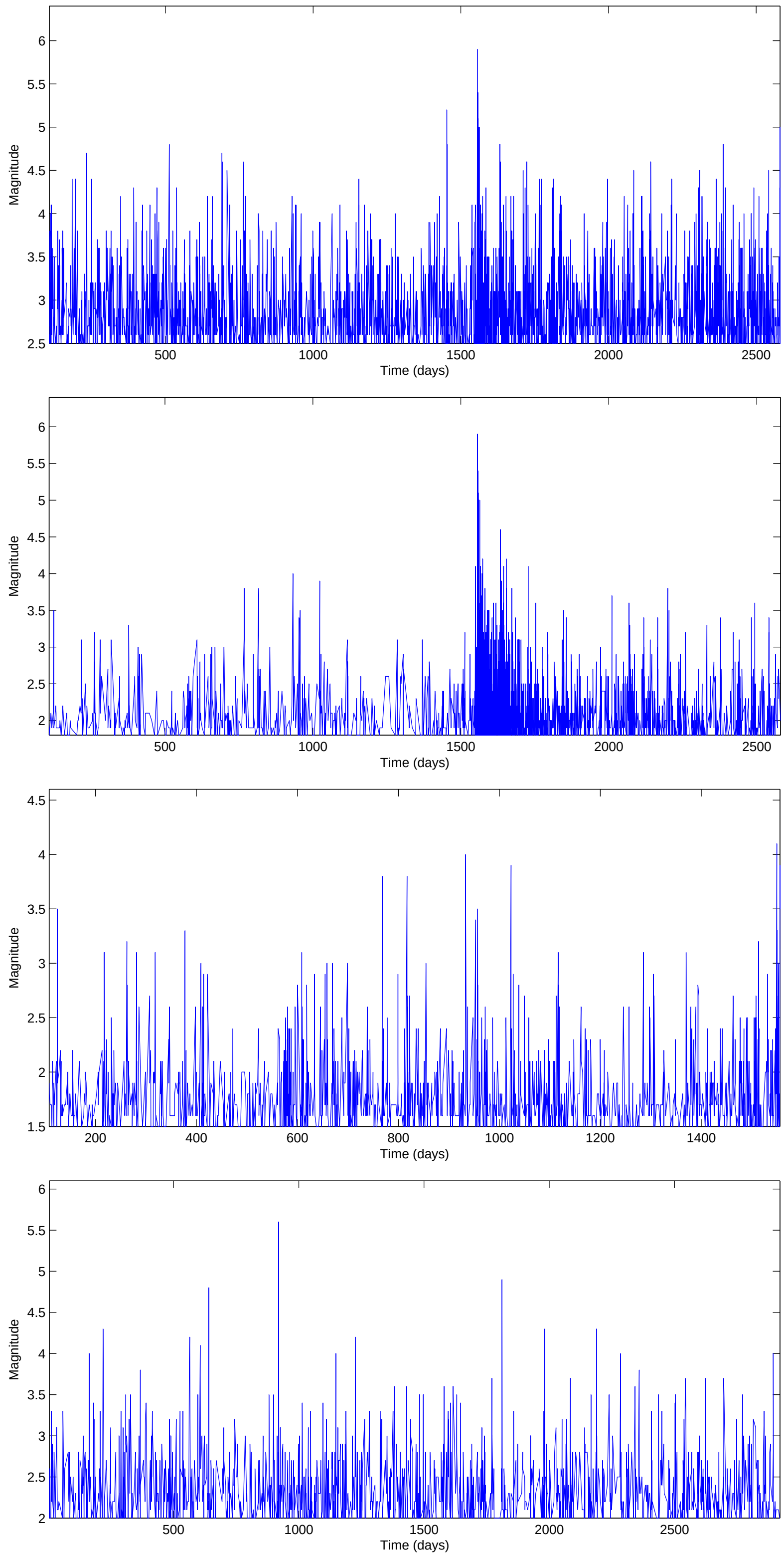


Figure 3.6: Times versus magnitudes of the events in the four catalogs considered. Starting with the plot at the top and running down, the plots concern the first, second, third and fourth catalogs, respectively relative to the whole Italy, L'Aquila till 2012, L'Aquila till April the 5^{th}, 2009 and the Southern California.

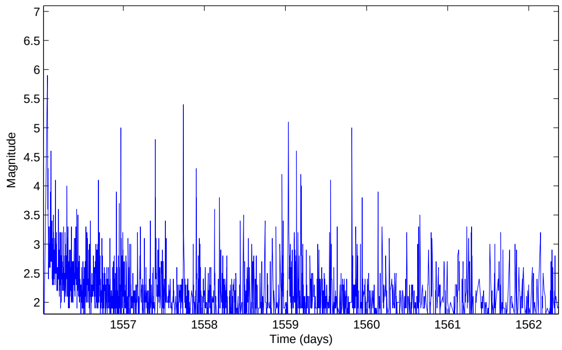


Figure 3.7: Zoom on the days after the strong shock on April the 6^{th}, 2009, of the times versus magnitudes plot relative to the second catalog, concerning L'Aquila till 2012. We can see the absence of events with small magnitudes in the period just after this shock.

3.2 The two types of analysis

In order to have coherent and consistent data, we have excluded from the analysis all the events in the four catalogs not measured with the local or moment magnitude scales. Generally, the events' magnitudes measured with other scales are very low, very often smaller than 1. Then, the events of this kind, not already excluded by cutting according to the completeness magnitude, are very few. Again, the analysis is not affected by this exceptions.

In the two types of analysis, performed for each catalog, the first step is the same.

1. Consider four magnitude subintervals contained in the magnitude range, from the completeness value m_c to the maximum one m_{max} (both included). The first and the last subintervals considered are of kind $[m_c, \bar{m}_1]$ and $[\bar{m}_2, m_{max}]$, for two chosen values $\bar{m}_1$ and $\bar{m}_2$; the two intermediate ones are instead suitably chosen between them; the size of each subinterval is also chosen, for each catalog, in such a way to have a comparable number of triggered events in all the four subintervals considered.

We then proceed differently for the two types of analysis.

3.2.1 Algorithm 1: first analysis

The first approach consists of the following steps, repeated for each catalog.

2a. Group the events whose magnitudes belong to each of the four subintervals of *Step 1*.

3a. For each event $Ev_{1,i}$ in the first subinterval, with magnitude $m_{1,i}$ and time of occurrence $t_{1,i}$, find all the shocks occurring in the time interval $[t_{1,i}, t_{1,i} + \delta^*]$. The choice of the amplitude δ^* is explained in Subsection 3.2.2.

4a. Group all the magnitudes $m_{1,i,j}$ of all the shocks $t_{1,i,j}$ belonging to all the previous time intervals, i.e., $t_{1,i,j} \in [t_{1,i}, t_{1,i} + \delta^*]$. If one event belongs for example to both $[t_{1,1}, t_{1,1}+\delta^*]$ and $[t_{1,2}, t_{1,2}+\delta^*]$, it is counted twice.

5a. Repeat the previous two steps for the other three magnitude subintervals considered.

6a. Consider the set $G_m = \{m_{1,i,j}, m_{2,i,j}, m_{3,i,j}, m_{4,i,j}\}_{i,j}$. Estimate the probability density function by the kernel density estimation method.

For the MATLAB code, see the section **Algorithm 1** in the file **Algorithms.pdf**.

The kernel density estimation method is a non-parametric approach very used in statistics [Parzen, 1962; Rosenblatt, 1956]. We proceed with this method as follows. Firstly, for simplicity of notation, we relabel the elements of G_m as $(m_1, m_2, \dots, m_{n_{tot}})$ and we consider the corresponding frequencies $f = (f_1, f_2, \dots, f_{n_{tot}})$, where n_{tot} is the total number of events considered to be triggered as explained before. As already said, one event can be counted more than one time and then these frequencies are not "true". We consider also the set m of 1000 magnitudes equispaced from the completeness threshold to the maximum value. Then, we compute the *kernel density estimator* of the empirical magnitude distribution M, for $m \in \{m_c + k(m_{max} - m_c)/1000,\ k = 0, 1, ..., 1000\}$, as

$$\hat{M}_\gamma(m) = \frac{1}{F}\sum_{i=1}^{n_{tot}} f_i K\left(\frac{m - m_i}{\gamma}\right), \quad \text{with} \quad F = \sum_{i=1}^{n_{tot}} K\left(\frac{m - m_i}{\gamma}\right), \tag{3.1}$$

where $K(\cdot)$ is known as *kernel* and the positive parameter γ is the *bandwidth* [Parzen, 1962]. The above formula is obtained by adapting the *Nadaraya-Watson* kernel for kernel regression [Scott, 1992].

For the ease of the reader, we recall the following definition.

Definition 16. A kernel $K(\cdot)$ is a non-negative, real-valued function such that

$$\int_{-\infty}^{\infty} K(x)dx = 1, \quad \text{and} \quad K(x) = K(-x) \quad \forall x \in \mathbb{R}.$$

As very often done, here we use the Gaussian kernel

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

The value of the bandwidth is chosen using the leave-one-out cross-validation method [Scott, 1992], opportunely modified and implemented by us. More precisely, we choose the value γ that minimizes the quantity $\sum_{i=1}^{n_{tot}} |\hat{f}_i - f_i|$, where

$$\hat{f}_i = \frac{1}{\bar{F}_i} \sum_{j \neq i} f_j K\left(\frac{m_i - m_j}{\gamma}\right), \quad \text{with} \quad \bar{F}_i = \sum_{j \neq i} K\left(\frac{m_i - m_j}{\gamma}\right).$$

Let's observe that, differently from formula (3.1), the value f_i doesn't contribute to $\hat{f}_i$.

3.2.2 Algorithm 1.1: choice of the time amplitude δ^*

The time window δ^* is chosen here in such a way that two seismic events, separated by a time larger than δ^*, are not in causal relation. The algorithm consists of the following steps.

- Divide the whole time interval in daily subintervals and count the number of events that occur in each subinterval. Denote this temporal sequence with X_t, i.e., X_t is the number of events occurred in day t..
- Starting from this temporal sequence, compute an estimate $\widehat{R}(\delta)$ of the autocorrelation function at different integer values of time lag δ:

$$\widehat{R}(\delta) = \frac{1}{(n_s - \delta)\widehat{V}} \sum_{t=1}^{n_s - \delta} (X_t - \widehat{\mu})(X_{t+\delta} - \widehat{\mu}), \tag{3.2}$$

  where n_s is the temporal dimension of the sample, $\delta = 0, 1, 2, \ldots$ and

$$\widehat{\mu} = \frac{1}{n_s} \sum_{t=1}^{n_s} X_t, \qquad \widehat{V} = \frac{1}{n_s - 1} \sum_{t=1}^{n_s} (X_t - \widehat{\mu})^2$$

  are the sample mean and variance, respectively [Priestley, 1982].
- Model $\widehat{R}(\delta)$ by the power law $\left(1 + \frac{t}{b_1}\right)^{-b_2}$ containing the two parameters b_1, b_2, which are estimated by least squares.

- Find the value δ^* such that the power law model is equal to 0.05. In the cases examined, this choice produces *p-values* always smaller than 0.01.

For the MATLAB code, see the section **Algorithm 1.1** in the file **Algorithms.pdf**.

The autocorrelation function is a very important tool already used in previous studies. It has been used for example to analyze seismic time sequences like in our case [Caputo and Sebastiani, 2011], to test the climate-seismicity [Molchanov, 2010] or in the matter of ambient seismic noise [Cezar, 2002; Ohmi and Hirahara, 2009]. Recall that the ambient seismic noise is a relative persistent vibration of the ground, caused by ≪a diversity of different, spatially distributed, mostly unrelated and often continuous sources [...] without a defined phase spectrum≫, as defined in [Bormann, 2009].

Due to the strong shock on April the 6^{th}, 2009, the second catalog shows a clear non-stationary pattern. Because of this, in order to find the δ^* value for this catalog and then to perform the first kind of analysis, we transform the original dataset by considering the following well-known random time-change, already discussed in (2.26):

$$\int_0^t \lambda(s|\mathcal{H}_t)ds, \tag{3.3}$$

where $\lambda(\cdot|\mathcal{H}_t)$ is the time-magnitude ETAS rate for seismic events given in (2.8) (see for example [Ogata, 1988]). It follows that, in the case of the second catalog, even the first analysis depends on the ETAS model in the way of transforming the occurrence times. More precisely, we obtain the catalog with the transformed times by using the FORTRAN program for the residual analysis [**RESIDUALS_ETAS.f**], in which we have considered as precursory period the one that ends at the time of the last event occurred on April the 5^{th}, 2009. The same period has been used for the estimation of the parameters appearing in the above rate. This estimation has been computed with the FORTRAN program [**etas.f**] written by Ogata and already presented in the previous chapter (see Section 2.3.1), by which statistical inference on the parameters of the ETAS model is performed [Ogata, 2006]. Due to the presence of the learning period, the catalog with the transformed times concerns only the part of the original catalog starting from April the 6^{th}, 2009. As explained before, by the transformation (3.3) the process becomes stationary. The consequent plot of magnitudes versus occurrence times is shown in Fig. 3.8. In this figure it can be again clearly seen that there is the lack of events with low magnitudes, let's say less than 2.5, in the first days after the strong shock.

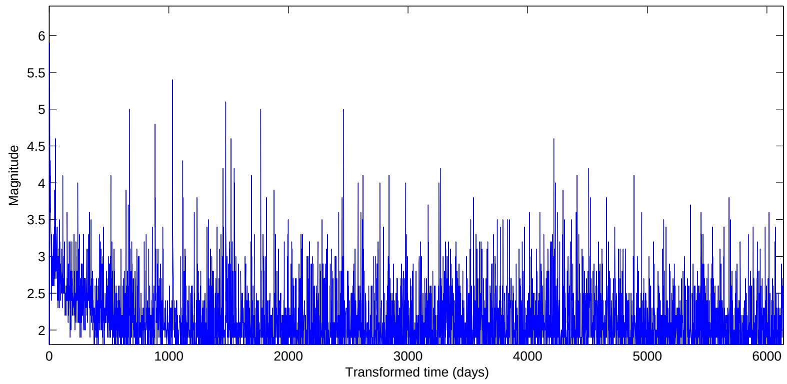


Figure 3.8: Times versus magnitudes of the events in the second catalog, concerning L'Aquila from April the 16^{th}, 2005 till January the 25^{th}, 2012, where the times have been transformed with the random time-change (3.3).

If we consider catalogs with different magnitudes, we get different estimations of the parameters c, p, κ, a of the ETAS model. It follows that, even if the Omori-Utsu law is independent of the magnitudes, the times of occurrence of the process described by the ETAS model depend on the magnitudes on the mother events, in fact in the triggered rate of the latter model the triggering events' magnitudes influence the temporal decay through the productivity law. However, since δ^* is used in the first analysis, which do not include the ETAS modeling, the times are here completely independent with respect to the magnitudes. Then, the value of δ^* doesn't depend on the magnitude of the events considered and is taken equal for each shock in the catalog. It differs only from catalog to catalog.

The time-correlations plots obtained for the four catalogs considered are shown in Fig. 3.9.

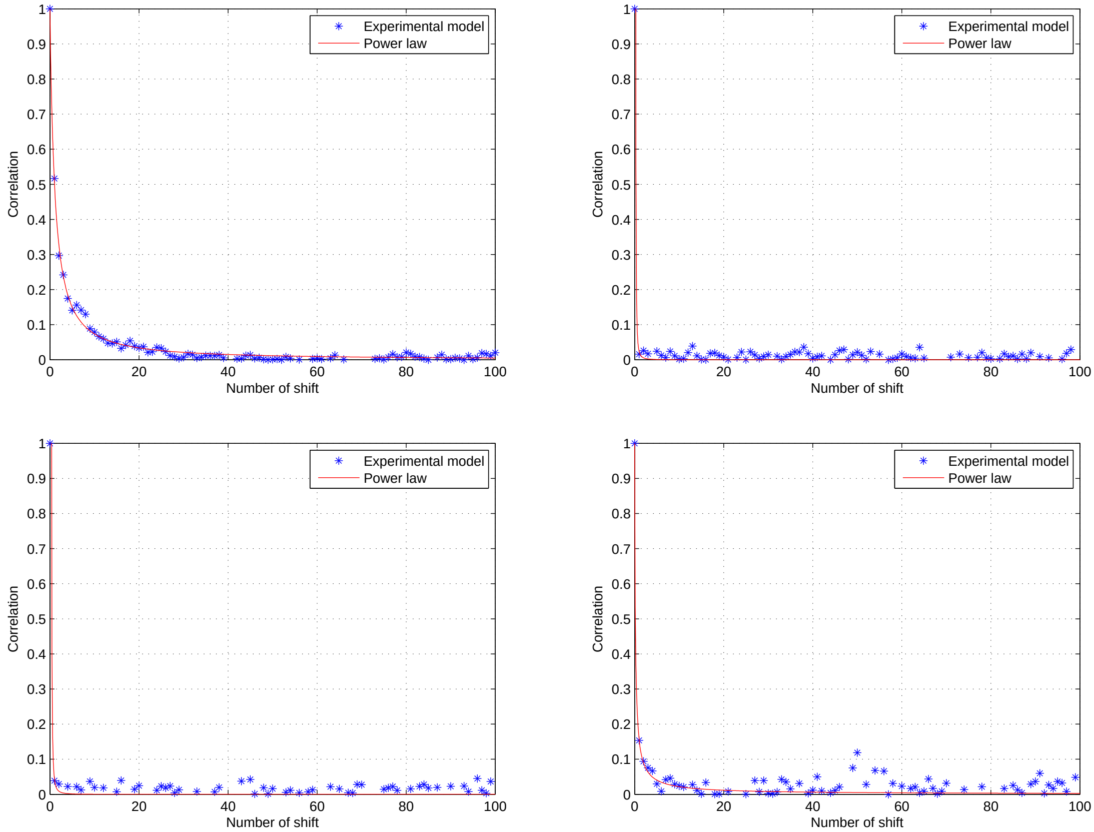


Figure 3.9: Time-correlations of the four catalogs considered. Starting with the upper-left plot and proceeding clockwise, the plots concern the first, second, third and fourth catalogs, respectively relative to the whole Italy, L'Aquila till 2012, L'Aquila till April the 5^{th}, 2009 and the Southern California.

3.2.3 Algorithm 2: second analysis

The second type of analysis, performed again for each catalog, consists of the following steps.

2b. Apply the FORTRAN program [**etas.f**] to estimate the parameters of the temporal ETAS model.

3b. For each event of the catalog, find the mother shock that most likely triggered it. The method to this aim is briefly explained just below.

4b. Consider the first of the four magnitude subintervals of *Step 1*.

5b. Group the triggered events whose triggering shock's magnitude belongs to the considered subinterval. Call the set of these events' magnitudes $\tilde{M}_1$.

6b. Repeat the previous step for the other three magnitude subintervals.

7b. Consider the set of all the magnitudes $\tilde{G}_{\tilde{m}} = \left\{ \tilde{M}_1 \cup \tilde{M}_2 \cup \tilde{M}_3 \cup \tilde{M}_4 \right\}$ to estimate the probability density function relative to triggered events' magnitude by using the Gaussian kernel density estimation method described above in Subsection 3.2.1. The value of the bandwidth is determined as Subsection 3.2.1, too.

For the MATLAB code, see the section **Algorithm 2** in the file **Algorithms.pdf**.

The method we use in order to find the mother events of *Step 3b*, is a variation of the Ogata's criterion. More precisely, the latter considers as mother of the i^{th} event the shock occurred in the smallest time t_J such that the ratio between the ETAS rate till t_J and the ETAS rate over all the previous $t_j < t_i$ is bigger than a uniform random number U generated in $(0, 1]$; instead we consider as mother the preceding event, among $p(m_i)\varpi$ and $p(m_i)\rho(m_i) \int_0^{t_J - t_i} \Phi(s)ds$, $i \leq J$, that gives the highest contribute to the ETAS rate.

We do not apply this analysis to the first catalog, relative to the whole Italy. In fact, it is not meaningful to use the pure temporal ETAS model in estimating the parameters for such a large region like the one in the first catalog.

3.3 Results

We present here some results obtained by the two types of the above-mentioned analysis for the catalogs considered.

In Fig. 3.10, we plot the estimated densities of triggered events' magnitude, obtained by the first type of analysis for the first catalog (whole Italy). By looking at the four curves, one can see that there are no apparent differences among the densities. As said before, this can be explained by the fact that there are many pairs of events that are close to each other along time, but spatially very separated. In fact, the mean distance between the events of the pairs in causal relation is 140 Km. The elements of these pairs are erroneously put in relation in the analysis.

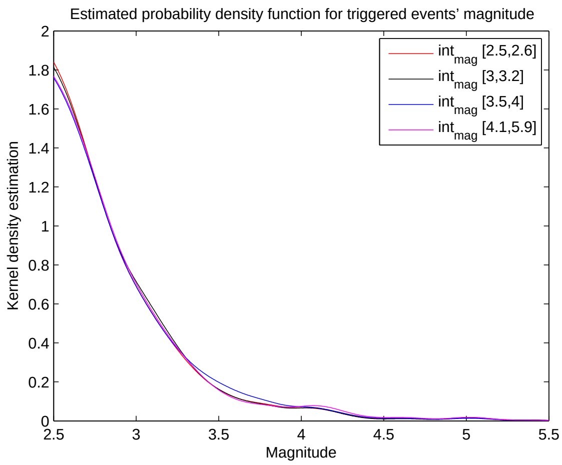


Figure 3.10: Kernel density estimation of triggered events' magnitude in the first catalog (whole Italy), concerning the first type of analysis. The considered intervals in which triggering events' magnitudes fall are: $[2.5, 2.6]$, $[3, 3.2]$, $[3.5, 4]$ and $[4.1, 5.9]$ (red, black, blue and magenta curves, respectively). The δ^* value is equal to fifteen days. The optimal bandwidth value for the Normal kernel density estimation is 0.12, 0.11, 0.12, 0.12 for the four intervals considered, respectively.

In Fig. 3.11, there are the estimated densities of triggered events' magnitude obtained by the first and the second types of analysis (plots at the top and at the bottom, respectively), relative to the second catalog (L'Aquila till 2012). Here, the spatial extension of the region analyzed is far smaller than the one of the previous catalog. Let's recall that, in this case, the first analysis has been applied to the dataset transformed by the random time-change (3.3), due to the non-stationary pattern of the process caused by the presence of the strong shock on April the 6^{th}, 2009.

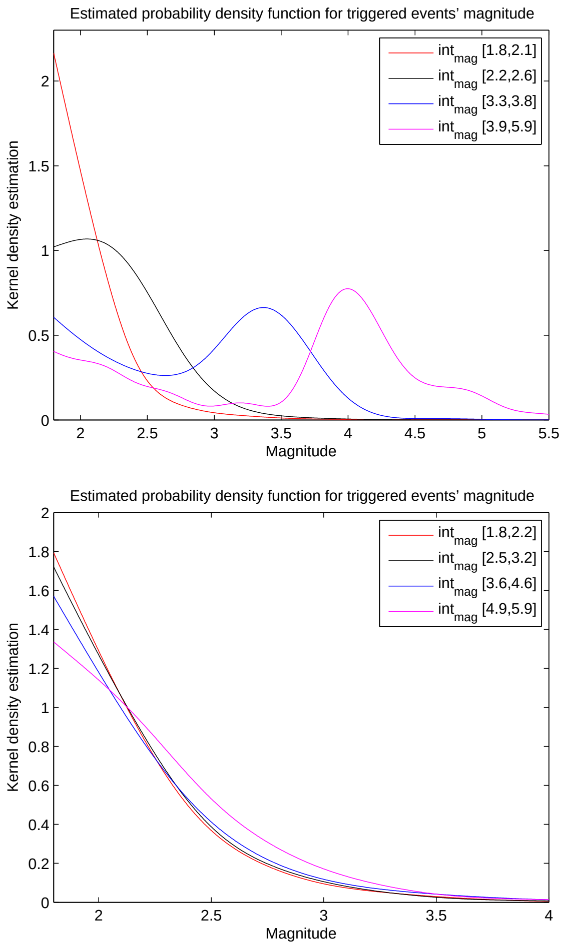


Figure 3.11: Kernel density estimation of triggered events' magnitude in the second catalog (L'Aquila till 2012), concerning the first and the second types of analysis (plots at the top and at the bottom, respectively). The considered intervals in which triggering events' magnitudes fall are the following. First analysis, plot at the top: $[1.8, 2.1]$, $[2.2, 2.6]$, $[3.3, 3.8]$ and $[3.9, 5.9]$; second analysis, plot at the bottom: $[1.8, 2.2]$, $[2.5, 3.2]$, $[3.6, 4.6]$ and $[4.9, 5.9]$ (in both cases, the curves are red, black, blue and magenta, respectively). The δ^* value is equal to one day. The optimal bandwidth value for the Normal kernel density estimation is, respectively for the four intervals considered, equal to: 0.22, 0.33, 0.28, 0.19 in the plot at the top and 0.25, 0.26, 0.31, 0.27 in the one at the bottom.

The means of the distances between the events of the pairs are about 7 Km and 13 Km for the first and the second types of analysis, respectively. From the first analysis (plot at the top), one can notice that the increase of the referential mothers' magnitude corresponds to a qualitative variation of the density in agreement with our hypothesis, indeed the density increases (decreases) for high (low) values of the magnitude when the mother event's magnitude increases (decreases). The results for the second analysis (plot at the bottom) show the same qualitative variations. The learning (precursory) period, chosen to estimate the parameters, ends at the time of the last event occurred on April the 5^{th}, 2009. We get

$$(\varpi, \kappa, c, a, p) = (0.304, 0.06, 0.104, 1.57, 1.39). \tag{3.4}$$

In order to test the goodness of fit of the model with the set of parameters used, we have performed the residual analysis for which, as explained in Subsection 2.3.3, it is considered the residual process obtained with the random time-change (2.26) in order to get a stationary Poisson process with rate one. The results of the tests, given in Tab. 3.1, show that the model slightly underestimates the number of target events.

Table 3.1: Results of the tests obtained with the residual analysis, concerning the catalog relative to the square centered at L'Aquila till 2012. The learning period ends at the time of the last event occurred on April the 5^{th}, 2009. The parameters are the ones obtained for this learning period.

Number of events expected by the model	6138.27
Number of events without the learning period	6229
Runs test	4.48E-001
Kolmogorov-Smirnov test	5.02E-002

Only the Runs test gives back a probability bigger than 5%. The fact that this is not true for the Kolmogorov-Smirnov test, even if the probability has a value very little lower than 5%, could be explained by the value of the completeness magnitude used for this catalog. As explained before, due to the very high number of events occurred during the days just after the strong shock on April the 6^{th}, 2009, several events of magnitude 1.8 and a bit more have not been recorded. This means that the real completeness value for these days is higher than 1.8, value estimated with ZMAP. The fact that there are some events not recorded implies that the random time-change (2.26) doesn't guarantee that the residual process is a Poisson with rate one. It follows the underestimation of the number of events and the fact that the standard exponential law is not well fitted by the interevents between transformed times at small values (see the left plot of Fig. 3.12). However, since the Runs test gives good results and considering that this real catalog has a very strong shock, that implies an unusual amount of data to be recorded for the Italian case, we think that the set of parameters chosen can be considered good to be used in the analysis.

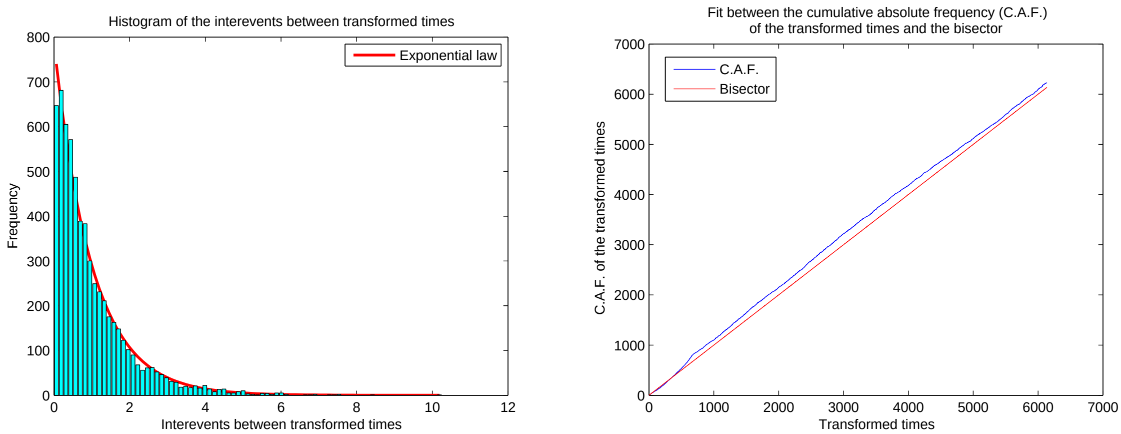
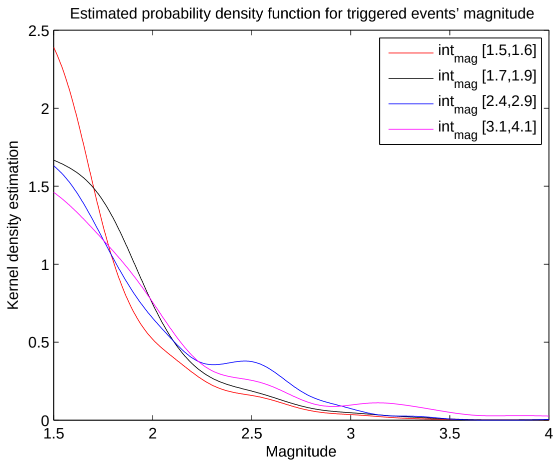


Figure 3.12: Plots concerning the second catalog, relative to L'Aquila till 2012. In the left plot the histogram of the interevent times of the transformed values is shown, together with the standard exponential law. The fit is not good for small values, as one can see from the non-agreement between the histogram and the red exponential law. In the right plot one can see how much the cumulative distribution function of the transformed times varies with respect to the bisector. There's not a really good agreement. This fact can be explained by the fact that the value 1.8 estimated for the completeness magnitude is probably too low for the days just after the strong shock present in the catalog considered.

In Fig. 3.13, one can see the results for the third catalog. In this case the temporal period is shorter than for the second one and ends the day before the strong shock on April the 6^{th}, 2009. The means of the distances are 23 Km and 17 Km for the first and the second types of analysis, respectively. The behaviors of the estimated densities of the triggered events' magnitude are qualitatively very close to those obtained for L'Aquila till 2012 in Fig. 3.11. This shows that our hypothesis of dependence is not related to the presence of a strong shock. Furthermore, we can conclude according to [Lippiello et al., 2012], that our hypothesis is not even connected to the incompleteness of the catalog, as instead proposed by [Corral, 2006]. In fact, we obtain evidences not only when we consider a catalog with a higher completeness magnitude m_c and with a strong shock that may cause problems with completeness in the days just after it. The evidences are obtained also when considering a catalog with a low m_c and a law seismic activity, then without any kind of incompleteness problem.



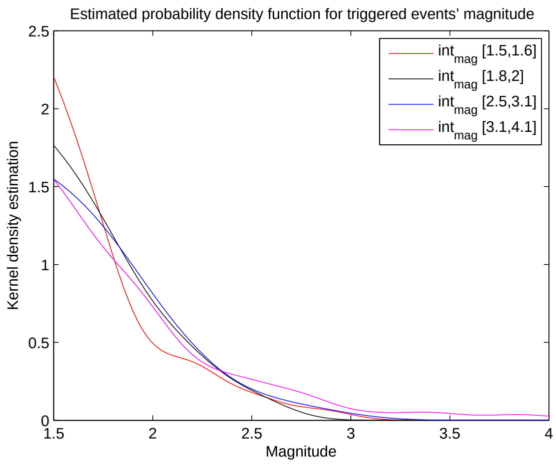


Figure 3.13: Kernel density estimation of triggered events' magnitude in the third catalog (L'Aquila till April the 5^{th}, 2009), concerning the first and the second types of analysis (plots at the top and at the bottom, respectively). The considered intervals in which triggering events' magnitudes fall are the following. First analysis, plot at the top: $[1.5, 1.6]$, $[1.7, 1.9]$, $[2.4, 2.9]$ and $[3.1, 4.1]$; second analysis, plot at the bottom: $[1.5, 1.6]$, $[1.8, 2]$, $[2.5, 3.1]$ and $[3.1, 4.1]$ (in both cases, the curves are red, black, blue and magenta, respectively). The δ^* value is equal to four days. The optimal bandwidth value for the Normal kernel density estimation is, respectively for the four intervals considered, equal to: 0.11, 0.14, 0.11, 0.15 in the plot at the top and 0.14, 0.16, 0.22, 0.15 in the one at the bottom.

The parameters used here are the averages over all the sets of parameters obtained by setting the learning period to 7%, 8%, ..., 20% (see Tab. 3.2).

Table 3.2: List of parameters obtained for the third catalog, concerning the square centered at L'Aquila till April the 5^{th}, 2009, by varying the precursory period. The values are rounded to the significant decimal places. The set of initial parameters, given in input, is $(\varpi, \kappa, c, a, p) = (0.13326, 0.05636, 0.0042524, 0.50173, 0.86843)$. If we change the initial values of p and ϖ, or the order of magnitudes of c and κ, the values obtained are generally very similar.

Input data Learning %	Output parameters (ϖ,κ,c,a,p)
7%	(0.54,0.024,0.014,1.64,1.08)
8%	(0.55,0.023,0.014,1.64,1.09)
9%	(0.52,0.023,0.012,1.7,1.06)
10%	(0.52,0.023,0.013,1.68,1.06)
11%	(0.53,0.024,0.013,1.66,1.07)
12%	(0.57,0.022,0.015,1.66,1.1)
13%	(0.6,0.022,0.016,1.65,1.12)
14%	(0.64,0.02,0.018,1.65,1.17)
15%	(0.63,0.021,0.017,1.65,1.16)
16%	(0.62,0.021,0.016,1.64,1.15)
17%	(0.64,0.021,0.017,1.63,1.16)
18%	(0.65,0.020,0.017,1.64,1.17)
19%	(0.62,0.021,0.015,1.64,1.13)
20%	(0.62,0.021,0.015,1.65,1.14)

This choice is due to the absence of a very strong shock in this catalog. We get

$$(\varpi, \kappa, c, a, p) = (0.5893, 0.0219, 0.0151, 1.6521, 1.1186). \tag{3.5}$$

The results obtained with the sets of parameters corresponding to the different precursory periods show small variations from the ones in (3.5). The residual analysis, performed for the residual process obtained with the random time-change (2.26), is shown in Tab. 3.3.

Table 3.3: Results of the tests obtained with the residual analysis, concerning the catalog relative to the square centered at L'Aquila till April the 5^{th}, 2009. The parameters considered are the mean ones. The precursory period is set at 13%, value at which we obtain the parameters closest to the mean ones.

Number of events expected by the model	1435.4
Number of events without the learning period	1466
Runs test	4.64E-003
Kolmogorov-Smirnov test	8.84E-001

We can see that the Runs test doesn't give a good result: the interevents between transformed times are not independent. This means that small values of interevent times follow small values. This could be due to the presence in the catalog of sequences near in time, but spatially quite separated, or of sequences not completely included in the area under examination, as can be seen in the seismicity map of Fig. 3.3. Instead, the Kolmogorov-Smirnov test gives back a good result, in fact in this case the probability is bigger that 5%. Concluding with the residual analysis, the number of expected events is very similar to that of the real target shocks. All these results are illustrated in Fig. 3.14. In the case of this catalog, in order to give a graphical support for the non-good result of the Runs test, this figure contains also the plot of the interevent transformed times in logarithmic scale. We can see for example a persistence of quite small interevent transformed times at about 250 days (red rectangle). However, since we have to recall that we are modeling a very complex phenomenon with a quite simple process, we can't expect that the results are always good for all the tests. Then, all things evaluated, we think that the model with the set of the above mean parameters can be considered good to fit the real case.

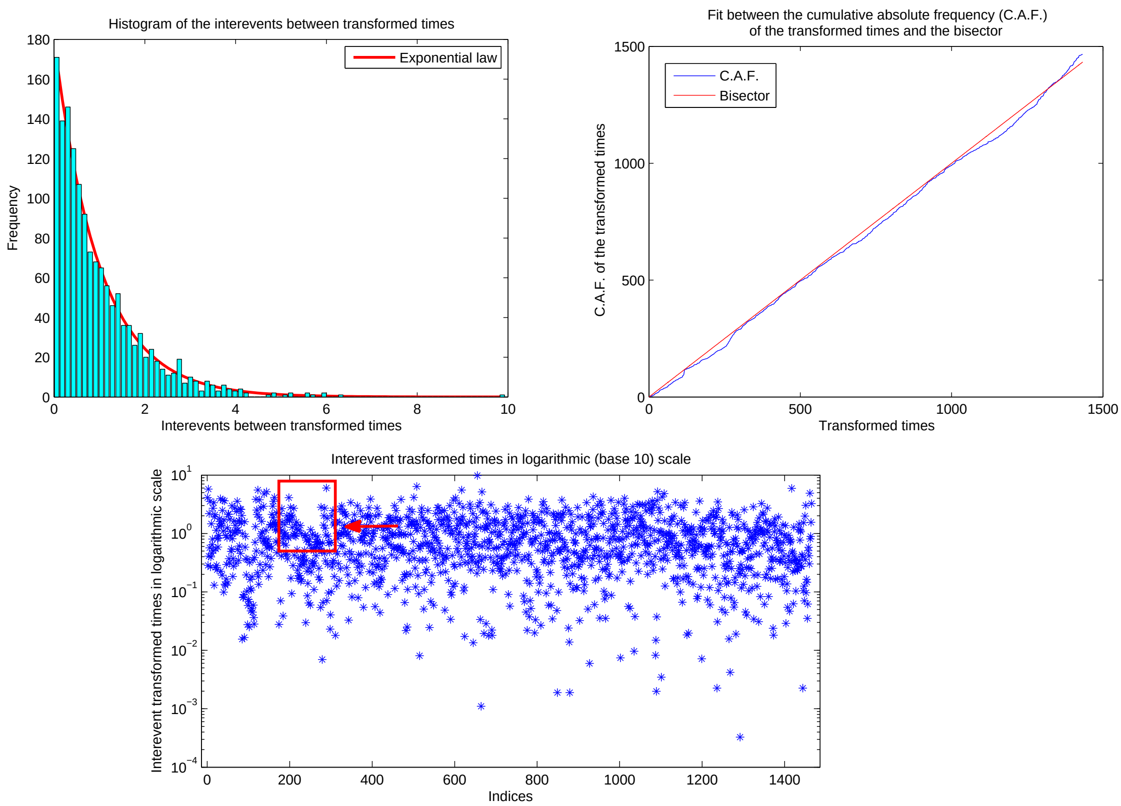
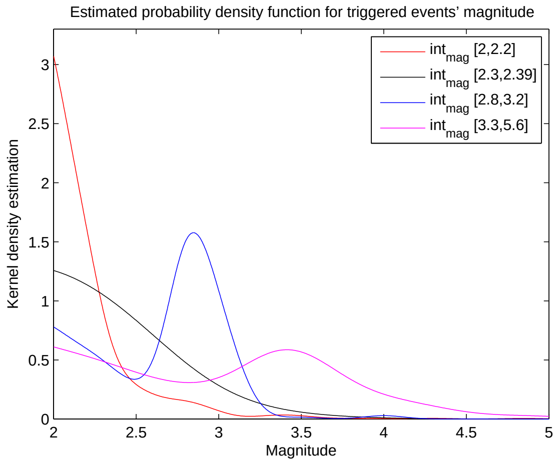


Figure 3.14: Plots concerning the third catalog, relative to L'Aquila till April the 5^{th}, 2009. In the left plot at the top, the histogram of the interevent times of the transformed values is shown, together with the standard exponential law. The fit is good. In the right plot at the top, one can see how much the cumulative distribution function of the transformed times varies with respect to the bisector. The deviation in the final part can be explained because that period dates back to the seismic swarm preceding the strong shock occurred on April the 6^{th}, 2009. At the beginning and in the central part, the non perfect agreement may be instead due to the presence of small clusters not included by the model. However, the fit is generally good, in fact so is the result of the Kolmogorov-Smirnov test. The plot at the bottom contains the interevent transformed times in logarithmic scale. In the red rectangle, indicated by the arrow, we can see a short persistent trend of small interevent times, as we expected by looking at the non-good result for the Runs test.

Concluding with the figures containing the estimated densities of triggered events' magnitude in the real catalogs, in Fig. 3.15 one can see the results relative to the fourth catalog, that is the Californian one.



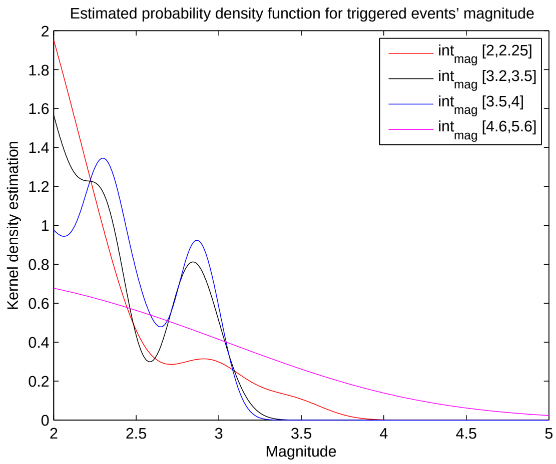


Figure 3.15: Kernel density estimation of triggered events' magnitude in the fourth catalog, concerning the first and the second types of analysis (plots at the top and at the bottom, respectively). The considered intervals in which triggering events' magnitudes fall are the following. First analysis, plot at the top: $[2, 2.2]$, $[2.3, 2.39]$, $[2.8, 3.2]$ and $[3.3, 5.6]$; second analysis, plot at the bottom: $[2, 2.25]$, $[3.2, 3.5]$, $[3.5, 4]$ and $[4.6, 5.6]$ (in both cases, the curves are red, black, blue and magenta, respectively). The δ^* value is equal to one day. The optimal bandwidth value for the Normal kernel density estimation is, respectively for the four intervals considered, equal to: 0.11, 0.44, 0.12, 0.25 in the plot at the top and 0.18, 0.13, 0.1, 1 in the one at the bottom.

Here, the mean of the distances is 13 Km for both the two types of analysis. Both from the plots at the top and the bottom, respectively obtained with the first and the second types of analysis, we get results in agreement with the above behaviors. It follows that, even if we analyze the events of a region in another continent, the hypothesis is still supported by the results of the two types of analysis. The parameters are again obtained by averaging over the sets estimated for a learning period fixed at 7%, 8%, ..., 20% (see Tab. 3.4).

Table 3.4: List of parameters obtained for the Southern Californian catalog by varying the precursory period. The values are rounded to the significant decimal places. The set of initial parameters, given in input, is $(\varpi, \kappa, c, a, p) = (0.13326, 0.05636, 0.0042524, 0.50173, 0.86843)$. If we change the initial values of p and ϖ, or the order of magnitudes of c and κ, the values obtained are generally very similar.

Input data Learning %	Output parameters (ϖ,κ,c,a,p)
7%	(0.37,0.012,0.0002,0.87,0.89)
8%	(0.37,0.011,0.0002,0.9,0.89)
9%	(0.39,0.011,0.0002,0.9,0.9)
10%	(0.39,0.011,0.0002,0.89,0.9)
11%	(0.38,0.011,0.0002,0.87,0.89)
12%	(0.38,0.011,0.0002,0.89,0.9)
13%	(0.39,0.011,0.00019,0.83,0.9)
14%	(0.39,0.011,0.00019,0.83,0.9)
15%	(0.37,0.012,0.00017,0.83,0.88)
16%	(0.37,0.012,0.00018,0.83,0.89)
17%	(0.37,0.012,0.00018,0.85,0.88)
18%	(0.36,0.012,0.00018,0.84,0.88)
19%	(0.35,0.013,0.00016,0.82,0.87)
20%	(0.34,0.013,0.0002,0.86,0.86)

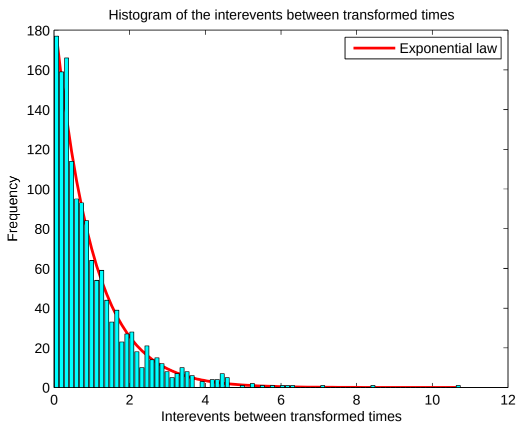
The catalog contains a quite strong shock of magnitude 5.6 too, but in the Californian region these events are more frequent and the technologies themselves are more suited to record these kind of earthquakes. Then, we decided to proceed with the research of the parameters as for the previous catalog. We get

$$(\varpi, \kappa, c, a, p) = (0.3729, 0.0116, 0.0002, 0.8579, 0.8879). \tag{3.6}$$

The values obtained for different learning periods are close to the mean values in (3.6). The results of the residual analysis are shown in Tab. 3.5; we recall again that this analysis considers for the residual process, obtained with the random time-change (2.26). Both the Kolmogorov-Smirnov and the Runs tests give back a probability bigger than 5%. The number of expected events is also close to the real number of target events. We deduce the goodness of fit of the model with the set of the above mean parameters (see Fig. 3.16).

Table 3.5: Results of the tests obtained with the residual analysis, concerning the Southern Californian catalog. The parameters considered are the mean ones. The precursory period is set at 7%, value at which we obtain parameters really very close to the mean ones.

Number of events expected by the model	1418.21
Number of events without the learning period	1428
Runs test	2.46E-001
Kolmogorov-Smirnov test	2.99E-001



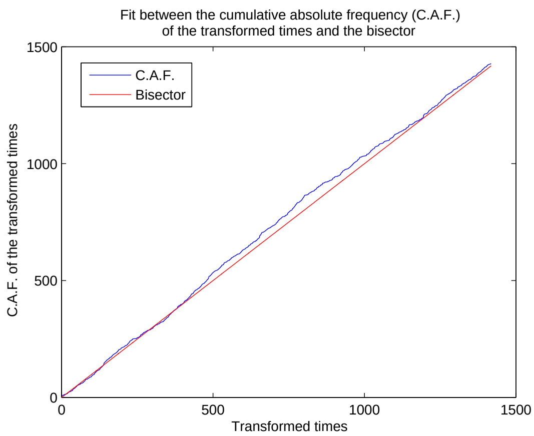


Figure 3.16: Plots concerning the fourth catalog, relative to Southern California. In the left plot the histogram of the interevent times of the transformed values is shown, together with the standard exponential law. The fit is good. In the right plot one can see how much the cumulative distribution function of the transformed times varies with respect to the bisector. The central part isn't in a really very good agreement, instead at the beginning and the end the lines are very close. This can be explained by the fact that on July the 8^{th}, 1986, a shock of magnitude 5.6 has occurred. Then, in the days just after, there may have been problems with the sequence, like for example a slight incompleteness of the catalog for magnitude 2. However, this comes under the variability of the process.

Before proceeding with the plots containing the means, we just would like to add that, as said before, the presence of a high magnitude event generally implies a very high number of aftershocks. This fact could cause two opposite behaviors. On the one hand, the dependence is more evident than for the catalog without a strong shock, due to the longer time-interval of consequent activity interested and to the larger number of events to be considered in the analysis. On the other hand, the very high number of events may induce errors when putting events in causal relation and this could affect the results of dependence. We suppose that the latter case concerns mainly the first analysis, in which the relations between events are established without an underlying model (with the exception of the catalog concerning L'Aquila till 2012, for which we perform the random time-change (2.26) based on the ETAS rate). These considerations may explain the small differences of behavior between the catalogs. Nevertheless, in our opinion it is important to notice that, even if the hypothesis of magnitude correlation is sometimes more, sometimes less evident, it is always supported by the results obtained.

Finally, let's consider the magnitude means. Figures 3.17 and 3.18 in the next two pages contain the plots of the averages of triggered events' magnitudes versus triggering events' magnitudes for the first and the second types of analysis, respectively. In each case, the four triggered magnitude averages are normalized by the averages of these four mean values. The results of the linear regression analysis and the error bars are also shown. The lengths of the latter are given by the normalized mean standard errors. Concerning Fig. 3.17, starting with the upper-left plot and proceeding clockwise, we illustrate the results obtained for the first, second, third and fourth catalogs, respectively relative to the whole Italy, L'Aquila till 2012, L'Aquila till April the 5^{th}, 2009 and the Southern California. In all the cases, with the exception of the first catalog as we expected, we can see a clearly increasing trend of the means, supporting our hypothesis of magnitude correlation. The no percentage variation in the plot of the Italian catalog are due to the presence here of many pairs of events that are close to each other along time but spatially very separated, as said before. The elements of these pairs are erroneously put in relation in the analysis. Concerning instead Fig. 3.18, the plots have the same order as in Fig. 3.17, but the one relative to Italian catalog is obviously absent. In fact, as already explained, this analysis has not been performed for this catalog. Again the means have an increasing trend in all of these three cases. The results obtained are statistically significant, as one can see in Tab. 3.6.

Table 3.6: List of correlation coefficients R and p-values p. Catalogs from 1 to 6 refer to the whole Italy, L'Aquila till 2012, L'Aquila till April the 5^{th}, 2009 and Southern California, respectively.

	First analysis		Second analysis	
	$R \simeq$	$p \simeq$	$R \simeq$	$p \simeq$
Catalog 1	0.88	0.11	/	/
Catalog 2	0.99	0.004	0.94	0.05
Catalog 3	0.96	0.03	0.94	0.05
Catalog 4	0.99	0.0005	0.95	0.04

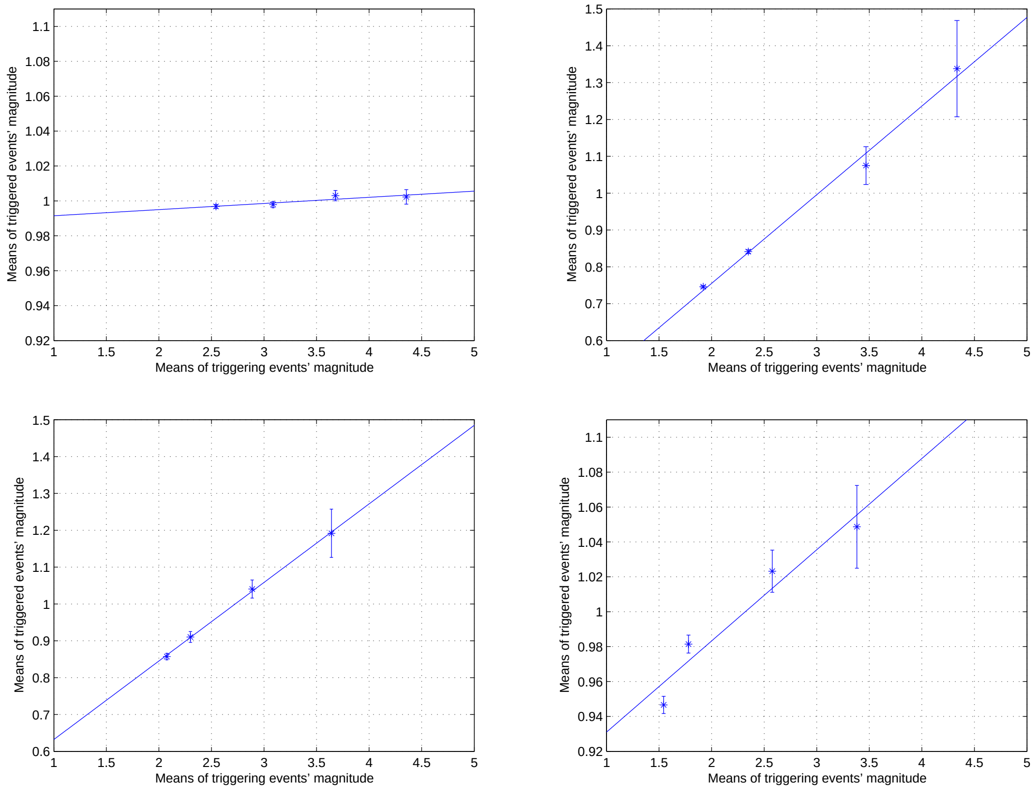


Figure 3.17: Averages of the normalized triggered events' magnitudes obtained with the first analysis. Starting with the upper-left plot and proceeding clockwise, the plots concern the first, second, third and fourth catalogs, respectively relative to the whole Italy, L'Aquila till 2012, L'Aquila till April the 5^{th}, 2009 and the Southern California. The percentage means for the four subintervals considered for the above four catalogs are the following. Catalog one: 0.9968, 0.9979, 1.0030, 1.0023 (corresponding to the triggering events' magnitude 2.5434, 3.0862, 3.6798, 4.3541, respectively); catalog two: 0.7460, 0.8413, 1.0748, 1.3379 (corresponding to the triggering events' magnitude 1.9199, 2.3491, 3.4687, 4.3342, respectively); catalog three: 0.9466, 0.9814, 1.0233, 1.0487 (corresponding to the triggering events' magnitude 1.5433, 1.7796, 2.5770, 3.3818, respectively); catalog four: 0.8572, 0.9104, 1.0406, 1.1918 (corresponding to the triggering events' magnitude 2.0754, 2.3000, 2.8876, 3.6415, respectively). The continuous lines correspond to the results of the linear regression and the semi-amplitude of the error bars are the normalized mean standard errors.

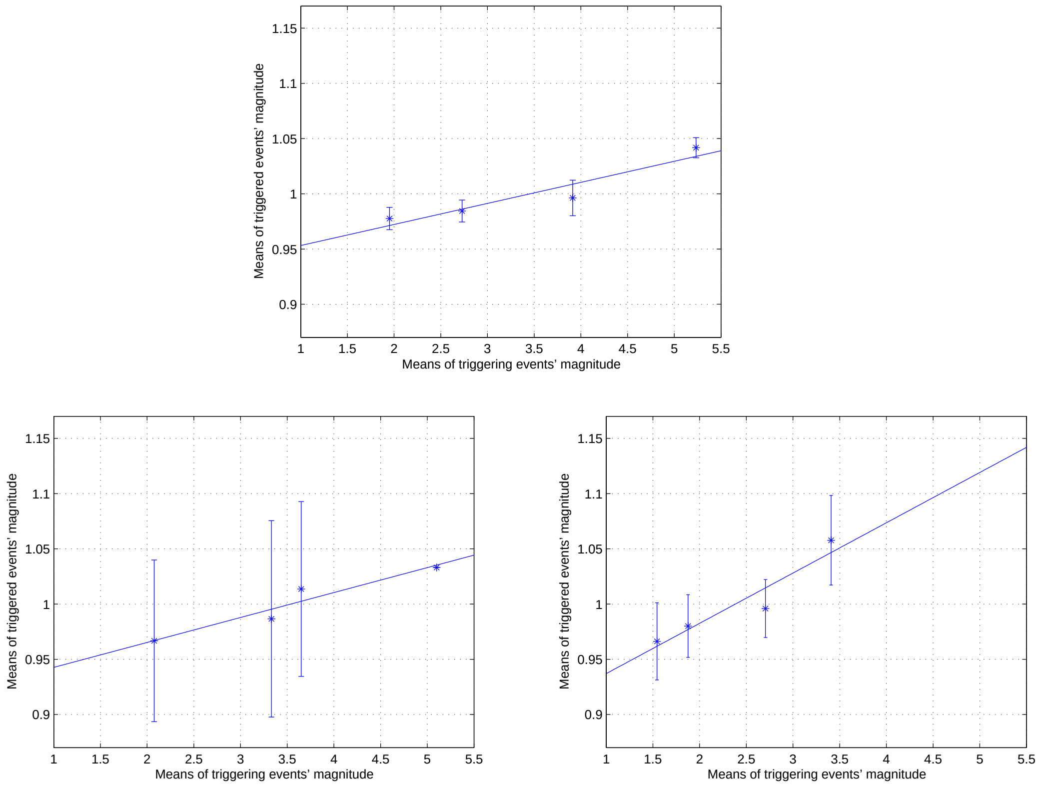


Figure 3.18: Averages of the normalized triggered events' magnitudes obtained with the second analysis. Starting with the plot at the top and proceeding clockwise, the plots concern the second, third and fourth catalogs, respectively relative to L'Aquila till 2012, L'Aquila till April the 5^{th}, 2009 and the Southern California. The percentage means for the four subintervals considered for the above three catalogs are the following. Catalog two: 0.9776, 0.9844, 0.9963, 1.0417 (corresponding to the triggering events' magnitude 1.9508, 2.7277, 3.9125, 5.2333, respectively); catalog three: 0.9662, 0.9801, 0.9959, 1.0578 (corresponding to the triggering events' magnitude 1.5433, 1.8763, 2.7056, 3.4083, respectively); catalog four: 0.9667, 0.9866, 1.0136, 1.0331 (corresponding to the triggering events' magnitude 2.0753, 3.3309, 3.6500, 5.1000, respectively). The continuous lines correspond to the results of the linear regression and the semi-amplitude of the error bars are the normalized mean standard errors.

3.4 Conclusions

Although the magnitudes of all the events are distributed according to the Gutenberg-Richter law when not taking into account the characteristics of past seismicity, the results obtained in this chapter with the two types of analysis above described show that the probability density of triggered events' magnitude changes when the mother events' magnitude changes. This supports the hypothesis of considering a new model, that is a variation of the classical temporal ETAS, in which the law of the triggered events' magnitudes depends on the triggering events' magnitude, conditionally on knowing its value.

The results of this chapter can be precisely summarized as follows.

Pr.1) The probability of "high" aftershocks magnitude increases with the mother event's magnitude. Furthermore, the triggered events' magnitude averages have an increasing trend, again with respect to the mother events' magnitudes. This is true for all the catalogs considered with the exception of the whole Italian catalog. The absence of the variation in this catalog is due to the fact that it contains many pairs of events temporally close to each other, but spatially very separated.

Pr.2) The magnitudes are not independent from each other, but there exists a transition probability density for the triggered events' magnitudes. Moreover, it has to change in shape with the triggering events' magnitude: when the latter increases, it may have a relative maximum and the higher the mother event's magnitude,

- the higher is the relative maximum,
- the higher is the expected value.

Chapter 4

The Epidemic Type Aftershock Sequence model with correlated magnitudes

In this chapter, we focus on our new version of the ETAS model with correlated triggered events' magnitudes.

In Section 4.1 we write explicitly the conditional intensity function characterizing the above new ETAS model. This function is then analyzed in order to give a precise description of the process, for which we also derive the conditions for the non-explosion. We will see that these conditions remain the same as for the classical ETAS.

In Section 4.2 we then obtain the Laplace functional for this model: it is a very useful tool for the study of several characteristics and properties of the model itself, allowing to leave also open questions that could be analyzed in future. In the first part of this section the results obtained are general, while in the last part we derive the explicit form relative to the precise functions of which consists the conditional intensity function written in the previous Section 4.1.

Finally, in the last Section 4.3, we find an explicit form of the probability density function $p(\cdot|m')$ for the triggered events' magnitude, depending on the magnitude m' of the corresponding mother event. The results obtained in the previous chapter are very useful for this purpose: they allow to understand which kind of transition function the $p(\cdot|m')$ should be. More precisely, we have to search for a function such that, when the mother events' magnitude m' increases, the probability of having events with high (low) magnitudes must increase (decrease). In addition, for high values of m', the function may also have a relative maximum.

4.1 The conditional intensity of the new ETAS model with correlated magnitudes

The main result of this thesis consists of the derivation of our new version of the temporal ETAS model: the Epidemic Type Aftershock Sequence model with correlated magnitudes. As already anticipated in Chapter 2 (see (2.9)), it is a variation of the temporal ETAS model. In fact, in our new model we assume that the probability density function relative to the triggered events' magnitude is a transition probability density depending on the magnitude of the mother events. It is nothing but the general ETAS model, characterized by the conditional intensity (2.3), but without considering the spatial locations of the events, i.e., (2.9).

To rigorously define our new model, let's give then its conditional intensity function:

$$\lambda(t, m|\mathcal{H}_t) = \varpi p(m) + \sum_{\{i|t_i<t\}} \varrho(m_i)\Phi(t-t_i)p(m|m_i), \tag{4.1}$$

where $p(m|m_i)$ is a transition probability density.

As we are going to see in the Section 4.3, we will take the transition probability density $p(\cdot|\cdot)$ depending on the parameters β, a and a new parameter $C_1 \in [0,1]$ as follows:

$$p(m|m_i) = p(m)\Big[1 + C_1\big(1 - 2e^{-(\beta-a)(m_i-m_0)}\big)\big(1 - 2e^{-\beta(m-m_0)}\big)\Big]. \tag{4.2}$$

Then, recalling also the explicit forms of the functions $\varrho(\cdot)$ and $\Phi(\cdot)$ (equations (2.1) and (2.5)), the conditional intensity function (4.1) can be rewritten as

$$\begin{aligned}\lambda(t, m|\mathcal{H}_t) = &p(m)\Big\{\varpi + \sum_{\{i|t_i<t\}} \kappa e^{a(m_i-m_0)}\frac{p-1}{c}\Big(1+\frac{t-t_i}{c}\Big)^{-p}\Big[1 \\ &+ C_1\big(1 - 2e^{-(\beta-a)(m_i-m_0)}\big)\big(1 - 2e^{-\beta(m-m_0)}\big)\Big]\Big\},\end{aligned} \tag{4.3}$$

where $0 \leq C_1 \leq 1$.

It is important to note that, clearly, the above ETAS model with correlated magnitudes is a direct generalization of the classical ETAS model with independent magnitudes: when $C_1 = 0$, the transition probability density in (4.2) reduces to $p(m)$ and we get the model (2.7) in Chapter 2.

By looking at the conditional intensity function completely characterizing the new model we are proposing (the above equation (4.3)), we can do the following considerations. The background component of the process is modeled by a homogeneous Poisson process with constant rate ϖ. The magnitudes of this component are described by the well-known Gutenberg-Richter law. Then, each event of the seismic sequence may give birth to its own cluster of aftershocks, described by a time-stationary, non-homogeneous Poisson process: the rate of the triggered process depends on times and magnitudes and the time dependence is given only by the distances between the times of occurrence of mother/daughter events. More precisely, the rate relative to the aftershocks component consists of:

- the normalized Omori-Utsu law
$$\Phi(t) = \frac{p-1}{c}\Big(1 + \frac{t}{c}\Big)^{-p}, \quad t > 0,$$
for the time decay;
- the productivity law
$$\varrho(m') = \kappa e^{a(m'-m_0)}, \quad m' \geq m_0,$$
for the fertility of the generating event;
- a new law, given in (4.2), for the magnitudes of the triggered events. It is a transition probability function depending on the magnitude of the triggering events, just like the mean number of aftershocks per mother event does. Let's notice also that the transition probability $p(\cdot|\cdot)$ is proportional to the Gutenberg-Richter law $p(m)$ with decay parameter β.

As already specified, the explicit form (4.2) of the above transition probability will be obtained in Section 4.3. Let's anticipate here just that it will be derived in accordance with the experimental results of Chapter 3 and imposing some further conditions it has to verify. Among these, the most important one is that, if we average over all the possible triggering events' magnitudes, that is if we don't take into account the characteristics of past seismicity, the $p(\cdot|m')$ becomes again the Gutenberg-Richter law, well supported by many experimental evidences. This condition is expressed in formula (4.28) (with (4.26)) of Section 4.3. Actually, if we write condition (4.28) in the equivalent form
$$\int_{m_0}^{\infty} p(m')\varrho(m')p(m''|m')dm' = \eta p(m''),$$
(see equation (4.29) in Section 4.3), then we can deduce that $p(m)$ is an eigenfunction corresponding to the eigenvalue $\eta = \kappa\frac{\beta}{\beta-a}$, as we are going

to see in Lemma 1 below when considering the case of the operator $\mathcal{K}_2$ (see (4.6), (4.7) and (4.8)). More precisely, Lemma 1 gives the conditions for the non-explosion of our process, that are obtained to be exactly the same as for the classical time-magnitude ETAS, i.e., in the case of independent magnitudes (see (2.19)). Furthermore, the above-mentioned Lemma 1 clearly shows that η is the maximum eigenvalue and then one can apply Zhuang's method of Subsection 2.2.1 in Chapter 2 (see also [Zhuang, 2002]).

Summarizing, the new model proposed is derived through several intermediate requests, both theoretical and experimental, which allow to describe the seismic sequences in a more realistic way. Our model improves the already well constructed time-magnitude ETAS by considering the natural hypothesis of correlated magnitudes. This new condition about the magnitudes doesn't complicate the statistical and theoretical analysis so much. In fact, besides the previous parameters $(\varpi, \kappa, c, a, p, \beta, m_0)$, there is only one more parameter to be considered. It is the constant C_1 in (4.3), to be chosen non-negative and less than or equal to one. Finally, by setting the classical conditions for the non-explosion of the time-magnitude ETAS (see (2.19)), that are $\beta > a$, $p > 1$ and $\kappa < \frac{\beta - a}{\beta}$ for (p, β, a, κ) the parameters of the Gutenberg-Richter, the Omori and the productivity laws, our new model is non-explosive, and then it represents a reasonable model for earthquakes.

Now, before analyzing the properties of the model and explaining the reasons for our choice of $p(\cdot|m')$, let's state the just-mentioned Lemma 1 for the non-explosion: we will give here the enunciate and a sketch of the proof, referring to Appendix B for the complete proof.

Lemma 1. *Let's assume that the parameters* (p, β, a, κ) *of the ETAS model with correlated magnitudes, completely characterized by intensity* (4.1) *and transition probability* (4.2)*, satisfy the conditions*

$$p > 1, \quad \beta > a, \quad \kappa \frac{\beta}{\beta - a} < 1. \tag{4.4}$$

Then, the point process is non-explosive, whatever the value of the correlation parameter $C_1 \in [0, 1]$ *is.*

Sketch of the proof. In order to prove the non-explosion of the ETAS model with correlated magnitudes, we have to show that, given the left and right eigenfunctions $\ell_1(m')$ and $\ell_2(m)$ of the rate for the triggered component of the sequence, the corresponding maximum eigenvalue $\tilde{\eta}$ is strictly less than one. We have then to consider the equations

$$\tilde{\eta}\ell_1(m') = \int_{m_0}^{\infty} \ell_1(m)\varrho(m')p(m|m')dm,$$
$$\tilde{\eta}\ell_2(m) = \int_{m_0}^{\infty} \varrho(m')p(m|m')\ell_2(m')dm'$$

as in Subsection 2.2.1 of Chapter 2, that have been obtained considering (2.13) and (2.14) and computing the integrals with respect to space and time. At first, let's notice that, as for the classical ETAS, the condition for the temporal integral to be finite is $p > 1$, that is exactly the first condition in (4.4) for the non-explosion.

We can now introduce the two operators

$$\mathcal{K}_1\ell(m') = \int_{m_0}^{\infty} \ell(m)\varrho(m')p(m|m')dm \tag{4.5}$$

and

$$\mathcal{K}_2\ell(m) = \int_{m_0}^{\infty} \varrho(m')p(m|m')\ell(m')dm'. \tag{4.6}$$

At first let's say that, substituting the expression (4.2) for the transition probability in the above two formulas, we find that the second condition $\beta > a$ in (4.4) has to be satisfied in order to get finite integrals. Furthermore, for both the two operators, we find that the eigenvalues are

$$\begin{cases} \tilde{\eta}^{(1)} & = \eta_a(\beta)\, C_1\, \frac{2\beta(\beta-a)}{(2\beta-a)(3\beta-2a)} \\ \\ \tilde{\eta}^{(2)} & = \eta_a(\beta), \end{cases} \tag{4.7}$$

where

$$\eta_a(\beta) = \eta = \kappa\frac{\beta}{\beta - a}.$$

It follows that, in order to have $\tilde{\eta}^{(1)} < \tilde{\eta}^{(2)} < 1$, we need to impose the last condition for the non-explosion of the process in (4.4), that is $\kappa\frac{\beta}{\beta-a} < 1$. Now, setting

$$\varrho_\vartheta(m) := \kappa e^{\vartheta(m-m_0)} \qquad \text{and} \qquad p_\vartheta(m) := \vartheta e^{-\vartheta(m-m_0)},$$

the row eigenvectors relative to the above eigenvalues are proportional to:

- in the case of the operator $\mathcal{K}_1$,

$$\begin{cases} \ell_1^{(1)}(m') & \propto \varrho_a(m') - 2\varrho_{2a-\beta}(m') \\ \\ \ell_1^{(2)}(m') & \propto \varrho_a(m') - 2\frac{C_1 a(3\beta-a)}{(2\beta-a)[3\beta-2a+C_1(2a-\beta)]}\varrho_{2a-\beta}(m'); \end{cases}$$

- in the case of the operator $\mathcal{K}_2$,

$$\begin{cases} \ell_2^{(1)}(m) & \propto p_\beta(m) - [1 - C_1\,\frac{\beta}{3\beta-2a}]\kappa\,\frac{2\beta}{2\beta-a}\,p_{2\beta}(m) \\ \\ \ell_2^{(2)}(m) & \propto p_\beta(m). \end{cases} \tag{4.8}$$

Finally, if $N_i^{tr}(\mathbb{R}^2)$ is the number of all the triggered events generated by the spontaneous shock $(S_i = s, M_i = m)$, with $N_i^{tr,n}(\mathbb{R}^2)$ indicating its n^{th} generation, we find that

$$\begin{aligned} \mathbb{E}\Big[N_i^{tr,n}(\mathbb{R}^2)\Big|S_i = s, M_i = m\Big] =& \mathcal{K}_1^{(n)}\mathbf{1}(m) \\ =& \mathcal{K}_1^{(n-1)}\varrho(m), \end{aligned}$$

where $\mathbf{1}(m) = \mathbb{1}_{[m_0,\infty)}(m)$. Then, considering the evolution of the equation

$$\ell(m) = \vartheta_1\varrho_a(m) + \vartheta_2\varrho_{2a-\beta}(m)$$

after n steps, i.e.,

$$\mathcal{K}_1^{(n)}\ell(m) = \vartheta_1^n\varrho_a(m) + \vartheta_2^n\varrho_{2a-\beta}(m)$$

and setting

$$x_1 = C_1\frac{a}{2\beta - a}, \qquad y_1 = C_1\frac{2a-\beta}{3\beta-2a}, \qquad z_1 = x_1 - y_1,$$

we get

$$\begin{pmatrix} \vartheta_1^n \\ \\ \vartheta_2^n \end{pmatrix} = \begin{pmatrix} \frac{1+y_1}{1-z_1}\left[\tilde{\eta}^{(2)}\right]^{n-1} - \frac{x_1}{1-z_1}\left[\tilde{\eta}^{(1)}\right]^{n-1} \\ \\ \frac{-2x_1}{1-z_1}\left[\tilde{\eta}^{(2)}\right]^{n-1} + \frac{2x_1}{1-z_1}\left[\tilde{\eta}^{(1)}\right]^{n-1} \end{pmatrix}.$$

Since

$$\begin{aligned} \mathbb{E}\Big[N_i^{tr}(\mathbb{R}^2)\Big|S_i = s, M_i = m\Big] &= \mathbb{E}\left[\sum_{n=1}^{\infty} N_i^{tr,n}(\mathbb{R}^2)\Big|S_i = s, M_i = m\right] \\ &= \varrho_a(m)\gamma_1 + \varrho_{2a-\beta}(m)\gamma_2 < \infty, \end{aligned}$$

where

$$\gamma_1 = \sum_{n=1}^{\infty}\left(\frac{1+y_1}{1-z_1}\left[\tilde{\eta}^{(2)}\right]^{n-1} - \frac{x_1}{1-z_1}\left[\tilde{\eta}^{(1)}\right]^{n-1}\right),$$
$$\gamma_2 = \sum_{n=1}^{\infty}\left(\frac{-2x_1}{1-z_1}\left[\tilde{\eta}^{(2)}\right]^{n-1} + \frac{2x_1}{1-z_1}\left[\tilde{\eta}^{(1)}\right]^{n-1}\right)$$

and

$$\begin{aligned}
\mathbb{E}\Big[N_i^{tr}(\mathbb{R}^2)\Big|S_i = s\Big] &= \int_{m_0}^{\infty}\left[\gamma_1\varrho_a(m) + \gamma_2\varrho_{2a-\beta}(m)\right]p_\beta(m)dm \\
&= \gamma_1\int_{m_0}^{\infty}\varrho_a(m)p_\beta(m)dm + \gamma_2\int_{m_0}^{\infty}\varrho_{2a-\beta}(m)p_\beta(m)dm \\
&= \eta\left[\gamma_1 + \frac{\gamma_2}{2}\right] < \infty,
\end{aligned}$$

we deduce that the mean cluster size is finite and then, recalling Remark 3 in Chapter 1, the resulting process does not explode. □

Remark 6. The same result on the finite mean cluster size can be obtained by integrating with respect to the spontaneous event's magnitude m already when computing the expected values $\mathbb{E}\Big[N_i^{tr,n}(\mathbb{R}^2)\Big|S_i = s, M_i = m\Big]$. More precisely, switching the order of integration where necessary and recalling that

$$\int_{m_0}^{\infty} p(m)\varrho(m)p(m'|m)dm = \mathcal{K}_2 p(m') = \eta p(m'),$$

one obtains

$$\begin{aligned}
\mathbb{E}\Big[N_i^{tr,1}(\mathbb{R}^2)|S_i = s\Big] &= \int_{m_0}^{\infty} dm p(m)\mathbb{E}\Big[N_i^{tr,1}(\mathbb{R}^2)\Big|S_i = s, M_i = m\Big] \\
&= \int_{m_0}^{\infty} dm p(m)\int_s^{\infty} dt\int_{m_0}^{\infty}\Phi(t-s)\varrho(m)p(m_1|m)dm_1 \\
&= \int_{m_0}^{\infty} dm_1\int_{m_0}^{\infty} p(m)\varrho(m)p(m_1|m)dm \\
&= \int_{m_0}^{\infty} dm_1\mathcal{K}_2 p(m_1) \\
&= \eta,
\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}\Big[N_i^{tr,2}(\mathbb{R}^2)|S_i=s\Big] &= \int_{m_0}^{\infty} dmp(m)\int_s^{\infty} dt_1 \int_{m_0}^{\infty} dm_1 \int_{t_1}^{\infty} dt_2 \int_{m_0}^{\infty} \Phi(t_1-s) \\
&\cdot \Phi(t_2-t_1)\varrho(m)p(m_1|m)\varrho(m_1)p(m_2|m_1)dm_2 \\
&= \int_{m_0}^{\infty} dmp(m)\int_{m_0}^{\infty} dm_1 \int_{m_0}^{\infty} dm_2 \varrho(m)p(m_1|m)\varrho(m_1)p(m_2|m_1) \\
&= \int_{m_0}^{\infty} dm_2 \int_{m_0}^{\infty} dm_1 \varrho(m_1)p(m_2|m_1)\mathcal{K}_2 p(m_1) \\
&= \int_{m_0}^{\infty} dm_2 \mathcal{K}_2^{(2)} p(m_2) \\
&= \eta^2
\end{aligned}
$$

and then

$$
\mathbb{E}\Big[N_i^{tr,n}(\mathbb{R}^2)|S_i=s\Big] = \eta^n.
$$

Nevertheless, the method we use in the proof of the previous lemma allows to show that the process doesn't explode even if the background events' magnitudes are not distributed according to the Gutenberg-Richter law, but with another law with a density $q(\cdot)$ such that the integrals of $\varrho_a(m)$ and $\varrho_{2a-\beta}(m)$, with respect to $q(m)$, are finite.

Independently of us, Roueff *et al.* analyzed similarly the non-explosion problem for a class of locally stationary Hawkes processes [Roueff et al., 2015]. Nevertheless, they present a condition for the non-explosion (Theorem 1, page 7 in their manuscript) that is not satisfied by our magnitude correlated model (4.3). More precisely, they suppose that

$$
\sup_{(s,m)} E\big[N_i^{tr,1}(\mathbb{R}^2)|S_i=s, M_i=m\big] < 1, \tag{4.9}
$$

where we recall that $N_i^{tr,1}(dt,dm)$ represents the marked point process of first generation aftershocks of the spontaneous event $(S_i=s, M_i=m)$. In our case, it holds

$$
\sup_{(s,m)} E\big[N_i^{tr,1}(\mathbb{R}^2)|S_i=s, M_i=m\big] = \sup_m \varrho(m) = \infty,
$$

and then Roueff's condition, written above in (4.9), is not satisfied. For our model (4.3) it holds instead

$$
E[N_i^{tr,1}(\mathbb{R}^2)] = \eta < 1,
$$

for each i.

Concluding, we want to discuss about the moment properties in the case of the ETAS model with correlated triggered events' magnitudes. The equations (2.20) and (2.21), derived in Subsection 2.2.2 of Chapter 2, are still valid because they have been obtained for a general function $h(\cdot|\cdot)$ representing the rate relative to the triggered component of the process, that is for the general space-time-magnitude ETAS, to which our new model belongs. The only thing to discuss is the branching ratio. In order to get the latter value for the ETAS model with correlated magnitudes, we start with equation (2.22) of the above-mentioned subsection, i.e.,

$$\begin{aligned} M_1(x, y, m) =& \varpi(x, y)p(m) + \int_{\mathfrak{S}} \varrho(m'')\Phi(t - t'')f(x - x'', y - y''|m'') \\ & \cdot p(m|m'')M_1(x'', y'', m'')d\bar{v}'', \end{aligned}$$

where

$$\mathfrak{S} = \mathbb{R}^2 \times (-\infty, +\infty) \times [m_0, \infty).$$

We recall that this is an equation for the first order moment $M_1(x, y, m)$, again in the case of the general space-time-magnitude ETAS. In our case, we are discarding the spatial component and then the above result becomes

$$M_1(m) =\varpi p(m) + \int_{-\infty}^{\infty}\int_{m_0}^{\infty} \varrho(m'')\Phi(t - t'')p(m|m'')M_1(m'')d\bar{v}''.$$

As in Subsection 2.2.2, we take the expectation with respect to time, obtaining the relationship between the total magnitude distribution $j_{tot}(\cdot)$, the distribution of the background events' magnitude $p(\cdot)$ and the one of the aftershocks' magnitude $p(\cdot|\cdot)$:

$$j_{tot}(m) = \frac{\varpi p(m)}{\overline{m}} + \int_{m_0}^{\infty} \varrho(m'')p(m|m'')j_{tot}(m'')dm'', \tag{4.10}$$

where

$$\overline{m} = \int_{m_0}^{\infty} M_1(m)dm.$$

Now, as already said in the current section, we are supposing that

$$\int_{m_0}^{\infty} p(m'')\varrho(m'')p(m|m'')dm'' = \eta p(m),$$

meaning that a generic event's magnitude, whatever generation it belongs to and without any other information about the magnitude of the mother event, is distributed according to the Gutenberg-Richter law. It follows that, in the case of our new model, the total magnitude distribution $j_{tot}(m)$ is exactly the Gutenberg-Richter $p(m)$. Then, equation (4.10) becomes

$$p(m) = \frac{\varpi p(m)}{\overline{m}} + \int_{m_0}^{\infty} \varrho(m'')p(m|m'')p(m'')dm'' = \frac{\varpi p(m)}{\overline{m}} + \eta p(m);$$

we obtain again equation (2.24) of Subsection 2.2.2 and the branching ratio remains equal to the one derived for the classical ETAS.

4.2 The Laplace functional

In this section we derive the Laplace functional relative to the ETAS model with correlated magnitudes. More precisely, we will consider the general Hawkes process with non-independent marks characterized by the conditional intensity function defined in (4.1), but we will not use the explicit form (4.2). Furthermore, we will use the fact that the background process is modeled as a homogeneous Poisson.

The analysis of the Laplace functional we propose in the current section (and subsection) is interesting, since the particular choice of our model has not been treated before in the literature to obtain the relative Laplace functional.

Let's start then by considering the random number $N(\omega, dt, dm)$ of all the events in the process. If $N^{sp}(\omega, dt, dm)$ is the total number of spontaneous events and $N^{tr}(\omega, dt, dm)$ is the total number of the triggered ones, it obviously holds

$$N(\omega, dt, dm) = N^{sp}(\omega, dt, dm) + N^{tr}(\omega, dt, dm), \tag{4.11}$$

where

$$N^{sp}(\omega, dt, dm) = \sum_i \delta_{(S_i, M_i)}$$

and

$$N^{tr}(\omega, dt', dm') = \sum_{i=1}^{N^{sp}} N_i^{tr}(\omega, dt', dm'|S_i, M_i).$$

The random number $N^{sp}(\omega, dt, dm)$ is a homogeneous Poisson process with rate $\varpi p(m)dtdm$, while all the variables $N_i^{tr}(\omega, dt', dm'|S_i = s, M_i = m)$, representing the progenies generated by the spontaneous event occurred in $S_i = s$ and with magnitude $M_i = m$, are independent marked Hawkes processes with rate $\varrho(m)\Phi(t'-s)\mathbb{1}_{(s,\infty)}(t')p(m'|m)$. It follows that the evolution of each $N_i^{tr}(\omega, dt', dm'|S_i = s, M_i = m)$ is modeled again as a Hawkes process in which:

- the immigrants are the first generation aftershocks of the background event (S_i, M_i) and have intensity $\varrho(m)\Phi(t'-s)\mathbb{1}_{(s,\infty)}(t')p(m'|m)$,
- the triggered events are the aftershocks of the background event (S_i, M_i) belonging to the generations from the second one on, with intensity $\varrho(m')\Phi(t''-t')\mathbb{1}_{(t',\infty)}(t'')p(m''|m')$.

For the spontaneous events, we will set

$$S_0 := \sup\{t \le 0 : \ N(\{t\}) = 1\},$$
$$S_{-1} < S_0 \le 0 < S_1 < S_2 < \cdots < S_n < S_{n+1} \cdots$$

Now, setting

$$\begin{aligned} N(\varphi) : &= \int_{\mathbb{R}\times[m_0,\infty)} \varphi(t,m)N(\omega, dt, dm) \\ &= \int_{\mathbb{R}\times[m_0,\infty)} \varphi(t,m)N^{sp}(\omega, dt, dm) + \int_{\mathbb{R}\times[m_0,\infty)} \varphi(t,m)N^{tr}(\omega, dt, dm), \end{aligned}$$

the Laplace functional of $N(\omega, dt, dm)$ is defined, for $\varphi(t,m) \ge 0$, as

$$\mathcal{L}(\varphi) := \mathbb{E}\left[\exp\{-N(\varphi)\}\right]. \tag{4.12}$$

For simplicity of notation, form now on we will omit the dependence on ω in the random numbers.

Theorem 9. *The Laplace functional of the total number of events $N(dt, dm)$ in the process completely characterized by the conditional intensity* (4.1)*, is*

$$\mathcal{L}(\varphi) = \exp\left\{-\int_{\mathbb{R}}\int_{m_0}^{\infty} \left(1 - e^{-\varphi(s,m)+\log(\mathcal{L}^{tr}(\varphi|s,m))}\right)\varpi\, ds\, p(m)dm\right\} \tag{4.13}$$

$$= \exp\left\{-\int_{\mathbb{R}}\int_{m_0}^{\infty} \left(1 - e^{-\Theta(s,m)}\right)\varpi\, ds\, p(m)dm\right\}, \tag{4.14}$$

where

$$\begin{aligned} \Theta(s,m) =& \varphi(s,m) + \int_{s}^{\infty}\int_{m_0}^{\infty} \left(1 - e^{-\varphi(t',m')}\mathcal{L}^{tr}(\varphi|t',m')\right) \\ &\cdot \varrho(m)\Phi(t'-s)dt'p(m'|m)dm' \end{aligned} \tag{4.15}$$

and

$$\begin{aligned} \mathcal{L}^{tr}(\varphi|s,m) &:= \mathbb{E}\left[\exp\{-N_i^{tr}(\varphi|S_i,M_i)\}\Big|S_i = s, M_i = m\right] \\ &= \mathbb{E}\left[\exp\left\{-\int \varphi(t,m)N_i^{tr}(dt,dm|S_i,M_i)\right\}\Big|S_i = s, M_i = m\right] \end{aligned} \tag{4.16}$$

is the conditional Laplace functional of the number $N_i^{tr}(dt, dm|S_i, M_i)$ of triggered events generated by the spontaneous shock in $(S_i = s, M_i = m)$.

Remark 7. Let's notice that the results obtained in the above theorem doesn't take into account the explicit forms of $\varrho(\cdot)$, $\Phi(\cdot)$, $p(\cdot|\cdot)$ and then are valid in general.

Proof. Recalling (4.11), we have that

$$\begin{aligned}
\mathcal{L}(\varphi) &= \mathbb{E}\left[\exp\left\{-\int \varphi(t,m)N(dt,dm)\right\}\right] \\
&= \mathbb{E}\left[\exp\left\{-\int \varphi(t,m)N^{sp}(dt,dm) - \int \varphi(t,m)N^{tr}(dt,dm)\right\}\right] \\
&= \mathbb{E}\left[\exp\left\{-\int \varphi(t,m)N^{sp}(dt,dm)\right\}\exp\left\{-\int \varphi(t,m)N^{tr}(dt,dm)\right\}\right] \\
&= \mathbb{E}\left[\exp\left\{-\int \varphi(t,m)N^{sp}(dt,dm)\right\}\mathbb{E}\left[\exp\left\{-\int \varphi(t,m)N^{tr}(dt,dm)\right\}\Big|N^{sp}\right]\right] \\
&= \mathbb{E}\left[\exp\{-N^{sp}(\varphi)\}\mathbb{E}\left[\exp\{-N^{tr}(\varphi)\}\Big|N^{sp}\right]\right]. \qquad (4.17)
\end{aligned}$$

Let's focus on the internal expected value in the final line of the previous formula:

$$\begin{aligned}
\mathbb{E}\left[\exp\{-N^{tr}(\varphi)\}\Big|N^{sp}\right] &= \mathbb{E}\left[\exp\left\{-\int \varphi(t,m)N^{tr}(dt,dm)\right\}\Big|N^{sp}\right] \\
&= \mathbb{E}\left[\exp\left\{-\sum_{i=1}^{N^{sp}} N_i^{tr}(\varphi|S_i,M_i)\right\}\Big|N^{sp}\right] \\
&= \mathbb{E}\left[\prod_{i=1}^{N^{sp}}\exp\left\{-\int \varphi(t,m)N_i^{tr}(dt,dm|S_i,M_i)\right\}\Big|N^{sp}\right] \\
&= \mathbb{E}\left[\prod_{i=1}^{N^{sp}}\exp\{-N_i^{tr}(\varphi|S_i,M_i)\}\Big|N^{sp}\right] \\
&= \prod_{i=1}^{N^{sp}}\mathbb{E}\left[\exp\{-N_i^{tr}(\varphi|S_i,M_i)\}\Big|N^{sp}\right] \\
&= \prod_{i=1}^{N^{sp}}\mathbb{E}\left[\exp\{-N_i^{tr}(\varphi|S_i,M_i)\}\Big|S_i,M_i\right], \qquad (4.18)
\end{aligned}$$

where the last two equalities follow from the independence of the processes $N_i^{tr}(dt,dm|S_i,M_i)$ and conditional on the knowledge of N^{sp}.

We then have to compute

$$\mathbb{E}\left[\exp\{-N_i^{tr}(\varphi|S_i,M_i)\}\Big|S_i,M_i\right]$$
$$=\mathbb{E}\left[\exp\left\{-\int\varphi(t,m)N_i^{tr}(dt,dm|S_i,M_i)\right\}\Big|S_i,M_i\right].$$

Actually, for any $i\in\mathbb{Z}$, we can define the conditional Laplace functional of $N_i^{tr}(dt,dm|S_i,M_i)$ as in (4.16), i.e.,

$$(\varphi,s,m)\to\mathcal{L}^{tr}(\varphi|s,m):=\mathbb{E}\left[\exp\{-N_i^{tr}(\varphi|S_i,M_i)\}\Big|S_i=s,M_i=m\right]$$
$$=\mathbb{E}\left[\exp\left\{-\int\varphi(t,m)N_i^{tr}(dt,dm|S_i,M_i)\right\}\Big|S_i=s,M_i=m\right],$$

which is independent of i, and then

$$\mathbb{E}\left[\exp\{-N_i^{tr}(\varphi|S_i,M_i)\}\Big|S_i,M_i\right]=\mathcal{L}^{tr}(\varphi|S_i,M_i).$$

Nor, for every $i\in\mathbb{Z}$, we can write

$$N_i^{tr}(dt,dm|S_i,M_i)=N_i^{tr,1}(dt,dm|S_i,M_i)+N_i^{tr,>1}(dt,dm|S_i,M_i),$$

where, given the spontaneous event occurred in S_i and with magnitude M_i, $N_i^{tr,1}(dt,dm|S_i,M_i)$ represents the marked point process of its first generation aftershocks and $N_i^{tr,>1}(dt,dm|S_i,M_i)$ represents the marked point process of its progeny belonging to the generations from the second one on. Setting $(T_j^{(i,1)},M_j^{(i,1)})$ the time and magnitude of the j^{th} daughter, we have that

$$N_i^{tr,1}(dt,dm|S_i,M_i)=\sum_j\delta_{(T_j^{(i,1)},M_j^{(i,1)})}(dt,dm),$$

and

$$N_i^{tr,>1}(dt,dm|S_i,M_i)=\sum_{j=1}^{N_i^{tr,1}}N_j^{tr,>1}(dt,dm|T_j^{(i,1)},M_j^{(i,1)}).$$

For $j=1,\dots,N_i^{tr,1}$, conditional on the knowledge of (S_i,M_i), the random variables $\{M_j^{(i,1)}\}_j$ are independent and identically distributed according to $p(m'|M_i)dm'$ and the ones $\{T_j^{(i,1)}\}_j$ consist of pure temporal point processes with intensity $\varrho(M_i)\Phi(t'-S_i)\mathbb{1}_{(S_i,\infty)}(t')$, independent of the magnitudes $M_j^{(i,1)}$. Furthermore, for each j and conditional on the knowledge of (S_i,M_i) and $(T_j^{(i,1)},M_j^{(i,1)})$, the marked point processes $\{N_j^{tr,>1}(dt,dm|T_j^{(i,1)},M_j^{(i,1)})\}_j$

are independent copies of Hawkes processes with the same kind of conditional intensity as for $N_i^{tr}(dt, dm|S_i, M_i)$, that is the function $\varrho(m')\Phi(t'' - t')\mathbb{1}_{(t',\infty)}(t'')p(m''|m')$. Then,

$$\begin{aligned}
&\mathbb{E}\left[\exp\{-N_j^{tr,>1}(\varphi\,|T_j^{(i,1)}, M_j^{(i,1)})\}\Big|S_i = s_0, M_i = m_0, T_j^{(i,1)} = s, M_j^{(i,1)} = m\right]\\
&= \mathbb{E}\left[\exp\{-N_j^{tr,>1}(\varphi\,|T_j^{(i,1)}, M_j^{(i,1)})\}\Big|T_j^{(i,1)} = s, M_j^{(i,1)} = m\right]\\
&=: \mathcal{L}^{tr}(\varphi|s, m).
\end{aligned}$$

This means that we have obtained the same function $\mathcal{L}^{tr}(\varphi|s, m)$ as in (4.16). Since we have noticed the independence of i, we can also write

$$\mathbb{E}\left[\exp\{-N_j^{tr,>1}(\varphi\,|T_j^{(i,1)}, M_j^{(i,1)})\}\Big|S_i, M_i, T_j^{(i,1)}, M_j^{(i,1)}\right] = \mathcal{L}^{tr}(\varphi|T_j^{(i,1)}, M_j^{(i,1)}).$$

Following the same reasoning as for $\mathcal{L}(\varphi)$, we have that

$$\begin{aligned}
&\mathcal{L}^{tr}(\varphi|s, m) := \mathbb{E}\left[\exp\{-N_i^{tr,1}(\varphi|S_i, M_i) - N_i^{tr,>1}(\varphi\,|S_i, M_i)\}|S_i = s, M_i = m\right]\\
&= \mathbb{E}\Bigg[\exp\{-N_i^{tr,1}(\varphi\,|S_i, M_i)\}\mathbb{E}\Big[\exp\{-N_i^{tr,>1}(\varphi\,|S_i, M_i)\}|N_i^{tr,1}(\cdot, \cdot|S_i, M_i),\\
&\quad S_i = s, M_i = m\Big]\Big|S_i = s, M_i = m\Bigg]\\
&= \mathbb{E}\left[\exp\{-N_i^{tr,1}(\varphi\,|S_i, M_i)\}\prod_{j=1}^{N_i^{tr,1}}\mathcal{L}_j^{tr}(\varphi|T_j^{(1,i)}, M_j^{(1,i)})\Big|S_i = s, M_i = m\right]\\
&= \mathbb{E}\left[\exp\left\{-\int\Big(\varphi(t, m) - \log\big(\mathcal{L}^{tr}(\varphi|t, m)\big)\Big)N_i^{tr,1}(dt, dm)\right\}\Big|S_i = s, M_i = m\right]\\
&= \mathbb{E}\left[\exp\left\{-N_i^{tr,1}\Big(\varphi - \log\mathcal{L}^{tr}\big(\varphi|S_i, M_i\big)\Big)\right\}\Big|S_i = s, M_i = m\right].
\end{aligned}$$

Finally, since the process $N_i^{tr,1} = \{M_j^{1,i}, T_j^{1,i}, j \geq 1\}$, conditional on $S_i = s, M_i = m$, is a marked Poisson process on $\mathbb{R} \times [m_0, \infty)$ with intensity $\lambda_j^{1,i}(dt', dm') = \varrho(m)\Phi(t' - s)\mathbb{1}_{(s,\infty)}(t')dt'p(m'|m)dm'$, we can use the *second Campbell's formula* (see equation (4.20)).

It then follows that

$$\mathcal{L}^{tr}(\varphi|s, m) = \mathcal{L}^{tr,1}(\varphi - \log\mathcal{L}^{tr}(\varphi|\cdot, \cdot)\,|s, m),$$

where

$$\begin{aligned}
&\mathcal{L}^{tr,1}(\psi\,|s, m) := \mathbb{E}\left[\exp\{-N_i^{tr,1}(\psi)\}|S_i = s, M_i = m\right]\\
&\quad= \exp\left\{-\int_{\mathbb{R}}\int_{m_0}^{\infty}\big(1 - e^{-\psi(t',m')}\big)\varrho(m)\Phi(t' - s)\mathbb{1}_{(s,\infty)}(t')\,dt'\,p(m'|m)dm'\right\}
\end{aligned}$$

and

$$\psi := \varphi - \log \mathcal{L}^{tr}(\varphi|\cdot,\cdot).$$

Explicitly, we can write

$$\begin{aligned}
-\ln \mathcal{L}^{tr}(\varphi|s,m) :=& \int_{\mathbb{R}}\int_{m_0}^{\infty}\left(1-e^{-[\varphi(t',m')-\log \mathcal{L}^{tr}(\varphi|t',m')]}\right) \\
&\cdot \varrho(m)\Phi(t'-s)\mathbb{1}_{(s,\infty)}(t')\,dt'\,p(m'|m)dm' \\
=& \int_{s}^{\infty}\int_{m_0}^{\infty}\left(1-e^{-\varphi(t',m')}\mathcal{L}^{tr}(\varphi|t',m')\right) \\
&\cdot \varrho(m)\Phi(t'-s)\,dt'\,p(m'|m)dm'.
\end{aligned} \tag{4.19}$$

In conclusion, we can use the obtained results for deriving $\mathcal{L}(\varphi)$. Recalling (4.17) and (4.18), we get

$$\begin{aligned}
\mathcal{L}(\varphi) &= \mathbb{E}\left[\exp\Big\{-\int \varphi(t,m)N^{sp}(dt,dm)\Big\}\prod_{i=1}^{N^{sp}}\mathcal{L}^{tr}(\varphi|S_i,M_i)\right] \\
&= \mathbb{E}\left[\exp\Big\{-\int \varphi(t,m)N^{sp}(dt,dm)\Big\}\prod_{i=1}^{N^{sp}}\exp\{\log\left(\mathcal{L}^{tr}(\varphi|S_i,M_i)\right)\right] \\
&= \mathbb{E}\left[\exp\Big\{-\int \varphi(t,m)N^{sp}(dt,dm)\Big\}\exp\Big\{\sum_{i=1}^{N^{sp}}\log\left(\mathcal{L}^{tr}(\varphi|S_i,M_i)\right)\Big\}\right] \\
&= \mathbb{E}\left[\exp\Big\{-N^{sp}\Big(\varphi-\log\left(\mathcal{L}^{tr}(\varphi|\cdot,\cdot)\right)\Big)\Big\}\right] \\
&= \mathcal{L}^{sp}\Big(\varphi-\log\left(\mathcal{L}^{tr}(\varphi|\cdot,\cdot)\right)\Big),
\end{aligned}$$

where

$$\mathcal{L}^{sp}(\varphi) := \mathbb{E}\left[\exp\{-N^{sp}(\varphi)\}\right]$$

and then, again using the second Campbell's formula (4.20) and equation (4.19), it holds

$$\begin{aligned}
-\ln \mathcal{L}(\varphi) &= \int_{\mathbb{R}}\int_{m_0}^{\infty}\left(1-e^{-\varphi(s,m)+\log(\mathcal{L}^{tr}(\varphi|s,m))}\right)\varpi\,ds\,p(m)dm \\
&= \int_{\mathbb{R}}\int_{m_0}^{\infty}\left(1-e^{-\Theta(s,m)}\right)\varpi\,ds\,p(m)dm,
\end{aligned}$$

where

$$\begin{aligned}
\Theta(s,m) =& \varphi(s,m) + \int_{s}^{\infty}\int_{m_0}^{\infty}\left(1-e^{-\varphi(t',m')}\mathcal{L}^{tr}(\varphi|t',m')\right) \\
&\cdot \varrho(m)\Phi(t'-s)dt'p(m'|m)dm'.
\end{aligned}$$

The thesis of the theorem is then proved. □

In the above proof we have mentioned the second Campbell's formula, which can be stated as follows.

Theorem 10. *Let N be a Poisson process on $\mathbb{R}^m$ with mean measure $\nu(dx)$. Let ϕ be a non-negative measurable function on $\mathbb{R}^m$. Then, the Laplace functional of the process is*

$$\mathcal{L}(\phi) = \exp\left\{\int \big(e^{-\phi(x)} - 1\big)\nu(dx)\right\}. \tag{4.20}$$

4.2.1 The probability generating function obtained from the Laplace functional

As we are going to see in Chapter 6, the probability generating function (PGF) of the total number of events in the process with magnitude higher than the completeness threshold m_c and occurring in the temporal interval $[0;\tau]$, is very useful for the analysis of the time delay τ between two successive shocks. In this subsection, our aim is to obtain an equation for the probability generating function, starting from the Laplace functional. More precisely, we will derive the PGF for a general time-magnitude window.

Let's then start by considering a magnitude threshold $\underline{m} \geq m_0$, where we recall that m_0 is the reference magnitude, that is the minimum magnitude for an event to trigger its own progeny. We don't consider events with magnitude smaller than m_0 because, as we are going to see in Chapter 6, they will never be taken into account in the analysis. Now, for every z such that $|z| \leq 1$ and $\tau > 0$, we define

$$\varphi(t,m) = \underline{\varphi}(t,m) := \varphi_{z,\tau,\underline{m}}(t,m) = -\log(z)\mathbb{1}_{[0,\tau]}(t)\ \mathbb{1}_{[\underline{m},\infty)}(m). \tag{4.21}$$

Recalling (4.12), it then follows that

$$\mathcal{L}(\underline{\varphi}) = \mathbb{E}\left[\exp\{-N(\varphi_{z,\tau,\underline{m}})\}\right] = \mathbb{E}\left[z^{N\left([0,\tau]\times[\underline{m},\infty)\right)}\right],$$

where $N\big([0,\tau]\times[\underline{m},\infty)\big)$ is the total number of events with times and magnitudes in $[0,\tau]\times[\underline{m},\infty)$ and the probability generating function of the total number of events of the process in $[0,\tau]\times[\underline{m},\infty)$ is derived.

Now, in order to obtain the explicit expression of the above PGF, we have to substitute (4.21) in (4.14). Firstly, anticipating the notations of Chapter 6, we define here

$$G^{tr}(z;\tau|s,m) := \mathbb{E}\left[z^{N_i^{tr}\left([0,\tau]\times[\underline{m},\infty)\right)}\Big|S_i = s, M_i = m\right] = \mathcal{L}^{tr}(\underline{\varphi}|s,m), \tag{4.22}$$

where $N_i^{tr}\big([0,\tau]\times[\underline{m},\infty)\big)$ is the number of events in $[0,\tau]\times[\underline{m},\infty)$ triggered by the spontaneous shock (S_i, M_i).

Remark 8. Conditional on $(S_i = s, M_i = m)$, if $s > \tau$ it holds $N_i^{tr}\big([0,\tau]\times[\underline{m},\infty)\big) = 0$; then,

$$G^{tr}(z;\tau|s,m) = 1 \quad \text{for all } s > \tau. \tag{4.23}$$

Since

$$e^{-\underline{\varphi}(t,m)} = z^{\mathbb{1}_{[0,\tau]\times[\underline{m},\infty)}(t,m)} = \begin{cases} z & \text{if } t \in [0,\tau],\ m \in [\underline{m},\infty) \\ 1 & \text{otherwise,} \end{cases} \tag{4.24}$$

it follows that we can write

$$\begin{aligned} &-\ln \mathcal{L}^{tr}(\underline{\varphi}|s,m) = -\ln G^{tr}(z;\tau|s,m) \\ &= \int_s^\infty \int_{m_0}^\infty \big(1 - e^{-\underline{\varphi}(t',m')}\mathcal{L}^{tr}(\underline{\varphi}|t',m')\big)\,\varrho(m)\Phi(t'-s)\,dt'\,p(m'|m)dm' \\ &= \int_s^\infty \int_{m_0}^\infty \big(1 - z^{\mathbb{1}_{[0,\tau]\times[\underline{m},\infty)}(t',m')}G^{tr}(z;\tau|t',m')\big)\,\varrho(m)\Phi(t'-s)\,dt'\,p(m'|m)dm'. \end{aligned} \tag{4.25}$$

Let's notice that

$$1 - z^{\mathbb{1}_{[0,\tau]\times[\underline{m},\infty)}(t',m')}G^{tr}(z;\tau|t',m') = 0$$

if and only if $G^{tr}(z;\tau|t',m') = 1$ and $z^{\mathbb{1}_{[0,\tau]\times[\underline{m},\infty)}(t',m')} = 1$; then, we can discard the integral over $[\tau,\infty)\times[\underline{m},\infty)$ in (4.25). Therefore, taking into account also (4.23) and (4.24), we have that equation (4.25) becomes:

- **if** $s > \tau$,
$$-\ln \mathcal{L}^{tr}(\underline{\varphi}|s,m) = 0;$$
- **if** $0 < s \le \tau$,
$$\begin{aligned} &-\ln \mathcal{L}^{tr}(\underline{\varphi}|s,m) \\ &= \int_s^\tau \int_{m_0}^\infty \big(1 - z^{\mathbb{1}_{[\underline{m},\infty)}(m')}G^{tr}(z;\tau|t',m')\big)\,\varrho(m)\Phi(t'-s)\,dt'\,p(m'|m)dm' \\ &= \int_s^\tau \int_{\underline{m}}^\infty \big(1 - z\,G^{tr}(z;\tau|t',m')\big)\,\varrho(m)\Phi(t'-s)\,dt'\,p(m'|m)dm' \\ &\quad + \int_s^\tau \int_{m_0}^{\underline{m}} \big(1 - G^{tr}(z;\tau|t',m')\big)\,\varrho(m)\Phi(t'-s)\,dt'\,p(m'|m)dm'; \end{aligned}$$

- **if** $s < 0$

$$
\begin{aligned}
&-\ln \mathcal{L}^{tr}(\underline{\varphi}|s,m)\\
&= \int_s^0 \int_{m_0}^{\infty} \big(1 - G^{tr}(z;\tau|t',m')\big)\varrho(m)\Phi(t'-s)\,dt'\,p(m'|m)dm'\\
&\quad + \int_0^{\tau} \int_{m_0}^{\infty} \big(1 - z^{\mathbb{1}_{[\underline{m},\infty)}(m')} G^{tr}(z;\tau|t',m')\big)\varrho(m)\Phi(t'-s)\,dt'\,p(m'|m)dm'\\
&= \int_s^0 \int_{m_0}^{\infty} \big(1 - G^{tr}(z;\tau|t',m')\big)\varrho(m)\Phi(t'-s)\,dt'\,p(m'|m)dm'\\
&\quad + \int_0^{\tau} \int_{\underline{m}}^{\infty} \big(1 - z\,G^{tr}(z;\tau|t',m')\big)\varrho(m)\Phi(t'-s)\,dt'\,p(m'|m)dm'\\
&\quad + \int_0^{\tau} \int_{m_0}^{\underline{m}} \big(1 - G^{tr}(z;\tau|t',m')\big)\varrho(m)\Phi(t'-s)\,dt'\,p(m'|m)dm'.
\end{aligned}
$$

Exactly in the same way, we can derive the Laplace functional $\mathcal{L}(\underline{\varphi})$. Recalling equations (4.13), (4.22) and (4.24), we have that

$$
\begin{aligned}
-\ln \mathcal{L}(\underline{\varphi}) &= \int_{\mathbb{R}} \int_{m_0}^{\infty} \big(1 - e^{-\underline{\varphi}(s,m)+\log(\mathcal{L}^{tr}(\underline{\varphi}|s,m))}\big)\varpi\,ds\,p(m)dm\\
&= \int_{\mathbb{R}} \int_{m_0}^{\infty} \big(1 - z^{\mathbb{1}_{[0,\tau]\times[\underline{m},\infty)}(s,m)} G^{tr}(z;\tau|s,m)\big)\varpi\,ds\,p(m)dm.
\end{aligned}
$$

Then, with the same considerations as for $\mathcal{L}^{tr}(\underline{\varphi}|s,m)$, we get

- **if** $s > \tau$,

$$-\ln \mathcal{L}(\underline{\varphi}) = 0;$$

- **if** $0 < s \le \tau$,

$$
\begin{aligned}
&-\ln \mathcal{L}(\underline{\varphi})\\
&= \int_0^{\tau} \int_{m_0}^{\infty} \big(1 - z^{\mathbb{1}_{[\underline{m},\infty)}(m)} G^{tr}(z;\tau|s,m)\big)\varpi\,ds\,p(m)dm\\
&= \int_0^{\tau} \int_{\underline{m}}^{\infty} \big(1 - z\,G^{tr}(z;\tau|s,m)\big)\varpi\,ds\,p(m)dm\\
&\quad + \int_0^{\tau} \int_{m_0}^{\underline{m}} \big(1 - G^{tr}(z;\tau|s,m)\big)\varpi\,ds\,p(m)dm;
\end{aligned}
$$

- **if** $s < 0$

$$
\begin{aligned}
&-\ln \mathcal{L}(\underline{\varphi}) \\
&= \int_{-\infty}^{0}\int_{m_0}^{\infty} \left(1 - z^{\mathbb{1}_{[\underline{m},\infty)}(m)} G^{tr}(z;\tau|s,m)\right) \varpi \, ds \, p(m) dm \\
&= \int_{-\infty}^{0}\int_{\underline{m}}^{\infty} \left(1 - z \, G^{tr}(z;\tau|s,m)\right) \varpi \, ds \, p(m) dm \\
&\quad + \int_{-\infty}^{0}\int_{m_0}^{\underline{m}} \left(1 - G^{tr}(z;\tau|s,m)\right) \varpi \, ds \, p(m) dm.
\end{aligned}
$$

We anticipate that the same result will be obtained in Section 6.1 of Chapter 6, but with a different, less technical and more intuitive method, based on approximation arguments.

The Laplace functional (and then in particular the probability generating function) allows to study several aspects of the process, that could be very relevant in terms of seismic analysis. For example, if we choose

$$
\phi(t,m) = -\log(z_1) \mathbb{1}_{[0,\tau]\times[\underline{m}_1,\overline{m}_1)}(t,m) - \log(z_2) \mathbb{1}_{[0,\tau]\times[\underline{m}_2,\overline{m}_2)}(t,m),
$$

the Laplace functional becomes

$$
\mathbb{E}\left[z_1^{N\left([0,\tau]\times[\underline{m}_1,\overline{m}_1)\right)} \, z_2^{N\left([0,\tau]\times[\underline{m}_2,\overline{m}_2)\right)} \right]
$$

and one could study the relation between the magnitudes belonging to two different fixed magnitude subintervals of interest.

To our knowledge, in the literature there is only one work concerning a similar analysis: in their manuscript of 2015, independently of us, Roueff *et al.* analyze the Laplace functional of locally stationary Hawkes processes [Roueff et al., 2015]. Nevertheless, as already explained in Section 4.1, they present a condition for the non-explosion of their process that is not satisfied by our magnitude correlated model (4.3).

We conclude saying that the analysis made here could be really of interest for the study of marked Hawkes processes modeling seismic sequences.

4.3 Magnitude transition probability density function for triggered events' magnitude: an explicit form

In the new ETAS model with correlated magnitudes, completely characterized by the conditional intensity (4.3), the law of the magnitudes relative to the time-stationary, non-homogeneous Poisson process modeling the aftershocks component of a seismic sequence is given by the transition probability $p(\cdot|\cdot)$. We derive here the explicit form of the latter probability, anticipated in (4.2), in accordance with the experimental results of Chapter 3 (see the Properties *Pr.1* and *Pr.2* in Section 3.4).

To begin with, in what follows we will indicate with m' the triggering event's magnitude and with $p_{tr}(m'')$ the probability for an event to have magnitude m'', given that it is not a spontaneous event.

Lemma 2. *Let's consider a marked branching process describing an earthquake sequence, in which the immigrants are modeled by a homogeneous Poisson process, the offsprings are modeled by a time-stationary, non-homogeneous Poisson process and the marks are the events' magnitudes. Then, the probability for an event to have magnitude m'', given the fact that it isn't an immigrant, is*

$$p_{tr}(m'') := \int_{m_0}^{\infty} \frac{p(m')\varrho(m')p(m''|m')}{\int_{m_0}^{\infty} p(m')\varrho(m')dm'} dm', \tag{4.26}$$

where $p(m') = \beta e^{-\beta(m'-m_0)}$ is the Gutenberg-Richter law, assumed valid for the magnitudes of the immigrants, $\varrho(m') = \kappa e^{a(m'-m_0)}$ is the productivity law and m_0 is the reference magnitude, that is the minimum magnitude value for an event to trigger other shocks.

Proof. Let's consider the following three random variables:

- M' is the variable for the triggering events' magnitude, distributed according to the Gutenberg-Richter law $p(\cdot)$;
- N counts the number of shocks that a generic mother event triggers; it is such that $\mathbb{P}\{N = k\} = \int_{m_0}^{\infty} \mathbb{P}\{N = k|m'\}p(m')dm'$;
- $N^1(J)$ counts how many triggered shocks among the previous N have magnitude in the set J.

Since

$$N^1(J)|(M' = m', N = k) \sim B\Big(k, p(J|m')\Big),$$

where

$$p(J|m') = \int_J p(x|m')dx$$

and $B(k,p)$ indicates the law of a binomial random variable with parameters k and p, we have

$$\mathbb{P}\{N^1(J)=q\} = \int_{m_0}^{\infty}\sum_{k=0}^{\infty}\binom{k}{q}p(J|m')^q(1-p(J|m'))^{k-q}\mathbb{P}\{N=k|m'\}p(m')dm'.$$

Let's consider now $J=[m'',m''+\delta)$ and $\widetilde{n}$ i.i.d. realizations of the tern of variables $\Big\{M_i', N_i, N_i^1\big([m'',m''+\delta)\big)\Big\}_{i=1,\dots,\widetilde{n}}$. Let's focus on the ratio between the sample means of the random variables $N_i^1(J)$ and N_i, for $i=1,\dots,\widetilde{n}$, i.e.,

$$\frac{\frac{1}{\widetilde{n}}\sum_{i=1}^{\widetilde{n}}N_i^1(J)}{\frac{1}{\widetilde{n}}\sum_{i=1}^{\widetilde{n}}N_i}.$$

By looking at the ratio of the expected values of the i.i.d. random variables $N_i^1(J)$ and N_i, i.e.,

$$\frac{\mathbb{E}\big[N_1^1(J)\big]}{\mathbb{E}\big[N_1\big]} = \frac{\mathbb{E}\Big[\mathbb{E}\big[N_1^1(J)|M_1',N_1\big]\Big]}{\mathbb{E}\Big[\mathbb{E}\big[N_1|M_1'\big]\Big]},$$

since

$$\mathbb{E}\Big[N_1^1\big([m'',m''+\delta)\big)|M_1'=m',N_1=k\Big] = k\int_{m''}^{m''+\delta}p(x|m')dx$$

and

$$\mathbb{E}\big[N_1|M_1'=m'\big] = \sum_{k=0}^{\infty}k\mathbb{P}\{N=k|m'\} = \varrho(m'),$$

we can deduce that the probability measure

$$J \to \mu(J) := \frac{\mathbb{E}\Big[N_1^1(J)\Big]}{\mathbb{E}\big[N_1\big]}$$

has density $p_{tr}(m'')$. $\square$

Now, we are searching for a transition probability density function $p(\cdot|m')$, i.e.,

$$\int_{m_0}^{\infty}p(m''|m')dm'' = 1, \tag{4.27}$$

such that it holds

$$p_{tr}(m'') = p(m''). \tag{4.28}$$

This latter condition is a focal point of this study: it corresponds to the need of obtaining the Gutenberg-Richter law when considering the magnitudes of a generic event, without taking into account the generation to which it belongs and the relative characteristics of past seismicity. The requirement (4.28) represents exactly the invariance of the Gutenberg-Richter law, weighted by $\varrho(m')$, with respect to the transition probability density function assumed valid for triggered events' magnitude. More precisely, although we assume the existence of a density of the aftershocks' magnitude different from the Gutenberg-Richter law, at the same time the above condition (4.28) justifies its validity: it tells that, if we average the transition density $p(m''|m')$ over all the possible triggering events' magnitudes m', distributed according to the Gutenberg-Richter law, and taking into account the productivity model, we obtain again the Gutenberg-Richter law with the same parameter. This shows that first generation events' magnitude, without any other information about the magnitude of the mother events, follows the Gutenberg-Richter law as assumed for the background shocks' magnitudes. In addition, this property will be obviously true for events' magnitude of any generation. Hence, iterating the reasoning, a generic event's magnitude, whatever generation it belongs to, will be distributed according to the Gutenberg-Richter law.

Let's focus then on the specific problem of finding a suitable density $p(m''|m')$ such that (4.28) holds. Firstly, recalling that the branching ratio η, also defined as the average number of first generation triggered shocks $N_i^{tr,1}$ per triggering event i, is independent of i and is given by

$$\eta = \mathbb{E}[N_1^{tr,1}] = \int_{m_0}^{\infty} p(m')\varrho(m')dm' = \frac{\beta\kappa}{\beta - a},$$

it follows that the above condition (4.28) is equivalent to

$$\int_{m_0}^{\infty} p(m')\varrho(m')p(m''|m')dm' = \eta p(m''). \tag{4.29}$$

Furthermore, as already anticipated, the law $p(m''|m')$ must have a qualitative behavior in accordance with the results of the previous Chapter 3 (see the Properties *Pr.1* and *Pr.2* in Section 3.4). More precisely, the function we are looking for should be such that, when m' increases, the probability of having events with high magnitudes must increase and, at the same time, the one of having events with low magnitudes must decrease obviously. In addition, for high values of the mother events' magnitudes the function may also have a relative maximum.

In view of these considerations, the choice we adopt for $p(m''|m')$ is

$$p(m''|m') = p(m'')\big[1 + f(m', m'')\big], \tag{4.30}$$

with

$$f(m', m'') = -q(m') + s(m')\Big(1 - e^{-\beta(m''-m_0)}\Big). \tag{4.31}$$

Remark 9. The choice of the parameter β is arbitrary and we could take any parameter $\gamma > 0$ instead of β, obtaining similar results.

The functions $q(m')$ and $s(m')$ in (4.31) have to be found opportunely, by imposing the following conditions:

1) $\int_{m_0}^{\infty} p(m''|m')dm'' = 1$,

2) $p(m''|m') \geq 0$,

3) $p(m''|m')$ is such that condition (4.29) is verified,

4) $p(m''|m')$ has the qualitative behavior previously introduced.

Lemma 3. *Let's consider the function $p(m''|m')$ defined in* (4.30)*, with $f(m', m'')$ as in* (4.31)*. Then, the function $p(m''|m')$ is a non-negative probability density such that* (4.29) *holds, if and only if*

- $s(m') = 2q(m')$,
- $|q(m')| \leq 1$
- *and*

$$\int_{m_0}^{\infty} p(m')\varrho(m')q(m')dm' = 0. \tag{4.32}$$

Proof. The transition probability function $p(m''|m')$ is a density if and only if

$$\int_{m_0}^{\infty} p(m'')\big[1 + f(m', m'')\big]dm'' = 1 \quad \Leftrightarrow \quad \int_{m_0}^{\infty} p(m'')f(m', m'')dm'' = 0.$$

By substituting the expression (4.31) chosen for $f(m', m'')$, we obtain

$$\begin{aligned}
0 &= \int_{m_0}^{\infty} p(m'')\big[-q(m') + s(m') - s(m')e^{-\beta(m''-m_0)}\big]dm'' \\
&= -q(m') + s(m') - s(m')\int_{m_0}^{\infty} \beta e^{-2\beta(m''-m_0)}dm'' \\
&= -q(m') + s(m') - \frac{s(m')}{2} \\
&= -q(m') + \frac{s(m')}{2},
\end{aligned}$$

from which the first result in the lemma is proved:

$$s(m') = 2q(m').$$

If we substitute this latter result in equation (4.31), we get

$$\begin{aligned}
f(m', m'') &= q(m') - 2q(m')e^{-\beta(m''-m_0)} \\
&= q(m')\big(1 - 2e^{-\beta(m''-m_0)}\big).
\end{aligned}$$

Now, recalling equation (4.30), in order to have also $p(m''|m') \geq 0$, it should hold $f(m', m'') \geq -1$, that is

$$q(m')\big(2e^{-\beta(m''-m_0)} - 1\big) \leq 1.$$

Since $(2e^{-\beta(m''-m_0)} - 1)$ takes all the values in $(-1, 1]$ as m'' varies in $[m_0, \infty)$, the previous condition is equivalent to

$$|q(m')| \leq 1.$$

Finally, let's impose the condition (4.29):

$$\int_{m_0}^{\infty} p(m')\varrho(m')p(m'')\big[1 + f(m', m'')\big]dm' = \eta p(m'')$$

$$\Updownarrow$$

$$\int_{m_0}^{\infty} p(m')\varrho(m')f(m', m'')dm' = 0.$$

For

$$f(m', m'') = q(m')\big(1 - 2e^{-\beta(m''-m_0)}\big),$$

we get

$$\begin{aligned}
0 &= \int_{m_0}^{\infty} p(m')\varrho(m')q(m')\big(1 - 2e^{-\beta(m''-m_0)}\big)dm' \\
&= \big(1 - 2e^{-\beta(m''-m_0)}\big)\int_{m_0}^{\infty} p(m')\varrho(m')q(m')dm'.
\end{aligned}$$

A sufficient condition is then given by

$$\int_{m_0}^{\infty} p(m')\varrho(m')q(m')dm' = 0.$$

□

We have now to focus on the choice of the function $q(m')$, such that conditions (4.32) and $|q(m')| \leq 1$ hold. Furthermore, this function must be such that $p(m''|m')$ has the qualitative behavior in accordance with the results of Chapter 3, as previously explained. Then, let's begin with some qualitative comments.

Since the function $q(m')$ should be such that the law $p(m''|m')$ has the qualitative behavior introduced in Properties *Pr.1* and *Pr.2* of Section 3.4, we assume $q(m')$ to be continuous and increasing. Furthermore, we impose that $q(m')$ is negative for $m' < \overline{m}$, for a certain $\overline{m}$ at which it becomes zero, and positive elsewhere (see Fig. 4.1). Let's notice that, when $m' = \overline{m}$, it holds $f(m', m'') = 0$ and the transition probability density becomes the Gutenberg-Richter law. From the fact that

$$f(m', m'') = q(m')\big(1 - 2e^{-\beta(m''-m_0)}\big), \tag{4.33}$$

as we have seen in the previous proof, one can observe that, since the function

$$1 - 2e^{-\beta(m''-m_0)}$$

is always increasing, the increasing or decreasing of $f(m', \cdot)$ will be determined by the sign of $q(m')$. In particular, for $m' > \overline{m}$ where $q(m') > 0$, the function $f(m', m'')$ increases in m''. For smaller (higher) values of

$$m'' = \frac{1}{\beta}\ln 2 + m_0,$$

where $1 - 2e^{-\beta(m''-m_0)} = 0$, the function $f(m', m'')$ will have negative (positive) sign. Hence, the transition probability density $p(m''|m') = p(m'')[1 + f(m', m'')]$ is below $p(m'')$ for magnitude values smaller than $m'' = \frac{1}{\beta}\ln 2 + m_0$ and above it for higher magnitude values. Viceversa, for $m' < \overline{m}$ where

$q(m') < 0$, the function $f(m', m'')$ is decreasing in m''. With analogous reasonings we come to the conclusion that $p(m''|m')$ is above the Gutenberg-Richter law for values smaller than $m'' = \frac{1}{\beta}\ln 2 + m_0$, below for higher values. This property reflects the difference between the latter law and our transition probability density function: the Gutenberg-Richter law assigns a magnitude to each event independently of its past history, instead with the law $p(\cdot|m')$ we have that stronger events generates stronger aftershocks with a higher probability.

Remark 10. For triggered events' magnitude values such that $m'' > \frac{1}{\beta}\ln 2 + m_0$, the factor $1 - 2e^{-\beta(m''-m_0)}$ is positive. Hence, fixed m'' larger than the just-mentioned value, since $q(m')$ is increasing, the function $f(m', m'')$ is increasing with respect to the first argument, too. This means that when triggering event's magnitude increases, so does the transition probability density function at a point $m'' > \frac{1}{\beta}\ln 2 + m_0$. On the contrary, the value of the transition probability density decreases at a value $m'' < \frac{1}{\beta}\ln 2 + m_0$. Then, there is qualitative agreement with the results obtained in Chapter 3.

To sum up, the function $q(m')$ must be:

- less than one in absolute value,
- increasing,
- negative for magnitudes less than a certain value $\overline{m}$, positive otherwise,
- such that condition (4.32) is verified.

The choice we make for $q(m')$ is the following:

$$q(m') = -C_1 + C_2\big(1 - e^{-(\beta-a)(m'-m_0)}\big), \tag{4.34}$$

with $\beta > a$ and where the constants C_1 and C_2 are obtained imposing the above-cited conditions for $q(\cdot)$.

Remark 11. Again, the choice of the parameter $\beta - a$ is arbitrary: we could take any parameter $\xi > 0$ such that $\beta - a > \xi$ instead of $\beta - a$ and obtain similar results.

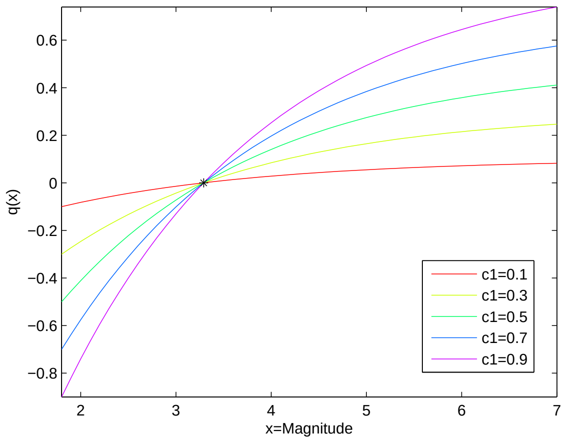


Figure 4.1: Plot of the function $q(x) = C_1(1 - 2\exp\{-0.4648(x - 1.8)\})$. The value at which the function becomes zero is $m^* = 3.2913$.

Now, in order to find the constants C_1 and C_2, let's consider at first equation (4.32). Recalling the expressions of $p(m')$ and $\varrho(m')$, we get

$$\begin{aligned} 0 &= \beta\kappa \int_{m_0}^{\infty} e^{-(\beta-a)(m'-m_0)} \Big[-C_1 + C_2\big(1 - e^{-(\beta-a)(m'-m_0)}\big)\Big]\, dm' \\ &= \beta\kappa \Big[(-C_1 + C_2)\int_{m_0}^{\infty} e^{-(\beta-a)(m'-m_0)} dm' - C_2 \int_{m_0}^{\infty} e^{-2(\beta-a)(m'-m_0)} dm'\Big] \\ &= \beta\kappa \Big[\frac{-C_1 + C_2}{\beta - a} - \frac{C_2}{2(\beta - a)}\Big] \\ &= \frac{\beta\kappa}{\beta - a}\Big(-C_1 + \frac{C_2}{2}\Big), \end{aligned}$$

for $\beta > a$, which is one of the conditions for the non-explosion of the process. It then follows

$$C_2 = 2C_1.$$

Hence

$$q(m') = C_1 - 2C_1 e^{-(\beta-a)(m'-m_0)} = C_1\big(1 - 2e^{-(\beta-a)(m'-m_0)}\big), \tag{4.35}$$

with $\beta > a$. In order to have also $|q(m')| \leq 1$, since $|1 - 2e^{-(\beta-a)(m'-m_0)}| \leq 1$, it is enough to choose $0 \leq C_1 \leq 1$. Using that $C_1 \geq 0$ and $\beta > a$, one can easily verify that the function $q(\cdot)$ is increasing (Fig. 4.1). Furthermore, if $C_1 \neq 0$, the value for which $q(m') = 0$ is

$$m' = \overline{m} = \frac{1}{\beta - a}\ln 2 + m_0.$$

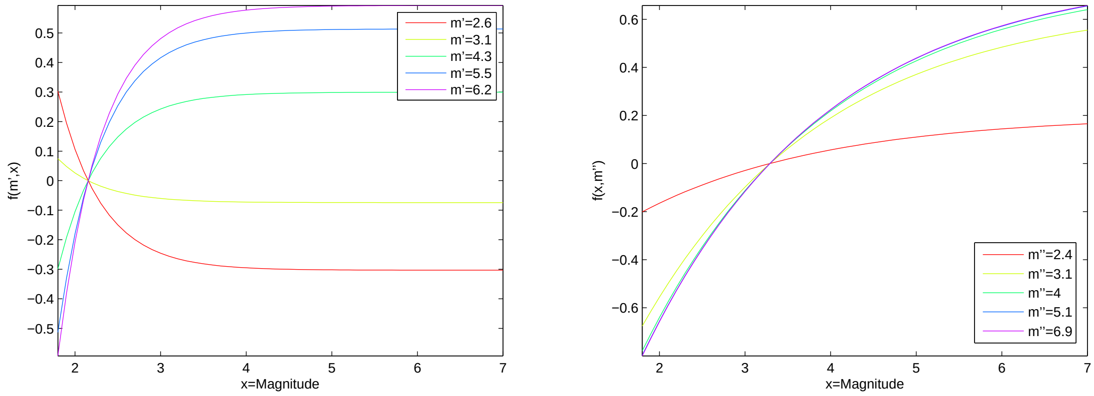


Figure 4.2: Functions $f(m', x) = 0.8(1-2\exp\{-0.4648(m'-1.8)\})(1-2\exp\{-1.9648(x-1.8)\})$ and $f(x, m'') = 0.8(1-2\exp\{-0.4648(x-1.8)\})(1-2\exp\{-1.9648(m''-1.8)\})$, respectively in the left and right plots. In the left plot the value that varies is the triggering events' magnitude. On the other hand, in the right one the varying value is the triggered events' magnitude.

Since $q(m')$ increases, it will be positive when m' is larger than this value and negative elsewhere.

At this point, by substituting the expression (4.35) of $q(m')$ in (4.33), we get

$$f(m', m'') = C_1\big(1 - 2e^{-(\beta-a)(m'-m_0)}\big)\big(1 - 2e^{-\beta(m''-m_0)}\big). \tag{4.36}$$

In the left plot of Fig. 4.2 we can see the function $f(m', \cdot)$ obtained with the values of parameters such that we get $\frac{\ln 2}{\beta-a} + m_0 = 3.2913$. This is the value at which $q(m')$ becomes zero. These values of parameters correspond to a real situation because they were estimated based on a real data set. As expected, $f(m', m'')$ decreases (increases) with respect to m'' when m' is smaller (higher) than the just-mentioned value. Furthermore, it holds $f(m', m'') = 0$ for $m'' = \frac{\ln 2}{\beta} + m_0 = 2.1527$. In the right plot of Fig. 4.2 one can observe that, for any fixed value of $m'' > 2.1527$, the function $f(m', m'')$ is increasing with respect to the first variable. As expected, different values of C_1 don't change the monotony of the function $f(m', m'')$. This parameter only controls the amplitude of the concavity, as one can see in Fig. 4.3.

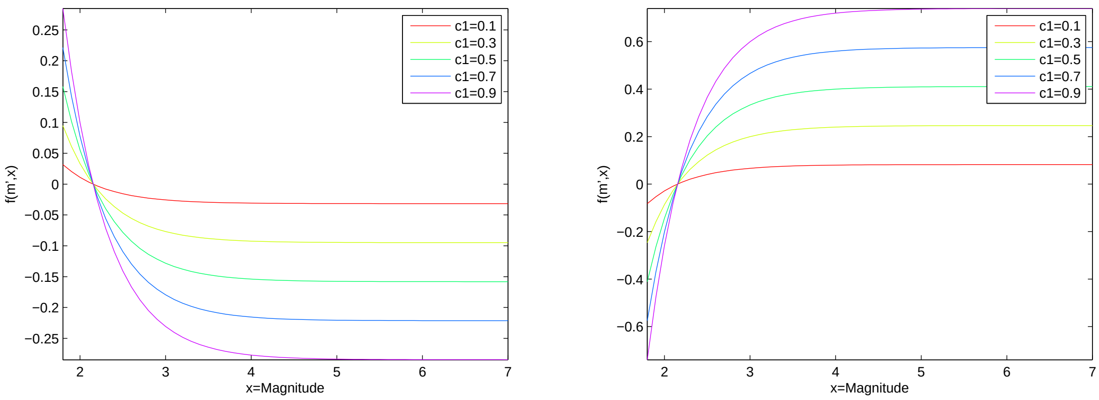


Figure 4.3: Function $f(m', x) = 0.8(1-2\exp\{-0.4648(m'-1.8)\})(1-2\exp\{-1.9648(x-1.8)\})$ respectively with $m' = 2.7$ and $m' = 7$ in the left and right plots, for different values of the parameter C_1.

In conclusion, by substituting equations (2.2) and (4.36) in (4.30), we obtain our proposal for the transition probability density function relative to triggered event's magnitude, with respect to the magnitude m' of its own triggering event:

$$p(m''|m') = \beta e^{-\beta(m''-m_0)}\left[1 + C_1\left(1-2e^{-(\beta-a)(m'-m_0)}\right)\left(1-2e^{-\beta(m''-m_0)}\right)\right], \tag{4.37}$$

for $0 \leq C_1 \leq 1$. Then, the above transition probability density function is the product between the Gutenberg-Richter law, computed in the magnitude of the triggered event considered, and another function of both the magnitude of this triggered event and the one of its mother, separated.

Let's study now the behavior of $p(m''|m')$ by looking at its derivative. Recalling equations (4.30) and (4.33), we have

$$\begin{aligned}
\frac{dp(m''|m')}{dm''} &= \frac{dp(m'')}{dm''}[1 + f(m', m'')] + p(m'')\frac{df(m', m'')}{dm''} \\
&= -\beta p(m'')[1 + f(m', m'')] + p(m'')\frac{df(m', m'')}{dm''} \\
&= \beta p(m'')\left[-1 - f(m', m'') + 2q(m')e^{-\beta(m''-m_0)}\right] \\
&= \beta p(m'')\left[-1 - q(m') + 2q(m')e^{-\beta(m''-m_0)} + 2q(m')e^{-\beta(m''-m_0)}\right] \\
&= \beta p(m'')\left[-1 - q(m') + 4q(m')e^{-\beta(m''-m_0)}\right].
\end{aligned}$$

It follows that if $q(m') \leq 0$, that is $m' \leq \frac{1}{\beta-a}\ln 2 + m_0$, since $|q(m')| < 1$, the density function $p(m''|m')$ is always decreasing in m''. If instead $q(m') > 0$, that is $m' > \frac{1}{\beta - a} \ln 2 + m_0$, the above-cited density increases in m'' till a certain maximum, reached in m_* such that

$$4q(m')e^{-\beta(m_* - m_0)} = 1 + q(m') \quad \Leftrightarrow \quad m_* = \frac{1}{\beta} \ln \frac{4q(m')}{1 + q(m')} + m_0,$$

and for values larger than m_* it decreases, in agreement with the experimental results obtained in the previous chapter. This trend is shown in Fig. 4.4, in which we can also verify the behavior described in Remark 10. Finally, in Fig. 4.5 we can see how the transition probability density function varies with the parameter C_1.

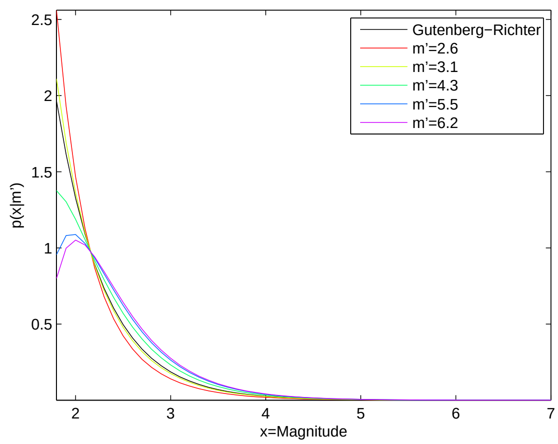


Figure 4.4: Plot of the probability density function relative to triggered events' magnitude when triggering events' magnitude varies and $C_1 = 0.8$.

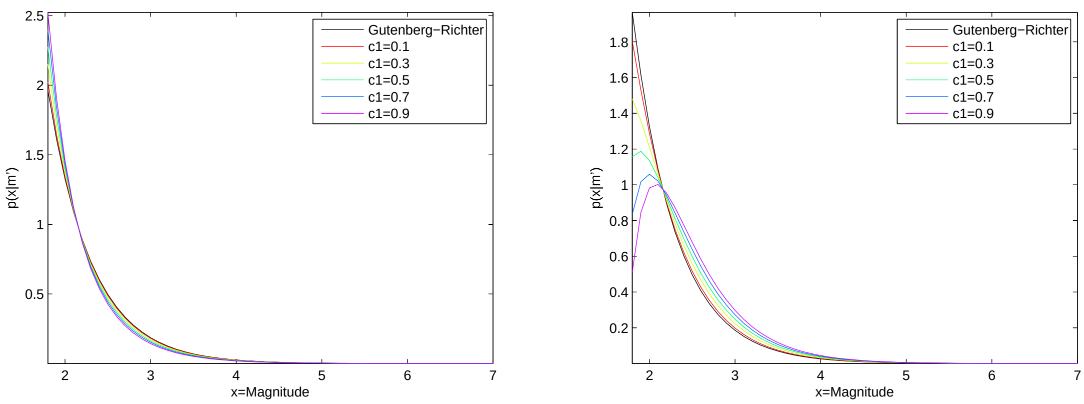


Figure 4.5: Transition probability density function $p(x|m')$ respectively with $m' = 2.7$ and $m' = 7$ in the left and right plots, for different values of the parameter C_1.

Chapter 5

The analysis of the simulated catalogs

In the latter chapter we have proposed an explicit form of the transition probability density function $p(m''|m')$ for the triggered events' magnitudes, with respect to the mother events' ones. In order to demonstrate that our hypothesis of magnitude correlation remains true even when we consider synthetic sets of data, not affected by any kind of "real effect" that may influence the analysis, we analyze some pure simulated catalogs, obtained with two different approaches. For brevity, we will show here the results concerning only two of the simulated catalogs. We want to specify that one of these two datasets is simulated by using the just proposed explicit form for $p(m''|m')$, as explained below. The simulation is solely temporal, in fact we recall that this study is based only on the temporal-magnitude aspect.

5.1 The catalogs simulation

The initial step for this analysis consists of simulation. We proceed in the following two different ways.

1. In the case of the first simulated catalog, we use the temporal version of the FORTRAN program [**etasim.f**], written by Ogata [Ogata, 1981, 1998, 2006]. It is the well-known algorithm for the simulation of seismic catalogs presented in Chapter 2. We just modify it in the fact that the number of events to be simulated is now random between the two input starting and ending times, instead of being fixed at the beginning without specifying the temporal interval for the simulation. We choose here the option of simulating the magnitudes with the Gutenberg-Richter law, instead of taking them from a given catalog. Then, we expect no

evidence of our hypothesis of magnitude correlation for the synthetic catalogs obtained, when performing the two types of analysis above described. The input parameters for the simulation in this catalog are

$$\beta = b \ln 10 = 1 \ln 10 = 2.3,$$
$$(\varpi, \kappa, c, a, p) = (0.55, 0.022, 0.014, 1.7, 1.09). \tag{5.1}$$

2. The second catalog is instead simulated with a program very similar to the Ogata's one: we have adapted it to our hypothesis of magnitude correlation. More precisely, we simulate only the magnitudes of background events with the Gutenberg-Richter law. Instead, the magnitudes of the triggered events are simulated with the new transition probability density function (4.37) obtained in Chapter 4. The algorithm is described in the next subsection. The input parameters for the simulation of this catalog are

$$\beta = b \ln 10 = 1 \ln 10 = 2.3, \qquad C_1 = 0.9,$$
$$(\varpi, \kappa, c, a, p) = (0.55, 0.022, 0.014, 0.8, 1.09). \tag{5.2}$$

In both the two simulated catalogs we consider only one threshold magnitude value, that is 1.5. Furthermore, we set a null learning period: since the parameters are given in input and the data are simulated, the influence of external (precursory) events affecting the seismicity in the target interval is not considered crucial for the analysis as for the real catalogs. Since the simulations aren't spatial, there are no conditions on the maximum depth to be considered.

The time-correlations are shown in Fig. 5.1, while Fig. 5.2 contains the magnitudes versus times plots for the above two simulated catalogs.

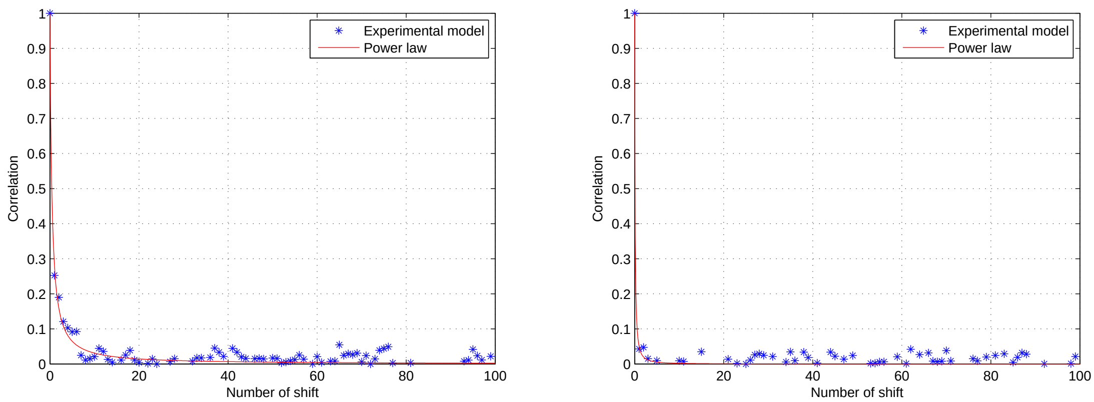


Figure 5.1: Time-correlations of the two simulated catalogs. The left plot contains the correlation of the catalog simulated with the classical Ogata's model, while the right one concerns the catalog simulated with our new method, based on the hypothesis of correlation between the magnitudes of triggered events and the ones of the corresponding mother events.

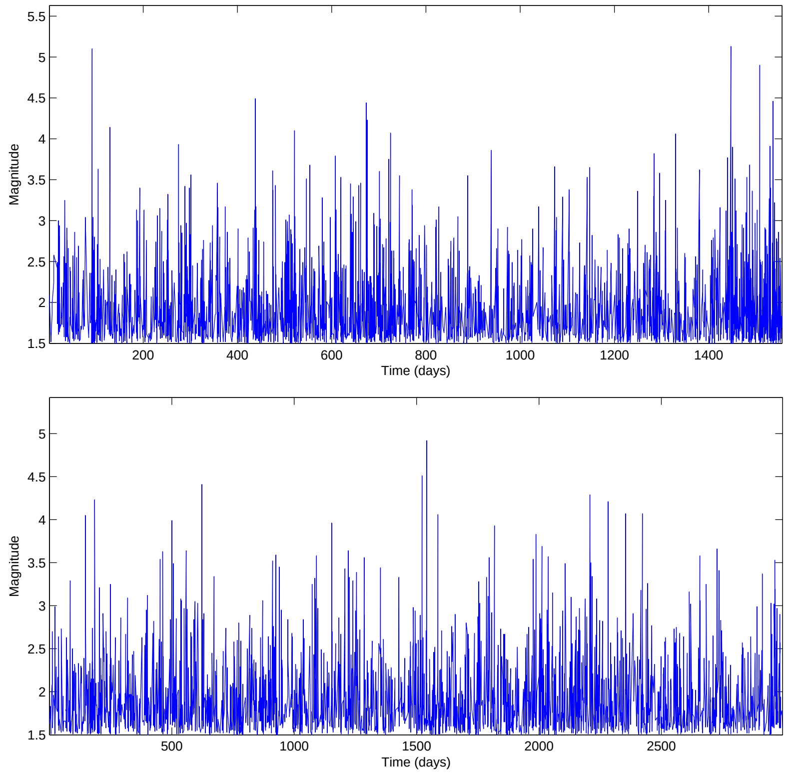


Figure 5.2: Times versus magnitudes of the events in the simulated catalogs. The plot at the top contains the results concerning the catalog simulated with the classical Ogata's model, while the one at the bottom contains the results relative to the catalog simulated with our new method, based on the hypothesis of correlation between the magnitudes of triggered events and the ones of the corresponding mother events.

5.1.1 A new algorithm for simulation

The new Algorithm for simulation is a modified version of the Ogata's FORTRAN program [**etasim.f**]. It requires an input file which contains six lines (see Tab. 5.1).

Table 5.1: Input file for the new algorithm of simulation.

$ic = 0$			b_{val}	
t_{start}			t_{end}	
m_c			m_0	
ϖ	κ	c	a	p
		C_1		
		$idnum$		

The first one contains a label ic and the *b-value* to be given in input. In the Ogata's program, the label ic can be set either zero, indicating that we want to simulate the magnitudes, or one, indicating that we want to take the latter from a specified catalog and we want to simulate only the times. In the case of our model, the choice $ic = 1$ makes no sense, since it is fundamental for us to simulate the magnitudes imposing the hypothesis of magnitude dependence; then, for our simulated program one must set $ic = 0$. The second line of the input file has the initial and ending times of the simulation interval; in the third line there are the completeness and the reference magnitudes; in the fourth and fifth lines we have to write the parameters $(\varpi, \kappa, c, a, p)$ and C_1, respectively, to be given in input; finally, the last line is the "seed" of a uniform pseudo-random numbers generator in $[0, 1]$.

Once given this input text file, the algorithm generates a random number of events between the starting and the ending times specified. Let's set

$$\xi_j(t) = \frac{\kappa}{(t - time_j + c)^p} e^{a(mag_j - mag_0)}$$

and

$$\overline{C}(mag_i) = C_1 \left[1 - 2e^{-(b_{val} \ln 10 - a)(mag_i - m_0)} \right], \tag{5.3}$$

where $time_i$ and mag_i are the time and magnitude of the i^{th} event, respectively.

Now we are ready for the algorithm. It is structured as follows.

Step 1 Set $t = 0$, $x_{int} = \varpi$, $i = 1$.

Step 2 Set $u_{int} = x_{int}$.

Step 3 Generate a uniform random number $U_1 \in (0, 1)$. Put $e = -\log(U_1)/u_{int}$ and $t = t + e$. If $t > t_{end}$, then go to *Step 11*. Otherwise, if $i > 1$ go to *Step 4*, else set $time_1 = t$ and $indexm_1 = 0$ and go to *Step 9*.

Step 4 Set $x_{int} = \varpi$. Compute $x_{int} = x_{int} + \sum_{j=1}^{i-1} \xi_j(t)$.

Step 5 Set $prob = x_{int}/u_{int}$ and generate a uniform random number $U_2 \in (0, 1)$. If $U_2 > prob$ go to *Step 2*, otherwise set $time_i = t$.

Step 6 Set $im = 0$, $xm_{int} = \varpi$ and $logic = 0$. Generate a uniform random number $U_3 \in (0, 1)$.

Step 7 If $U_3 \geq xm_{int}/x_{int}$, then set $im = im + 1$ and compute $xm_{int} = xm_{int} + \xi_{im}(time_i)$. Otherwise set $indexm_i = im$ and $logic = 1$.

Step 8 If $logic = 0$ go to *Step 7*, else continue.

Step 9 Generate a uniform random number $U_4 \in (0, 1)$. If $im = 0$ then $mag_i = -\log(U_4)/b_{val} + m_0$, else

$$\begin{aligned} mag_i = &\frac{1}{b_{val} \ln 10}\Big[\ln(2|\overline{C}(mag_i)|) - \ln\Big(\Big|1 + \overline{C}(mag_i) \\ &- \sqrt{(1 + \overline{C}(mag_i))^2 - 4U_4\overline{C}(mag_i)}\Big|\Big)\Big] + m_0. \end{aligned}$$

For the derivation of this formula see Lemma 4 below.

Step 10 Set $u_{int} = x_{int} + \xi_i(time_i)$, $i = i + 1$ and go to *Step 3*.

Step 11 Set $tt = time_i + e$.

Let's notice that the steps from 1 to 5 corresponds to the simulation of times. This part doesn't change from the Ogata's program. The different part is the consecutive one, in which at first we assign a mother event to the i^{th} shock considered. Then, we generate randomly the magnitudes in the following way. If the mother event is of background type, we use the Gutenberg-Richter law. Otherwise, we use the cumulative distribution function corresponding to $p(m''|m')$, in which appears the magnitude of the mother event found. For the FORTRAN code, see the section **Algorithm 3** in the file **Algorithms.pdf**.

Lemma 4. *Let U be a random variable uniform in $(0,1)$ and set*

$$\overline{C}(m') := C_1[1 - 2e^{-(\beta-a)(m'-m_0)}].$$

Then, the random variable

$$M := \frac{1}{\beta}\left[\ln|2\overline{C}(m')| - \ln\left|1 + \overline{C}(m') - \sqrt{(1+\overline{C}(m'))^2 - 4U\overline{C}(m')}\right|\right] + m_0 \tag{5.4}$$

has probability density function $p_M(m'')$ given by (4.2), i.e., we have $p_M(m'') = p(m''|m')$, where

$$\begin{aligned} p_M(m'') &= p(m''|m') \\ &= \beta e^{-\beta(m''-m_0)}\left[1 + C_1\big(1 - 2e^{-(\beta-a)(m'-m_0)}\big)\big(1 - 2e^{-\beta(m''-m_0)}\big)\right]. \end{aligned}$$

Proof. Let's compute the integral $\int_{m_0}^{m} p(m''|m')dm''$:

$$\begin{aligned} \tilde{F}_{m'}(m) = &\int_{m_0}^{m} \beta e^{-\beta(m''-m_0)}\Big[1 + C_1\big(1 - 2e^{-(\beta-a)(m'-m_0)}\big) \\ &\cdot\big(1 - 2e^{-\beta(m''-m_0)}\big)\Big]dm'' \\ =&\left[1 + C_1\big(1 - 2e^{-(\beta-a)(m'-m_0)}\big)\right]\int_{m_0}^{m} \beta e^{-\beta(m''-m_0)}dm'' \\ &- C_1\big(1 - 2e^{-(\beta-a)(m'-m_0)}\big)\int_{m_0}^{m} 2\beta e^{-2\beta(m''-m_0)}dm'' \\ =&(1+\overline{C}(m'))\big[1 - e^{-\beta(m-m_0)}\big] - \overline{C}(m')\big[1 - e^{-2\beta(m-m_0)}\big]. \end{aligned} \tag{5.5}$$

We notice that $\overline{C}(m') < C_1 \leq 1$. Now, the random variables $\tilde{F}_{m'}^{-1}(1-U)$, where $\tilde{F}_{m'}^{-1}(\cdot)$ is the inverse function of $\tilde{F}_{m'}(\cdot)$, has the desired distribution, since the random variable $1-U$ is uniform in $(0,1)$. Then, in order to compute it, we use the expression of $\tilde{F}_{m'}$ obtained in the last row of (5.5): setting $x(m) = e^{-\beta(m'-m_0)}$, and $\overline{C} = \overline{C}(m')$, for each $U \in (0,1)$, we need to find an m such that

$$\overline{C}x(m)^2 - (1+\overline{C})x(m) + 1 = 1 - u \quad SSE \quad \overline{C}x(m)^2 - (1+\overline{C})x(m) = u$$

Using the uniform random variable $1-U$ in place of U, we find the solutions in $x = x(m)$:

$$x_1 = \frac{1+\overline{C}-\sqrt{(1+\overline{C})^2-4U\overline{C}}}{2\overline{C}}, \quad x_2 = \frac{1+\overline{C}+\sqrt{(1+\overline{C})^2-4U\overline{C}}}{2\overline{C}}.$$

We can easily deduce that:

- $0 < x_1 < 1$ for all $\overline{C} < 1$;
- if $0 < \overline{C} < 1$ then $x_2 > 1$ and if $\overline{C} < 0$ then $x_2 < 0$, too.

Since $0 < e^{-\beta(m-m_0)} < 1$, we choose the first solution. Then, we get

$$e^{-\beta(m-m_0)} = \frac{1+\overline{C}-\sqrt{(1+\overline{C})^2-4U\overline{C}}}{2\overline{C}}$$

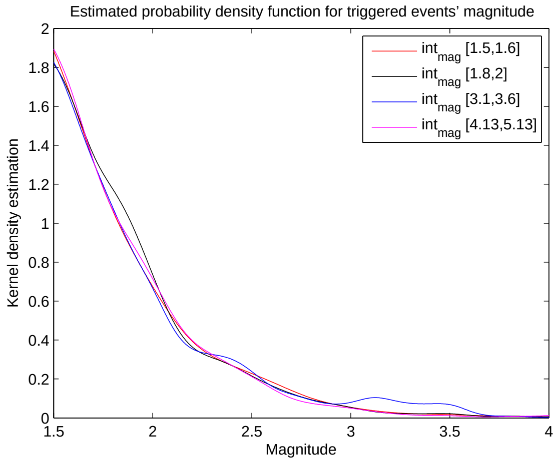
$$\Updownarrow$$

$$m = \frac{1}{\beta}\left[\ln|2\overline{C}| - \ln\left|1+\overline{C}-\sqrt{(1+\overline{C})^2-4U\overline{C}}\right|\right] + m_0.$$

Recalling that $\beta = b\ln 10$, we have thus obtained equation (5.4). □

5.2 Results

We analyze here the results obtained for the two simulated catalogs. Fig. 5.3 contains the estimated densities of triggered events' magnitude, obtained by the first and the second types of analysis (plots at the top and at the bottom, respectively), relative to the catalog simulated with the classical Ogata's model. We recall that, in this case, the magnitudes are simulated randomly (i.e., independently of each others) with the Gutenberg-Richter law. According to this law, triggered events' magnitudes aren't correlated with their respective mothers' magnitudes. This is reflected in the absence of variations of the densities in the four magnitude subintervals considered, both in the plots at the top and at the bottom (first and second types of analysis, respectively). The results are very similar to the ones concerning the Italian catalog.



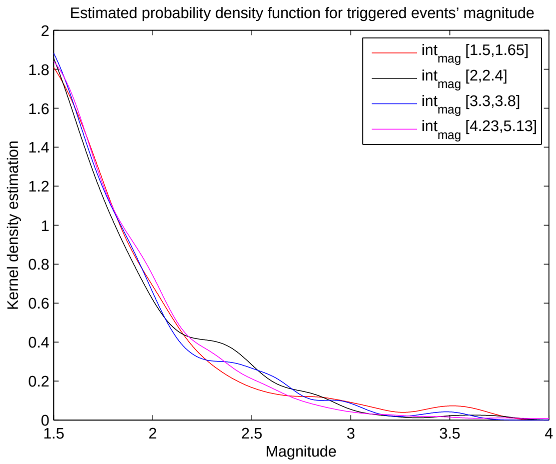


Figure 5.3: Kernel density estimation of triggered events' magnitude in the catalog simulated with the classical Ogata's model, concerning the first and the second types of analysis (plots at the top and at the bottom, respectively). The considered intervals in which triggering events' magnitudes fall are the following. First analysis, plot at the top: $[1.5, 1.6]$, $[1.8, 2]$, $[3.1, 3.6]$ and $[4.13, 5.13]$; second analysis, plot at the bottom: $[1.5, 1.65]$, $[2, 2.4]$, $[3.3, 3.8]$ and $[4.23, 5.13]$ (in both cases, the curves are red, black, blue and magenta, respectively). The δ^* value is equal to seven days. The optimal bandwidth value for the Normal kernel density estimation is, respectively for the four intervals considered, equal to: 0.14, 0.09, 0.08, 0.11 in the plot at the top and 0.12, 0.11, 0.1, 0.1 in the one at the bottom.

The parameters, estimated by setting the precursory at about 10%, are

$$(\varpi, \kappa, c, a, p) = (0.62, 0.02, 0.013, 1.72, 1.11). \tag{5.6}$$

The above parameters can be compared with the ones in (5.1), which are used as input for the simulation.

Performing the residual analysis, for which it is considered the residual process obtained with the random time-change (2.26) in order to get a stationary Poisson process with rate one (see Subsection 2.3.3), we obtain good results: the probabilities for the Kolmogorov-Smirnov and the Runs tests are bigger that 5% (see Tab. 5.2 and Fig. 5.4).

Table 5.2: Results of the tests obtained with the residual analysis, concerning the catalog simulated with the classical Ogata's model. The parameters considered are the ones obtained for a precursory period set at about 10%.

Number of events expected by the model	3121.46
Number of events without the learning period	3144
Runs test	6.03E-001
Kolmogorov-Smirnov test	6.74E-001

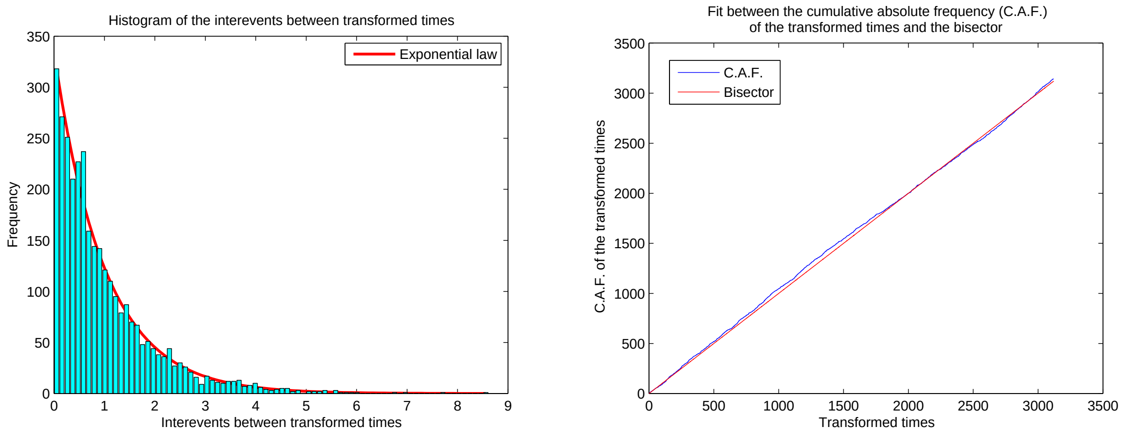
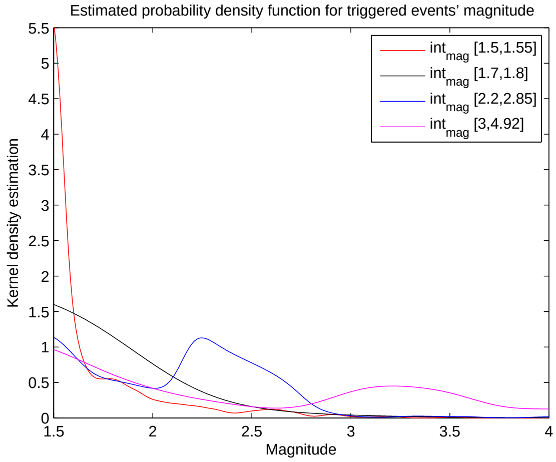


Figure 5.4: Plots concerning the catalog simulated with the classical Ogata's model. In the left plot the histogram of the interevent times of the transformed values is shown, together with the standard exponential law. The fit is good. In the right plot one can see how much the cumulative distribution function of the transformed times varies with respect to the bisector. There is a very good agreement.

A result that instead strongly supports our hypothesis of magnitude correlation is the one obtained in Fig. 5.5, where the estimated densities behaves exactly the same as the for the real data of catalog two, three and four in Chapter 3. This strongly supports our hypothesis.



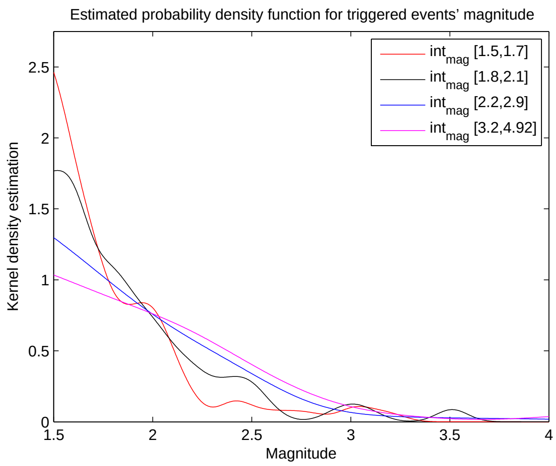


Figure 5.5: Kernel density estimation of triggered events' magnitude in the catalog simulated with our conditional model, concerning the first and the second types of analysis (plots at the top and at the bottom, respectively). The considered intervals in which triggering events' magnitudes fall are the following. First analysis, plot at the top: $[1.5, 1.55]$, $[1.7, 1.8]$, $[2.2, 2.85]$ and $[3, 4.92]$; second analysis, plot at the bottom: $[1.5, 1.7]$, $[1.8, 2.1]$, $[2.2, 2.9]$ and $[3.2, 4.92]$ (in both cases, the curves are red, black, blue and magenta, respectively). The δ^* value is equal to one day. The optimal bandwidth value for the Normal kernel density estimation is, respectively for the four intervals considered, equal to: 0.05, 0.34, 0.07, 0.18 in the plot at the top and 0.08, 0.08, 0.26, 0.28 in the one at the bottom.

In the above Fig. 5.5 we show the estimated densities of triggered events' magnitude for the catalog simulated with our new model. That is, the catalog in which the magnitudes are computed with the transition probability density function (4.37). Once simulated this catalog, we estimate again the parameters with the classical Ogata's FORTRAN program [**etas.f**], fixing again the learning period at about 10%. We get

$$(\varpi, \kappa, c, a, p) = (0.58, 0.022, 0.017, 0.83, 1.12). \tag{5.7}$$

The parameters in (5.7) can be compared with the one used as input for the simulation, i.e., the ones in (5.2).

The results of the residual analysis for the stationary Poisson residual process (with rate one), obtained by considering the random time-change (2.26), are shown in Tab. 5.3 and Fig. 5.6. These results highlight that the set of parameters used is good to fit the phenomenon. In fact, both for the Kolmogorov-Smirnov and the Runs tests we obtain a probability bigger that 5%. The number of expected events is also very close to the one of the events in the target period.

Table 5.3: Results of the tests obtained with the residual analysis, concerning the catalog simulated with our new model. The parameters considered are the ones obtained for a precursory period set at about 10%.

Number of events expected by the model	2390.14
Number of events without the learning period	2387
Runs test	1.32E-001
Kolmogorov-Smirnov test	8.7E-001

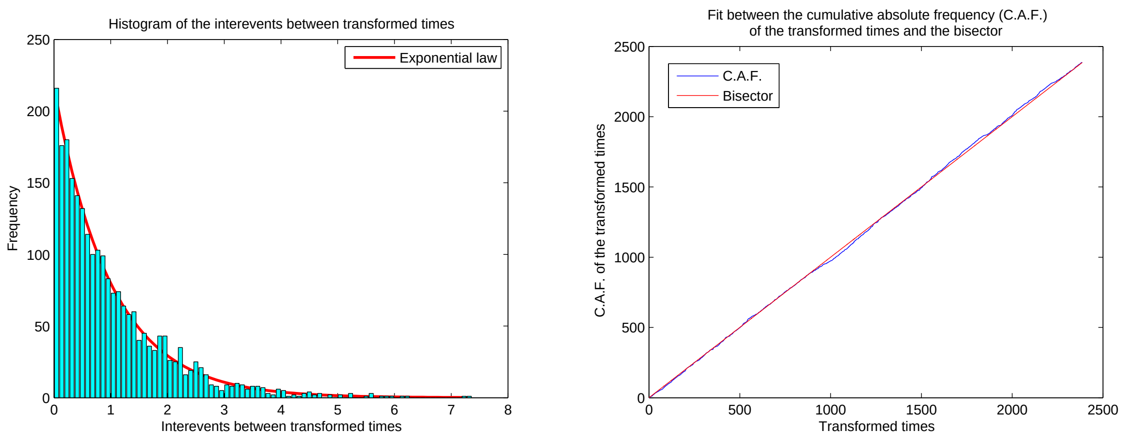


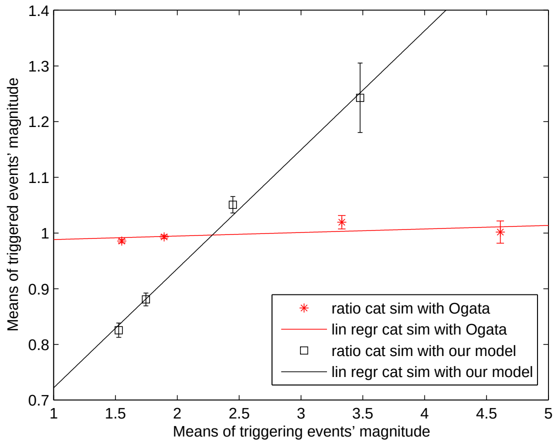
Figure 5.6: Plots concerning the catalog simulated with our new model. In the left plot the histogram of the interevent times of the transformed values is shown, together with the standard exponential law. The fit is good. In the right plot one can see how much the cumulative distribution function of the transformed times varies with respect to the bisector. There is really a very good agreement.

Concluding, Fig. 5.7 in the next page contains the lots of the averages of triggered events' magnitudes versus triggering events' magnitudes concerning the simulated catalogs, for the first and the second types of analysis (plots at the top and at the bottom, respectively). Red is used for the catalog simulated with the Ogata's model, while black for the one simulated with our new model. As for the real cases, for each of the two catalogs considered, the four triggered magnitude averages are normalized by the averages of these four mean values. Furthermore, we plot again the results of the linear regression analysis and the error bars. We recall that the lengths of the latter are given by the normalized mean standard errors. Regarding both the plot at the top and the one at the bottom, one can see that there is almost no percentage variation of the triggered events' magnitude in the catalog simulated with the Ogata's model. Instead, a clear increasing trend is evident for the other simulated catalog considered.

By looking at the correlation coefficients R and p-values p_{val} in Tab. 5.4 below, we can deduce that the results obtained are again statistically significant.

Table 5.4: List of correlation coefficients R and p-values p. Catalog 1 and 2 are simulated with the Ogata's model and the with our new model, respectively.

	First analysis		Second analysis	
	$R \simeq$	$p \simeq$	$R \simeq$	$p \simeq$
Catalog 1	0.61	0.38	-0.93	0.07
Catalog 2	0.99	0.002	0.98	0.01



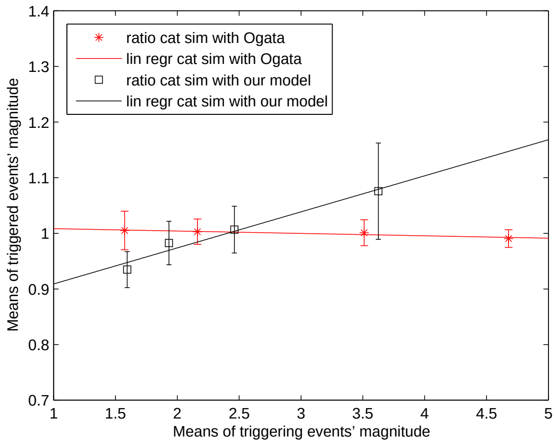


Figure 5.7: Averages of the normalized triggered events' magnitudes. The results are obtained with the first and the second types of analysis (plots at the top and at the bottom, respectively) and are relative to the catalog simulated with the Ogata's model (red) and the one simulated with our new model (black). Regarding the plot at the top, the percentage means for the four subintervals considered for the above two catalogs are the following. Catalog simulated with Ogata's model: 0.9856, 0.9932, 1.0194, 1.0018 (corresponding to the triggering events' magnitude 1.5506, 1.8931, 3.3302, 4.6113, respectively); catalog simulated with our new model: 0.8256, 0.8808, 1.0508, 1.2428 (corresponding to the triggering events' magnitude 1.5268, 1.7467, 2.4491, 3.4779, respectively). Regarding the plot at the bottom, the normalized means for the four subintervals considered for the two catalogs are the following. Catalog simulated with Ogata's model: 1.0052, 1.0029, 1.0011, 0.9907 (corresponding to the triggering events' magnitude 1.5742, 2.163, 3.5112, 4.6786, respectively); catalog simulated with our new model: 0.9349, 0.9827, 1.0067, 1.0757 (corresponding to the triggering events' magnitude 1.595, 1.9324, 2.4614, 3.6254, respectively). The continuous lines correspond to the results of the linear regression and the semi-amplitude of the error bars are the normalized mean standard errors.

5.3 Conclusions

The results obtained with the two types of analysis for the two synthetic catalogs, agree with those obtained with the same kind of analysis performed for the real datasets. More precisely, the catalog simulated by the classical Ogata's model gives the same result of the one concerning the whole Italian catalog. In the latter catalog, the results give a good evidence of magnitude independence, as was expected due to pairs of events temporally close, but spatially very separated. Instead, in the case of the simulated catalog just-mentioned, the variation is absent since it is obtained using the standard Gutenberg-Richter law, which models the magnitudes independently and without conditioning with respect to past seismicity. On the other hand, the results of the statistical analysis concerning the catalog simulated with our new magnitude model show a clear evidence of the magnitude correlation, qualitatively similar to those of the other real catalogs. In fact, the probability of triggered events with "high" magnitude increases with the mother event's magnitude. Regarding the triggered events' magnitude means, we can see again an agreement between synthetic and real data. More precisely, the means have an increasing trend only in the case of the catalog simulated with our new model. Furthermore, recalling that the explicit form of the probability density function $p(m''|m')$ obtained in Chapter 4 has been used to simulate the magnitudes of the second synthetic catalog above described, we deduce that this particular choice for $p(m''|m')$ is good to model the magnitudes of the triggered events in a catalog, when considering past seismicity.

Chapter 6

Mathematical characterization of the new ETAS model with correlated magnitudes

In this chapter we focus on the mathematical characterization of our new version of the ETAS model: the Epidemic Type Aftershock Sequence model with correlated magnitudes, described in Chapter 4. As already explained in the latter, in this model we assume that: the background events' magnitudes are distributed according to the Gutenberg-Richter law, that is an exponential distribution, and the magnitudes of the triggered events are modeled by a probability density function that depends on the mother events' magnitudes and with respect to which the Gutenberg-Richter law itself remains invariant. Thanks to this property, the well-known validity of the latter law is still respected when not taking into account past seismicity.

The theoretical analysis of the model allows to study the behavior of the interevent times, that are the times between consecutive shocks. More precisely, we will consider the variable associated with a single interevent time. In fact, due to the characteristics of the branching process involved and the homogeneity of the background component, one can deduce that interevent times have all the same distribution. The motivation for the study of this random variable lies in the fact that it is of interest for results concerning seismic hazard. Indeed, in the recent literature, much attention has been paid on the time delay between successive events [Bak et al., 2002; Corral, 2003, 2004; Davidsen and Goltz, 2004; Molchan, 2005].

In order to study the distribution of the interevent time, let's start saying that in what follows we will use the terms "observable events" to indicate the shocks with a magnitude higher than the completeness threshold m_c. We recall in fact that, for each seismic catalog, one can consider a completeness magnitude such that, only the events with magnitude above this threshold, are surely recorded in the catalog itself. As already explained, the completeness magnitude is the minimum value after which there's agreement with the exponential Gutenberg-Richter law: since it is expected that the magnitude process is well modeled by the latter law, the deviations from the exponential decay are interpreted like missing measurements.

We will use here the notation $N(\tau)$ for the total number of observable events in $[0, \tau]$, zero included. Let's notice that the same notation $N(t)$ has been used in Chapter 1, but to indicate the number of all the events in $(0, t]$. In order not to introduce one more notation, we decided to use the same symbol here, but we stress that in this chapter the events counted in $N(\tau)$ are only the observable ones and zero is included in the time interval considered. Furthermore, with the notations of Section 4.2, it holds

$$N(\tau) = N([0, \tau] \times [m_c, \infty)).$$

Now, to the aim of finding the interevent time's distribution, let's consider the probability generating function (PGF) of the above random number $N(\tau)$, that is the quantity

$$G_{N(\tau)}(z) = \mathbb{E}[z^{N(\tau)}] =: \Omega(z; \tau). \tag{6.1}$$

The value of this function in zero gives the probability $\mathbf{P}\{\tau\}$ of having zero observable events in $[0, \tau]$. The interevent time density is then obtained through the Palm theory (see equation (1.13) in Subsection 1.5.1, Chapter 1) by a double time derivative of $\mathbf{P}\{\tau\}$ and scaling [Cox and Isham, 2000].

In what follows, we will assume that $\bar{\lambda}$ is the average rate of the total observable events and ϖ is the average rate corresponding to the events of the spontaneous component. Since we recall that the branching ratio is the average number of first generation aftershocks for a given triggering event and that it is equal for both the classical ETAS and the new version of the ETAS we are studying (see the last part of Section 4.1), it holds

$$\bar{\lambda} = \bar{\lambda}(m_c) = \varpi \int_{m_c}^{\infty} p(m)dm \sum_{k=0}^{\infty} \eta^k = \frac{\varpi}{1-\eta} \int_{m_c}^{\infty} p(m)dm. \tag{6.2}$$

Remark 12. The above expression follows because, repeating similar computations as in Remark 6, switching the integrals where necessary and recalling that $\mathcal{K}_2 p(m) = \int_{m_0}^{\infty} p(m')\varrho(m')p(m|m')dm' = \eta p(m)$, it holds

$$
\begin{aligned}
&\mathbb{E}\Big[N_i^{tr,1}(\mathbb{R}\times[m_c,\infty))|S_i=s\Big]\\
&=\int_{m_0}^{\infty} dmp(m)\mathbb{E}\Big[N_i^{tr,1}(\mathbb{R}\times[m_c,\infty))\Big|S_i=s,M_i=m\Big]\\
&=\int_{m_0}^{\infty} dmp(m)\int_{s}^{\infty} dt_1\int_{m_0}^{\infty} dm_1\varrho(m)p(m_1|m)\mathbb{1}_{[m_c,\infty)}(m_1)\\
&=\int_{m_0}^{\infty} dmp(m)\int_{m_c}^{\infty} dm_1\varrho(m)p(m_1|m)\\
&=\int_{m_c}^{\infty} dm_1\mathcal{K}_2p(m_1)\\
&=\eta\int_{m_c}^{\infty} dm_1p(m_1),
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}\Big[N_i^{tr,2}(\mathbb{R}\times[m_c,\infty))|S_i=s\Big]\\
&=\int_{m_0}^{\infty} dmp(m)\mathbb{E}\Big[N_i^{tr,2}(\mathbb{R}\times[m_c,\infty))\Big|S_i=s,M_i=m\Big]\\
&=\int_{m_0}^{\infty} dmp(m)\int_{s}^{\infty} dt_1\int_{m_0}^{\infty} dm_1\int_{t_1}^{\infty} dt_2\int_{m_0}^{\infty}\Phi(t_1-s)\Phi(t_2-t_1)\varrho(m)\\
&\quad\cdot p(m_1|m)\varrho(m_1)p(m_2|m_1)\mathbb{1}_{[m_c,\infty)}(m_2)dm_2\\
&=\int_{m_0}^{\infty} dmp(m)\int_{m_0}^{\infty} dm_1\varrho(m)p(m_1|m)\int_{m_c}^{\infty}\varrho(m_1)p(m_2|m_1)dm_2\\
&=\int_{m_c}^{\infty} dm_2\int_{m_0}^{\infty} dm_1\mathcal{K}_2p(m_1)\varrho(m_1)p(m_2|m_1)\\
&=\int_{m_c}^{\infty} dm_2\mathcal{K}_2^{(2)}p(m_2)\\
&=\eta^2\int_{m_c}^{\infty} dm_2p(m_2)
\end{aligned}
$$

and then

$$\mathbb{E}\Big[N_i^{tr,2}(\mathbb{R}\times[m_c,\infty))|S_i=s\Big]=\eta^n\int_{m_c}^{\infty} dm_2p(m_2).$$

Since the average rate of the total number of observable events is the

$$\varpi\int_{m_c}^{\infty} dmp(m)+\sum_{n=1}^{\infty}\mathbb{E}\Big[N_i^{tr,n}(\mathbb{R}\times[m_c,\infty))|S_i=s\Big],$$

the expression (6.2) is obtained.

It is important to notice that the total average rate will be used only for the final step concerning the derivation of the interevent time density; instead, the average rate of the spontaneous events will be used straight away for the derivation of the above probability generating function.

Finally, we will consider here the parametrization (2.5) in Chapter 1 of the Omori-Utsu law for the rate of first generation triggered events from a given earthquake, with $p = \theta + 1$. More precisely, we will consider

$$\Phi(t) = \frac{\theta c^{\theta}}{(c+t)^{1+\theta}}, \quad t > 0, \tag{6.3}$$

with $\theta > 0$. Let's recall that this law can be interpreted as the probability density function of random times at which first generation shocks independently occur, when the mother event has occurred in $t = 0$.

6.1 Probability generating function for earthquake sequences

In this section, we want to derive the probability generating function $\Omega(\cdot;\tau)$ of the total number $N(\tau)$ of observable events in $[0;\tau]$.

As explained in Chapter 4, this PGF could be obtained by using the Laplace functional (see Subsection 4.2.1). Nevertheless, we will follow here a different approach, based on the properties of the probability generating function. In fact, in the case of seismic modeling we are studying, the PGF $\Omega(\cdot;\tau)$ can be obtained also without considering the Laplace functional and following a less technical and more intuitive method, based on approximation arguments. Obviously, the final results are the same as that obtained with the Laplace functional. In this way, we will also have illustrated two different approaches for the same goal.

Let's then proceed with the derivation of the probability generating function $\Omega(\cdot;\tau)$. In what follows, we will indicate with $G(\cdot;v|s,m')$ the PGF of the number of events with magnitude larger than m_c, triggered in a generic interval $[0,v]$ by a triggering event with magnitude m' occurred in a time $-s$, where $s \geq 0$. Recalling the notation in Section 4.2.1 of Chapter 4, it holds

$$G(\cdot;v|s,m') = G^{tr}(\cdot;v|-s,m'), \quad \text{for } s \geq 0.$$

This new notation will allow us to compute integrals always in positive time intervals. We anticipate also that, as we are going to see, in what follows the argument s in $G(\cdot;v|s,m')$ will be either positive, and then the time of occurrence of the triggering shock will be $-s$, or zero.

Theorem 11. *Let's consider the seismic process described by the ETAS model with correlated magnitudes, completely characterized by the conditional intensity* (4.3)*. The probability generating function* $\Omega(\cdot;\tau)$ *of the total number* $N(\tau)$ *of observable events belonging to* $[0;\tau]$ *is*

$$\Omega(z;\tau) = e^{-\varpi L(z;\tau)}, \tag{6.4}$$

where

$$\begin{aligned} L(z;\tau) = &\int_0^\infty \left[1 - \int_{m_0}^\infty G(z;\tau|t,m')p(m')dm'\right]dt \\ &+ \int_0^\tau \left[1 - \int_{m_0}^{m_c} G(z;t|0,m')p(m')dm' \right.\\ &\left. - z\int_{m_c}^\infty G(z;t|0,m')p(m')dm'\right]dt, \end{aligned} \tag{6.5}$$

and furthermore

$$\begin{aligned} -\frac{\ln G(z;\tau|t,m')}{\varrho(m')} =& b(t+\tau) - \int_{m_0}^\infty (\Phi(\cdot)\otimes G(z;\tau|\cdot,m''))(t)p(m''|m')dm'' \\ &- \int_{m_0}^{m_c} (\Phi(t+\cdot)\otimes G(z;\cdot|0,m''))(\tau)p(m''|m')dm'' \\ &- z\int_{m_c}^\infty (\Phi(t+\cdot)\otimes G(z;\cdot|0,m''))(\tau)p(m''|m')dm''. \end{aligned} \tag{6.6}$$

In the previous equation, $\otimes$ *is the convolution operator and*

$$b(t) = \int_0^t \Phi(x)dx = 1 - \frac{c^\theta}{(c+t)^\theta}. \tag{6.7}$$

Before giving the proof, let's say that the above theorem allows us to derive the probability of having zero observable events in $[0,\tau]$, that is our main aim. In fact, the just-mentioned probability is obtained by substituting equation (6.6) in (6.5) opportunely and recalling that

$$\mathbf{P}\{\tau\} = \Omega(0;\tau) = e^{-\varpi L(0;\tau)}.$$

In particular, it will be useful the function

$$
\begin{aligned}
-\frac{\ln G(0;\tau|t,m')}{\varrho(m')} =& b(t+\tau) - \int_{m_0}^{\infty} (\Phi(\cdot)\otimes G(0;\tau|\cdot,m''))(t)p(m''|m')dm'' \\
&- \int_{m_0}^{m_c} (\Phi(t+\cdot)\otimes G(0;\cdot|0,m''))(\tau)p(m''|m')dm''.
\end{aligned}
\tag{6.8}
$$

In the next sections, we will work separately on two cases. Firstly on the particular case $m_c = m_0$, then on the general one $m_c \geq m_0$.

Let's then conclude the current section by proving Theorem 11.

Proof of Theorem 11. We are interested in $N(\tau)$, that is the number of all the observable events in $[0,\tau]$ (spontaneous and triggered). Let's consider at first the observable triggered events, whose total number in $[0,\tau]$ is indicated with $N_1(\tau)$. We can associate to each of them the spontaneous shock from which the branching process reaches that event. Let's then consider all the spontaneous events. The triggered events can be partitioned according to their associated spontaneous triggering shocks. More precisely, the total number $N_1(\tau)$ of observable aftershocks in $[0,\tau]$ will be the sum of the observable shocks triggered by each spontaneous event. Let's consider now the following non-overlapping small subintervals of $(-\infty,\tau]$: for all $k\in\mathbb{Z}$, $k\leq n$,

$$
\begin{aligned}
I_k &= [x_{k-1}, x_k), \quad \text{where } x_0 = 0, x_k - x_{k-1} = \Delta x, \forall k < n \\
I_n &= [x_{n-1}, x_n], \quad \text{where } x_n = \tau.
\end{aligned}
$$

It obviously holds that $I_k \subseteq (-\infty, 0)$ if $k \leq 0$, instead $I_k \subseteq [0,\tau]$ if $k = 1,\dots,n$. The probability generating function of the number $N_1(\tau)$ can be obtained as the limit, with respect to Δx, of the corresponding probability generating function relative to the above subintervals.

Recall now both the property of the PGF expressed in equation (A.2), Appendix A.1, and the fact that there's independence between the numbers of aftershocks triggered by spontaneous events, the latter belonging to disjoint subintervals. We can therefore deduce that, setting

$$N_k^{tr}(\tau) = \#\{\text{triggered events in } [0,\tau] \text{ with mother in } I_k\},$$

$$N_k^{sp} = \#\{\text{spontaneous events in } I_k\},$$

and, if $s_k \in I_k$,

$$N_{k,s_k}^{tr}(\tau) = \#\{\text{triggered events in } [0,\tau] \text{ with mother occurred in } s_k\},$$

the PGF of the total number of observable events triggered in $[0,\tau]$, i.e.,

$$N_1(\tau) = \sum_{k\le 0} N_k^{tr}(\tau) + \sum_{k=1}^{n} N_k^{tr}(\tau),$$

is the limit of the product between the respective probability generating functions relative to different subintervals, i.e.,

$$G_{N_1(\tau)}(z) = \lim_{\Delta x\to 0} \prod_{k\le 0} G_{N_k^{tr}(\tau)}(z) \prod_{k=1}^{n} G_{N_k^{tr}(\tau)}(z),$$

Let's focus now on $G_{N_k^{tr}(\tau)}(z)$. Thanks to the properties of the probability generating function, for Δx tending to zero, this PGF is obtained by composing the PGF of the number N_k^{sp} of background events occurred in the generic subinterval considered I_k, with the PGF relative to the number $N_{k,s_k}^{tr}(\tau)$ of events in $[0,\tau]$ triggered by just one spontaneous event, that is for example s_k, (see Proposition 6, Appendix A.1), i.e., as $\Delta x \to 0$,

$$G_{N_k^{tr}(\tau)}(z) \approx G_{N_k^{sp}}[G_{N_{k,s_k}^{tr}(\tau)}(z)].$$

Finally, in order to obtain the probability generating function of the total number $N(\tau)$ of observable events in $[0,\tau]$, we have to take into account the spontaneous events, too. Obviously, the background events to be considered in $N(\tau)$ are only those which fall in subintervals I_k for $k = 1,\dots,n$, that are $I_k \subseteq [0,\tau]$. In this case, for each of these events we will count all the shocks triggered by it, plus the event itself. This latter is equivalent to multiply by z the probability generating function of the number of events triggered in $[0,\tau]$ by a single spontaneous event of magnitude m', itself belonging to $[0,\tau]$. More precisely, since this event has to be observable, the multiplication by z will be done only if its magnitude is larger than the threshold m_c. Therefore, instead of the multiplication by z, we will multiply by

$$f(z,m',m_c) = \mathbb{1}_{[0,m_c)}(m') + z\mathbb{1}_{[m_c,+\infty)}(m'), \tag{6.9}$$

where we recall that $\mathbb{1}_A(x)$ is the indicator function having value one when $x \in A$, zero otherwise.

Before proceeding, it is useful to notice that, for $s \in [0,\tau]$, it holds

$$\begin{aligned}
\mathbb{E}\left[z^{N_i([0,\tau]\times[m_c,\infty))}\middle| S_i = s, M_i = m\right] &= \mathbb{E}\left[z^{N_i([s,\tau]\times[m_c,\infty))}\middle| S_i = s, M_i = m\right] \\
&= \mathbb{E}\left[z^{N_i([0,\tau-s]\times[m_c,\infty))}\middle| S_i = 0, M_i = m\right] \\
&= G(z;\tau - s|0,m),
\end{aligned}$$

where $N_i([0,\tau]\times[m_c,\infty))$ is the number of events in $[0,\tau]$, with magnitude in $[m_c,\infty)$, triggered by the i^{th} event occurred in $S_i = s$ and with magnitude $M_i = m$. Then, when we will have to integrate the PGF

$$\mathbb{E}\Big[z^{N_i([0,\tau]\times[m_c,\infty))}\Big|S_i = s, M_i = m\Big]$$

in $[0,\tau]$, we will consider the integral

$$\int_0^\tau G(z;\tau-s|0,m)ds = \int_0^\tau G(z;t|0,m)dt,$$

obtained by the change of variable $t = \tau - s$.

Let's now compute explicitly the probability generating function $\Omega(\cdot;\tau)$ of $N(\tau)$. As anticipated before in this section, we have to consider the number of events with magnitude larger than m_c, triggered in a generic interval $[0,\upsilon]$ by a background event with magnitude m' occurred in a time $-s$, where $s \geq 0$. We have already specified that $G(\cdot;\upsilon|s,m')$ is the PGF of such a number. Recalling that the random variable associated with the number of spontaneous events in a generic subinterval of length Δx follows a Poisson distribution with parameter $\varpi\Delta x$, we have that, for fixed $-s_k \in I_k \subseteq (-\infty,0)$ and $w_k \in I_k \subseteq [0,\tau]$,

$$\begin{aligned}
\Omega(z;\tau) = &\lim_{\Delta x\to 0}\prod_{k\leq 0}\exp\Big\{-\varpi\Delta x[1-G(z;\tau|s_k,m')]\Big\}\\
&\cdot\prod_{k=1}^{n}\exp\Big\{-\varpi\Delta x[1-f(z,m',m_c)G(z;\tau-w_k|0,m')]\Big\}\\
=&\exp\Big\{-\varpi\Big[\Big(\int_{-\infty}^{0}[1-G(z;\tau|-t',m')]dt'\\
&+\int_0^\tau[1-f(z,m',m_c)G(z;\tau-t'|0,m')]dt'\Big)\Big]\Big\}\\
=&\exp\Big\{-\varpi\Big[\Big(\int_0^\infty[1-G(z;\tau|t,m')]dt\\
&+\int_0^\tau[1-f(z,m',m_c)G(z;t|0,m')]dt\Big)\Big]\Big\}.
\end{aligned}\qquad(6.10)$$

The formulae obtained till here are relative to observable events with magnitude larger than m_c, triggered by a spontaneous shock with a fixed but generic magnitude equal to m'. In order to consider all the possible values for the spontaneous event's magnitude, we have to integrate (6.10) also with respect to m'. More precisely, this integral must be computed between m_0 and infinity, in fact we recall that m_0 is the minimum value for an event to be able to trigger its own aftershocks. Furthermore, since the integration concerns the spontaneous events' magnitude, the density to be considered is the Gutenberg-Richter law (equation (2.2)). Then, with abuse of notation because of the further integral with respect to m' not yet considered in (6.10), we can rewrite the probability generating function $\Omega(z;\tau)$ of the total number of observable events in $[0,\tau]$, as

$$\Omega(z;\tau) = e^{-\varpi L(z;\tau)},$$

where we have set

$$\begin{aligned} L(z;\tau) = \int_{m_0}^{\infty} dm' p(m') \Big(& \int_0^{\infty} [1 - G(z;\tau|t,m')]dt \\ & + \int_0^{\tau} [1 - f(z,m',m_c)G(z;t|0,m')]dt \Big). \end{aligned} \tag{6.11}$$

Let's substitute now the above-cited function $f(z,m',m_c)$ (equation (6.9)) in the last expression of (6.11). Recalling that the reference magnitude m_0 is usually set less than or equal to the completeness one m_c [Sornette and Werner, 2005], in fact one can reduce to this situation not considering the data relative to observable events with magnitude less than m_0, we obtain

$$\begin{aligned} L(z;\tau) &= \int_{m_0}^{\infty} dm' p(m') \Big(\int_0^{\infty} [1 - G(z;\tau|t,m')]dt \\ &\quad + \int_0^{\tau} \Big[1 - [\mathbb{1}_{[0,m_c)}(m') + z\mathbb{1}_{[m_c,+\infty)}(m')]G(z;t|0,m')\Big]dt \Big) \\ &= \int_0^{\infty} \Big[1 - \int_{m_0}^{\infty} G(z;\tau|t,m')p(m')dm'\Big]dt \\ &\quad + \int_0^{\tau} \Big[1 - \int_{m_0}^{m_c} G(z;t|0,m')p(m')dm' \\ &\quad - z\int_{m_c}^{\infty} G(z;t|0,m')p(m')dm'\Big]dt. \end{aligned}$$

At this point, we want to compute explicitly the above probability generating function $\Omega(\cdot;\tau)$. To do that, let's find an expression for $G(z;\tau|t,m')$ that is, as said before, the PGF of the number of observable events in $[0,\tau]$ triggered by a fixed spontaneous shock with magnitude m' occurred in $-t$, where $t \geq 0$. Let's consider now any observable event triggered in $[0,\tau]$ belonging to any generation from the second one on and let $N_2(\tau)$ denote their total number. As before, we can associate it with a first generation triggered shock occurred between the spontaneous one and the considered event itself. The number $N_2(\tau)$ will be the sum of the observable events triggered by each first generation triggered shock. Thanks to the auto-similarity of the process, the above-mentioned probability generating function $G(z;\cdot|\cdot,\cdot)$ is the same for both a background event and a triggered one. Therefore, as previously done, we obtain that, for fixed $-s_k \in I_k \subseteq [-t,0)$ and $w_k \in I_k \subseteq [0,\tau]$,

$$\begin{aligned} G(z;\tau|t,m') = \lim_{\Delta x\to 0} \prod_{k\leq 0: I_k\subset[-t,0)} &\exp\left\{\Phi(-s_k+t)\Delta x \varrho(m')[G(z;\tau|s_k,m'')-1]\right\} \\ \cdot \prod_{k=1}^{n} \exp\left\{\Phi(w_k+t)\Delta x\varrho(m')[f(z,m'',m_c)G(z;\tau-w_k|0,m'')-1]\right\}, \end{aligned} \tag{6.12}$$

where m'' is the magnitude of the first generation triggered shock. Recall that $\Phi(\cdot)$ is the Omori-Utsu law and $\varrho(\cdot)$ is the productivity one (respectively equations (6.3) and (2.1)). Again, we have to average over the first generation triggered shock's magnitude. To this end, we integrate equation (6.12) with respect to m''. Since any first generation triggered event considered is also a triggering one, the just-mentioned integral must be computed again between the reference magnitude m_0 and infinity. Furthermore, we have to consider now the transition probability density function of the first generation triggered shock's magnitude m'' with respect to the background event's magnitude m'. More precisely, we have to consider the model $\mathbb{P}\{m''|m'\} = p(m''|m')$ for the magnitude m'', for which we have obtained an explicit form in Chapter 4 (see (4.37)). Then, recalling equation (6.9) for $f(z,m'',m_c)$, again with abuse of notation because of the further integration with respect to first generation aftershock's magnitude m'', we can rewrite (6.12) as:

$$\begin{aligned} -\frac{\ln G(z;\tau|t,m')}{\varrho(m')} &= \int_{m_0}^{\infty} dm'' p(m''|m') \Bigg[\int_{-t}^{0} \Phi(t+x)[1-G(z;\tau|-x,m'')]dx \\ &+ \int_0^{\tau} \Phi(t+x)[1-f(z,m'',m_c)G(z;\tau-x|0,m'')]dx\Bigg] \\ &= \int_{m_0}^{\infty} dm'' p(m''|m') \Bigg[\int_{-t}^{0}\Phi(t+x)dx - \int_{-t}^{0}\Phi(t+x)G(z;\tau|-x,m'')dx \\ &+ \int_0^{\tau}\Phi(t+x)dx - f(z,m'',m_c)\int_0^{\tau}\Phi(t+x)G(z;\tau-x|0,m'')dx\Bigg] \end{aligned}$$

$$
\begin{aligned}
=&\int_{m_0}^{\infty} dm'' p(m''|m') \bigg[\int_0^t \Phi(x)dx - \int_0^t \Phi(t-x)G(z;\tau|x,m'')dx \\
&+ \int_t^{t+\tau} \Phi(x)dx - f(z,m'',m_c)\int_0^{\tau} \Phi(t+\tau-x)G(z;x|0,m'')dx \bigg] \\
=&\int_{m_0}^{\infty} dm'' p(m''|m') \bigg[b(t+\tau) - (\Phi(\cdot)\otimes G(z;\tau|\cdot,m''))(t) \\
&- f(z,m'',m_c)(\Phi(t+\cdot)\otimes G(z;\cdot|0,m''))(\tau) \bigg] \\
=&b(t+\tau) - \int_{m_0}^{\infty} (\Phi(\cdot)\otimes G(z;\tau|\cdot,m''))(t)p(m''|m')dm'' \\
&- \int_{m_0}^{m_c} (\Phi(t+\cdot)\otimes G(z;\cdot|0,m''))(\tau)p(m''|m')dm'' \\
&- z\int_{m_c}^{\infty} (\Phi(t+\cdot)\otimes G(z;\cdot|0,m''))(\tau)p(m''|m')dm''.
\end{aligned}
$$

□

6.2 The case $m_c = m_0$

In this section, we want to focus on the particular case $m_c = m_0$. We will focus on this case because one can find a closed-form expression for the probability $\mathbf{P}\{\tau\}$ of having zero observable events in $[0, \tau]$, for small values of τ. Then, by using the Palm theory [Cox and Isham, 2000] (see equation (1.13) in Subsection 1.5.1, Chapter 1), we derive a closed-form expression for the density of the interevent time, again for small values of τ. The relevance of the case $m_c = m_0$ can be understood by the following reasoning. On one side, the reference magnitude m_0 is connected to the intrinsic physical features of soil and remains almost constant in a sufficiently long time frame. On the other side, the completeness magnitude is related to the number of seismic stations and their sensitivity. Then, the latter magnitude varies with seismic stations density. Hoping for a gradual improvement of technology, both in instrumental sensitivity and in the number of seismic stations available, we can certainly suppose to reduce m_c till it will be just equal to m_0.

Setting $m_c = m_0$ in the last expression of formula (6.5), which gives the function $L(z;\tau)$ at the exponent of $\Omega(z;\tau)$, and recalling (6.4), we find that the probability generating function $\Omega(z;\tau)$, relative to the total number of observable events in $[0, \tau]$, is

$$\Omega(z;\tau) = \exp\left\{-\varpi\left[\int_0^\infty \left[1 - \int_{m_0}^\infty G(z;\tau|t,m')p(m')dm'\right]dt \right.\right.$$
$$\left.\left. + \int_0^\tau \left[1 - z\int_{m_0}^\infty G(z;t|0,m')p(m')dm'\right]dt\right]\right\}.$$

As already explained, in order to find the probability $\mathbf{P}\{\tau\}$ of having zero observable events in $[0,\tau]$, we need to compute the above function in zero.

Theorem 12. *If $m_c = m_0$ and τ is small, the probability of having zero observable events in $[0,\tau]$ is*

$$\boldsymbol{P}\{\tau\} = \Omega(0;\tau) = e^{-\varpi L(0;\tau)} \approx \exp\left\{-\varpi\tau - \frac{\eta\varpi}{1-\eta}A(\tau)\right\}, \tag{6.13}$$

where

$$\eta = \frac{\beta\kappa}{\beta - a}$$

is the branching ratio (equation (2.18)*) and*

$$A(\tau) = \int_0^\tau a(x)dx = \frac{\left[c^\theta(\tau+c)^{1-\theta} - c\right]}{1-\theta}, \tag{6.14}$$

with

$$a(t) = 1 - b(t) = \int_t^\infty \Phi(x)dx = \frac{c^\theta}{(c+t)^\theta}. \tag{6.15}$$

Proof. If $m_c = m_0$ and $z = 0$, the function $L(z;\tau)$ in (6.5) becomes

$$\begin{aligned} L(0;\tau) &= \int_0^\infty \left[1 - \int_{m_0}^\infty G(0;\tau|t,m')p(m')dm'\right]dt + \int_0^\tau dt && (6.16)\\ &= \int_0^\infty \left[1 - \int_{m_0}^\infty G(0;\tau|t,m')p(m')dm'\right]dt + \tau \\ &= \int_0^\infty \overline{N}_{m_c=m_0}(t,\tau)dt + \tau, && (6.17)\end{aligned}$$

where we have set

$$\overline{N}_{m_c=m_0}(t,\tau) := 1 - \int_{m_0}^\infty G(0;\tau|t,m')p(m')dm'. \tag{6.18}$$

Let's notice that the argument in the second temporal integral of (6.16) is actually the constant 1. In fact, setting $z = 0$ and $m_c = m_0$ in the last temporal integral of (6.5), we get

$$1 - \int_{m_0}^{\infty} G(0;t|0,m')p(m')dm' + \int_{m_0}^{\infty} G(0;t|0,m')p(m')dm' = 1.$$

Then, let's compute in $z = 0$ the probability generating function $G(z;\tau|t,m')$, relative to the number of observable events in $[0;\tau]$ triggered by a fixed background shock with magnitude m' occurred in $-t$, where $t \geq 0$. Setting $m_c = m_0$ in equation (6.8) and considering the approximation of the exponential function for small τ, we obtain

$$\begin{aligned} G(0;\tau|t,m') =& \exp\Bigg\{ -\varrho(m') \Bigg[b(t+\tau) \\ & - \int_{m_0}^{\infty} (\Phi(\cdot) \otimes G(0;\tau|\cdot,m''))(t) p(m''|m')dm'' \Bigg] \Bigg\} \\ \approx& 1 - \varrho(m') \Bigg[b(t+\tau) \\ & - \int_{m_0}^{\infty} (\Phi(\cdot) \otimes G(0;\tau|\cdot,m''))(t) p(m''|m')dm'' \Bigg]. \end{aligned}$$

This approximation in Taylor series around the point zero makes sense since the exponential argument is zero when $\tau = 0$. In fact, $G(0;0|\cdot,m'') = 1$. If we substitute the previous formula in (6.18), we get for small τ that

$$\begin{aligned} \overline{N}_{m_c=m_0}(t,\tau) \approx& 1 - \int_{m_0}^{\infty} dm' p(m') \Bigg\{ 1 - \varrho(m') \Bigg[b(t+\tau) \\ & - \int_{m_0}^{\infty} (\Phi(\cdot) \otimes G(0;\tau|\cdot,m''))(t) p(m''|m')dm'' \Bigg] \Bigg\} \\ =& 1 - 1 + b(t+\tau) \int_{m_0}^{\infty} p(m')\varrho(m')dm' - \int_{m_0}^{\infty} dm' p(m')\varrho(m') \\ & \cdot \int_{m_0}^{\infty} (\Phi(\cdot) \otimes G(0;\tau|\cdot,m''))(t) p(m''|m')dm'' \\ =& \eta b(t+\tau) - \int_{m_0}^{\infty} dm' p(m')\varrho(m') \int_{m_0}^{\infty} dm'' p(m''|m') \\ & \cdot \int_0^t \Phi(x) G(0;\tau|t-x,m'')dx \end{aligned}$$

switching the order of integration

$$
\begin{aligned}
=&\eta b(t+\tau) - \int_0^t dx\, \Phi(x) \int_{m_0}^{\infty} dm'' G(0;\tau|t-x,m'') \\
&\cdot \int_{m_0}^{\infty} p(m')\varrho(m')p(m''|m')dm' \\
=&\eta b(t+\tau) - \int_0^t dx\, \Phi(x) \int_{m_0}^{\infty} G(0;\tau|t-x,m'') I_A(m'') dm'',
\end{aligned}
\tag{6.19}
$$

with

$$
I_A(m'') := \int_{m_0}^{\infty} p(m')\varrho(m')p(m''|m')dm'
$$

and

$$
\eta = \int_{m_0}^{\infty} p(m')\varrho(m')dm' = \frac{\beta\kappa}{\beta - a}.
$$

Let's notice that the previous formula (6.19) contains the model $\mathbb{P}\{m''|m'\} = p(m''|m')$ for the magnitudes of the aftershocks. Indeed, this is the model we hypothesize for the magnitudes of triggered events; an explicit form for it has been proposed in Chapter 4. As we are going to explain in a moment, this explicit form is not used in this section.

Let's notice also that the above integral $I_A(m'')$ is nothing but $\mathcal{K}_2 p(m)$, where $\mathcal{K}_2$ is the operator defined in the proof of Lemma 1.

We state now that the just-mentioned transition probability density function has to verify the following condition, representing a fundamental point for this analysis:

$$
\frac{\int_{m_0}^{\infty} p(m')\varrho(m')p(m''|m')dm'}{\int_{m_0}^{\infty} p(m')\varrho(m')dm'} = p(m'') \quad \Leftrightarrow \quad I_A(m'') = \eta p(m''). \tag{6.20}
$$

It is exactly the condition (4.28) of Chapter 4, written in terms of the integral $I_A(m'')$. In fact, we can recognize that the first member of the left equation in (6.20) is the probability of an event to have magnitude m'', given that it has been triggered (see (4.26)). As already explained, equation (6.20) corresponds to the need of obtaining the Gutenberg-Richter law when averaging over all the triggering event's magnitudes.

Let's notice that the explicit form of $p(m''|m')$ satisfying (6.20), found in Chapter 4, is not used in this section. This is due to the fact that in the above computations the function $p(m''|m')$ appears only in the integral $I_A(m'')$, which reduces to $\eta p(m'')$, as just explained. The explicit form will be instead used in the next section.

The condition (6.20) is crucial for the aim of finding the probability $\mathbf{P}\{\tau\}$ of having zero observable events in $[0,\tau]$. In fact, recalling equation (6.19), we can rewrite the approximation of $\overline{N}_{m_c=m_0}(t,\tau)$ for small τ as

$$\begin{aligned}
\overline{N}_{m_c=m_0}(t,\tau) \approx& \eta b(t+\tau) - \eta \int_0^t dx\,\Phi(x) \int_{m_0}^{\infty} G(0;\tau|t-x,m'')p(m'')dm'' \\
=&\eta\big[b(t+\tau)-b(t)\big] + \eta\Bigg[b(t) \\
&- \int_0^t dx\,\Phi(x) \int_{m_0}^{\infty} G(0;\tau|t-x,m'')p(m'')dm''\Bigg] \\
=&\eta\big[b(t+\tau)-b(t)\big] + \eta\int_0^t dx\,\Phi(x)\Bigg[1 \\
&- \int_{m_0}^{\infty} G(0;\tau|t-x,m'')p(m'')dm''\Bigg] \\
=&\eta\big[b(t+\tau)-b(t)\big] + \eta\int_0^t \Phi(x)\overline{N}_{m_c=m_0}(t-x,\tau)dx, \qquad (6.21)
\end{aligned}$$

where in the last equality we have used formula (6.18).

In order to obtain the expression to substitute in (6.17), let's compute the temporal integral of both the first and the last members of equation (6.21). We get

$$\begin{aligned}
\int_0^{\infty} \overline{N}_{m_c=m_0}(t,\tau)dt \approx& \eta \int_0^{\infty} \big[b(t+\tau)-b(t)\big]dt \\
&+ \eta \int_0^{\infty}\int_0^t \Phi(t-x)\overline{N}_{m_c=m_0}(x,\tau)dxdt \\
&\text{switch the last two temporal integrals} \\
=&\eta \int_0^{\infty} \big[b(t+\tau)-b(t)\big]dt \\
&+ \eta \int_0^{\infty}\int_x^{\infty} \Phi(t-x)\overline{N}_{m_c=m_0}(x,\tau)dtdx \\
=&\eta \int_0^{\infty} \big[b(t+\tau)-b(t)\big]dt + \eta\int_0^{\infty} \overline{N}_{m_c=m_0}(x,\tau)dx, \qquad (6.22)
\end{aligned}$$

where the last equality is obtained because

$$\int_x^\infty \Phi(t-x)dt = \int_0^\infty \Phi(y)dy = 1.$$

From (6.22), we then have

$$\begin{aligned}
\int_0^\infty \overline{N}_{m_c=m_0}(t,\tau)dt &\approx \frac{\eta}{1-\eta}\int_0^\infty \Big[b(t+\tau)-b(t)\Big]dt \\
&= \frac{\eta}{1-\eta}\int_0^\infty \left[1-\frac{c^\theta}{(c+t+\tau)^\theta}-1+\frac{c^\theta}{(c+t)^\theta}\right]dt \\
&= \frac{\eta c^\theta}{1-\eta}\int_0^\infty \left[\frac{1}{(c+t)^\theta}-\frac{1}{(c+t+\tau)^\theta}\right]dt \\
&= \frac{\eta c^\theta}{(1-\eta)(1-\theta)}\Big[(c+\tau)^{1-\theta}-c^{1-\theta}\Big], \qquad (6.23)
\end{aligned}$$

where the second equality follows from the explicit form of the function $b(t)$ (see (6.7)) and the last equality is obtained because the limit of the function $(c+t)^{1-\theta}-(c+t+\tau)^{1-\theta}$ is, for t tending to infinity, always zero. In fact, if $\theta > 1$ it is obvious; if instead $1 > \theta > 0$, applying De L'Hôpital's rule, we have

$$\lim_{t\to\infty}\Big[(c+t)^{1-\theta}-(c+t+\tau)^{1-\theta}\Big] = \lim_{t\to\infty}\frac{1-\left(1+\frac{\tau}{c+t}\right)^{1-\theta}}{(c+t)^{\theta-1}} = 0.$$

At this point, we can substitute equation (6.23) in (6.17):

$$L(0;\tau) \approx \frac{\eta}{(1-\eta)(1-\theta)}\Big[c^\theta(c+\tau)^{1-\theta}-c\Big]+\tau.$$

Then, introducing the function $A(\tau)$ as in (6.14), we get exactly the thesis of this theorem, i.e., for small τ,

$$\mathbf{P}\{\tau\} = e^{-\varpi L(0;\tau)} \approx \exp\left\{-\varpi\tau - \frac{\eta\varpi}{1-\eta}A(\tau)\right\}.$$

$\square$

Remark 13. The probability of having zero observable events in $[0,\tau]$, given in (6.13), coincides with formula (32), page 6, in [Saichev and Sornette, 2007]. In fact, they derive the above formula for the case of a simplified model which takes into account the impact of the Omori-Utsu law on recurrence times, i.e., considering a model in which the ratio of the number of observable spontaneous events over the total number of observable events is approximately independent of the threshold m_c. This correspond to the situation in which every triggering event is also observable, that is $m_c = m_0$, as exactly holds in the case we are considering in the current section.

Now, as already specified, using the Palm theory (see equation (1.13) in Subsection 1.5.1, Chapter 1) we obtain the density $F_{inter}(\tau)$ relative to the interevent time in the following way:

$$F_{inter}(\tau) = \frac{1}{\bar{\lambda}} \frac{d^2 \mathbf{P}\{\tau\}}{d\tau^2}. \tag{6.24}$$

Heuristically, we can suppose that, if

$$\hat{L}(0;\tau) = \frac{\eta}{(1-\eta)(1-\theta)} \Big[c^{\theta}(c+\tau)^{1-\theta} - c \Big] + \tau,$$

that is the approximation for $L(0;\tau)$ obtained in the previous Theorem, it holds also

$$\frac{d^k L(0;\tau)}{d\tau^k} \approx \frac{d^k \hat{L}(0;\tau)}{d\tau^k},$$

for $k = 1, 2$. Then, we can compute then the second derivative of the abovementioned probability $\mathbf{P}\{\tau\}$ as

$$\frac{d\mathbf{P}\{\tau\}}{d\tau} \approx \mathbf{P}\{\tau\} \Big[-\varpi - \frac{\eta\varpi}{1-\eta} a(\tau) \Big],$$

in fact

$$\frac{d}{d\tau} A(\tau) = \frac{d}{d\tau} \int_0^{\tau} a(x)dx = a(\tau). \tag{6.25}$$

Hence, since

$$\frac{d}{d\tau} a(\tau) = -\frac{d}{d\tau} b(\tau) = -\Phi(\tau),$$

it holds

$$\begin{aligned} \frac{d^2 \mathbf{P}\{\tau\}}{d\tau^2} \approx & \mathbf{P}\{\tau\} \left[\left(\varpi + \frac{\eta\varpi}{1-\eta} a(\tau) \right)^2 + \frac{\eta\varpi}{1-\eta} \Phi(\tau) \right] \\ = & \exp\left\{ -\varpi\tau - \frac{\eta\varpi}{1-\eta} A(\tau) \right\} \left[\left(\varpi + \frac{\eta\varpi}{1-\eta} a(\tau) \right)^2 + \frac{\eta\varpi}{1-\eta} \Phi(\tau) \right]. \end{aligned}$$

In conclusion,

$$F_{inter}(\tau) \approx \frac{1}{\lambda} \exp\left\{-\varpi\tau - \frac{\eta\varpi}{1-\eta}A(\tau)\right\}\left[\left(\varpi + \frac{\eta\varpi}{1-\eta}a(\tau)\right)^2 + \frac{\eta\varpi}{1-\eta}\Phi(\tau)\right]. \tag{6.26}$$

We can see that this density doesn't depend on the model we proposed for the transition probability of triggered events' magnitude. In fact, as we have seen before, this transition probability "vanishes" when applying condition

$$I_A(m'') = \eta p(m'').$$

Therefore, in the case $m_c = m_0$, in which all the events are observable, our hypothesis of dependence doesn't play any role. This could be justified by the fact that, if $m_c = m_0$, all the observable events may also generate aftershocks and there aren't external or background sources that could influence the process with their progeny: there is no distinction between the events that may or may not produce their own offsprings. The proportion of triggering/triggered events, on which depends the hypothesis of magnitude correlation, is exactly the one obtained by considering only the events in the process. Instead, as we are going to see in the next section, if $m_c > m_0$, the number of observable events in the process changes and consequently it changes also the proportion of triggering/triggered shocks.

6.3 The general case $m_c \geq m_0$

In order to find the interevent time density for the general case $m_c \geq m_0$, let's define at first the function

$$\Psi(h(\cdot), \widetilde{m}) := \int_{\widetilde{m}}^{\infty} p(m')e^{-\varrho(m')h(m')}dm', \tag{6.27}$$

where $\widetilde{m} \geq 0$ and $h(\cdot)$ is a continuous function depending on some variables; it is here considered as a function of only one of them, corresponding to the triggering event's magnitude m'.

Let's focus now on the function $L(z;\tau)$, given in equation (6.5), at the exponent of the PGF $\Omega(z;\tau)$, relative to the total number of observable events in $[0,\tau]$ (see also (6.4)). Recalling the expression (6.6) for $G(z;\tau|t,m')$, that is the PGF of the number of observable events in $[0,\tau]$ triggered by a fixed spontaneous shock with magnitude m' occurred in $-t$, where $t \geq 0$, we can rewrite the integral $\int_{m_0}^{\infty} G(z;\tau|t,m')p(m')dm'$ in the right member of (6.5), as

$$\int_{m_0}^{\infty} G(z;\tau|t,m')p(m')dm' = \Psi[y(z;t,\tau,m_0,m'),m_0], \tag{6.28}$$

where

$$\begin{aligned}
y(z;t,\tau,m_0,m') :=& b(t+\tau) - \int_0^t dx\,\Phi(x) \int_{m_0}^{\infty} G(z;\tau|t-x,m'')p(m''|m')dm'' \\
& - \int_0^{\tau} dx\,\Phi(t+\tau-x) \int_{m_0}^{m_c} G(z;x|0,m'')p(m''|m')dm'' \\
& - z\int_0^{\tau} dx\,\Phi(t+\tau-x) \int_{m_c}^{\infty} G(z;x|0,m'')p(m''|m')dm''
\end{aligned} \tag{6.29}$$

$$\begin{aligned}
=& b(t+\tau) - (\Phi(\cdot)\otimes D(z;\cdot,\tau,m_0,m'))(t) \\
& - \Big(\Phi(t+\cdot)\otimes \big[D_+(z;\cdot,m_0,m') - D_+(z;\cdot,m_c,m')\big]\Big)(\tau) \\
& - z(\Phi(t+\cdot)\otimes D_+(z;\cdot,m_c,m'))(\tau).
\end{aligned} \tag{6.30}$$

In the latter equality we have set

$$D(z;t,\tau,m_0,\overline{m}) := \int_{m_0}^{\infty} G(z;\tau|t,m')p(m'|\overline{m})dm' \tag{6.31}$$

and

$$D_+(z;t,\widetilde{m},\overline{m}) := \int_{\widetilde{m}}^{\infty} G(z;t|0,m')p(m'|\overline{m})dm'. \tag{6.32}$$

Let's notice that in our case we have $\widetilde{m} = m_0, m_c$.

Similarly, the other two integrals with respect to the magnitude in the right member of (6.5), i.e.,

$$\int_{m_0}^{m_c} G(z;t|0,m')p(m')dm' \quad \text{and} \quad \int_{m_c}^{\infty} G(z;t|0,m')p(m')dm',$$

can be expressed in terms of

$$\begin{aligned}
\int_{m_0}^{\infty} G(z;t|0,m')p(m')dm' &= \Psi[s(z;t,m_0,m'),m_0], \\
\int_{m_c}^{\infty} G(z;t|0,m')p(m')dm' &= \Psi[s(z;t,m_0,m'),m_c],
\end{aligned} \tag{6.33}$$

where

$$\begin{aligned}
s(z;t,m_0,m') :=& b(t) - \int_0^t dx\,\Phi(t-x)\int_{m_0}^{m_c} G(z;x|0,m'')p(m''|m')dm'' \\
&- z\int_0^t dx\,\Phi(t-x)\int_{m_c}^{\infty} G(z;x|0,m'')p(m''|m')dm'' \qquad (6.34)\\
=& b(t) - \Big(\Phi(\cdot)\otimes\big[D_+(z;\cdot,m_0,m') - D_+(z;\cdot,m_c,m')\big]\Big)(t) \\
&- z(\Phi(\cdot)\otimes D_+(z;\cdot,m_c,m'))(t). \qquad (6.35)
\end{aligned}$$

It follows that the formula (6.5) for the function $L(z;\tau)$ becomes

$$\begin{aligned}
L(z;\tau) =& \int_0^\infty \Big\{1 - \Psi[y(z;t,\tau,m_0,m'),m_0]\Big\}dt \\
&+ \int_0^\tau \Big\{1 - \Psi[s(z;t,m_0,m'),m_0] + (1-z)\Psi[s(z;t,m_0,m'),m_c]\Big\}dt.
\end{aligned} \tag{6.36}$$

As explained before, one can obtain the probability $\mathbf{P}\{\tau\}$ of having zero observable events in a temporal interval of length τ by computing the probability generating function $\Omega(z;\tau)$ in zero:

$$\mathbf{P}\{\tau\} = \Omega(0;\tau) = e^{-\varpi L(0;\tau)}. \tag{6.37}$$

By setting $z=0$ in (6.36), we easily obtain

$$\begin{aligned}
L(0;\tau) =& \int_0^\infty \Big\{1 - \Psi[y(0;t,\tau,m_0,m'),m_0]\Big\}dt \\
&+ \int_0^\tau \Big\{1 - \Psi[s(0;t,m_0,m'),m_0] + \Psi[s(0;t,m_0,m'),m_c]\Big\}dt \\
=& \int_0^\infty \overline{N}(t,\tau)dt + \int_0^\tau \overline{N}_-(t)dt,
\end{aligned} \tag{6.38}$$

where we have set

$$\overline{N}(t,\tau) := 1 - \Psi[y(0;t,\tau,m_0,m'),m_0] \tag{6.39}$$

and

$$\overline{N}_-(t) := 1 - \Psi[s(0;t,m_0,m'),m_0] + \Psi[s(0;t,m_0,m'),m_c]. \tag{6.40}$$

Remark 14. Recalling equations (6.28) and (6.33), it is clear that we can rewrite the last two functions just defined in the following way:

$$\overline{N}(t,\tau) = 1 - \int_{m_0}^{\infty} G(0;\tau|t,m')p(m')dm' \tag{6.41}$$

and

$$\overline{N}_{-}(t) = 1 - \int_{m_0}^{m_c} G(0;t|0,m')p(m')dm'. \tag{6.42}$$

Obviously, in the case $m_c = m_0$ we have that $\overline{N}(t,\tau) = \overline{N}_{m_c=m_0}(t,\tau)$ and $\overline{N}_{-}(t) = 1$.

At this point, in order to obtain the probability $\mathbf{P}\{\tau\}$ of having zero observable events in $[0,\tau]$, we have to find an expression for the functions $\overline{N}(t,\tau)$ and $\overline{N}_{-}(t)$ and then to substitute the obtained results in equation (6.38). Actually, since they are integrals rather difficult to compute, we can approximate them by means of the first order Taylor power series expansion of the function $\Psi(h(\cdot),\widetilde{m})$, with respect to the first argument, around the point zero. In our case it holds $\widetilde{m} = m_0, m_c$ and the function $h(\cdot)$ in $\overline{N}(t,\tau)$ and $\overline{N}_{-}(t)$ (equations (6.39) and (6.40)) is respectively equal to $y(0;t,\tau,m_0,m')$ and $s(0;t,m_0,m')$ where, recalling the expressions of $y(z;t,\tau,m_0,m')$ and $s(z;t,m_0,m')$ given respectively in (6.30) and (6.35),

$$\begin{aligned} y(0;t,\tau,m_0,m') =& b(t+\tau) - (\Phi(\cdot)\otimes D(0;\cdot,\tau,m_0,m'))(t) \\ & - \Big(\Phi(t+\cdot)\otimes\big[D_{+}(0;\cdot,m_0,m') - D_{+}(0;\cdot,m_c,m')\big]\Big)(\tau) \end{aligned} \tag{6.43}$$

and

$$s(0;t,m_0,m') = b(t) - \Big(\Phi(\cdot)\otimes\big[D_{+}(0;\cdot,m_0,m') - D_{+}(0;\cdot,m_c,m')\big]\Big)(t). \tag{6.44}$$

Let's notice that the above-mentioned Taylor approximation makes sense since, in both $\overline{N}(t,\tau)$ and $\overline{N}_{-}(t)$, the first argument of Ψ is small when $\tau \approx 0$. In fact, we have that

- for the argument $y(0;t,\tau,m_0,m')$, when $\tau = 0$, it holds

  $$y(0;t,0,m_0,m') = b(t) - b(t) = 0.$$

  This follows by setting $z = 0$ in (6.29) and observing that, since $G(z;\tau|t,m'')$ is the probability generating function of the number of events triggered in $[0,\tau]$ by a shock with magnitude m', occurred t seconds before time 0, it holds

$$G(0;0|t,m'') = 1.$$

The latter is due to the regularity of the process and the fact that the probability $\mathbb{P}\{N(0)=0\}$ of having zero observable events in an interval of length zero is equal to one;

- on the other hand, for the argument $s(0;t,m_0,m')$, since $t \in [0,\tau]$ and $\tau = 0$, it holds

$$s(0;0,m_0,m') = 0,$$

as one can easily deduce by setting $z=0$ in (6.34).

Let's compute then the above-cited expansion and the relative approximations of $\overline{N}(t,\tau)$ and $\overline{N}_-(t)$. Let's start with the first function.

Lemma 5. *For small τ, the function $\overline{N}(t,\tau)$ in (6.38) can be approximated as*

$$\begin{aligned}\overline{N}(t,\tau) \approx & \eta b(t+\tau) - \int_0^t dx\, \Phi(t-x) \int_{m_0}^{\infty} G(0;\tau|x,m'') I_A(m'') dm'' \\ & - \int_0^{\tau} dx\, \Phi(t+\tau-x) \int_{m_0}^{m_c} G(0;x|0,m'') I_A(m'') dm'', \qquad (6.45)\end{aligned}$$

where $\Phi(\cdot)$ is the Omori-Utsu law (equation (6.3)), η is the branching ratio (equation (2.18)), $b(\cdot)$ is the cumulative function relative to the Omori-Utsu law (equation (6.7)) and

$$I_A(m'') := \int_{m_0}^{\infty} p(m')\varrho(m')p(m''|m')dm'. \qquad (6.46)$$

Proof. Recalling (6.27), (6.39) and (6.43), it holds

$$\begin{aligned}&\overline{N}(t,\tau) = 1 - \Psi[y(0;t,\tau,m_0,m'),m_0] \\ &= 1 - \int_{m_0}^{\infty} p(m') e^{-\varrho(m')y(0;t,\tau,m_0,m')} dm' \\ &\approx 1 - \left[1 - \int_{m_0}^{\infty} p(m')\varrho(m')y(0;t,\tau,m_0,m')dm'\right]\end{aligned}$$

$$
\begin{aligned}
&= \int_{m_0}^{\infty} p(m')\varrho(m')dm'b(t+\tau) \\
&\quad - \int_{m_0}^{\infty} p(m')\varrho(m')(\Phi(\cdot) \otimes D(0;\cdot,\tau,m_0,m'))(t)dm' \\
&\quad - \int_{m_0}^{\infty} p(m')\varrho(m')\Big(\Phi(t+\cdot) \otimes \big[D_+(0;\cdot,m_0,m') - D_+(0;\cdot,m_c,m')\big]\Big)(\tau)dm'.
\end{aligned} \tag{6.47}
$$

Let's study now separately the integrals in the last equality of the above expression (6.47). In the matter of the first, recalling the first equality in (2.18), by definition we have

$$
\int_{m_0}^{\infty} p(m')\varrho(m')dm'b(t+\tau) = \eta b(t+\tau).
$$

For the second one, from (6.46) and (6.31), we have

$$
\begin{aligned}
&\int_{m_0}^{\infty} p(m')\varrho(m')(\Phi(\cdot) \otimes D(0;\cdot,\tau,m_0,m'))(t)dm' \\
&= \int_{m_0}^{\infty} dm'p(m')\varrho(m') \int_0^t dx\, \Phi(t-x) \int_{m_0}^{\infty} G(0;\tau|x,m'')p(m''|m')dm'' \\
&= \int_0^t dx\, \Phi(t-x) \int_{m_0}^{\infty} \int_{m_0}^{\infty} p(m')\varrho(m')G(0;\tau|x,m'')p(m''|m')dm''dm' \\
&= \int_0^t dx\, \Phi(t-x) \int_{m_0}^{\infty} dm''G(0;\tau|x,m'') \int_{m_0}^{\infty} p(m')\varrho(m')p(m''|m')dm' \\
&= \int_0^t dx\, \Phi(t-x) \int_{m_0}^{\infty} G(0;\tau|x,m'')I_A(m'')dm''.
\end{aligned}
$$

Furthermore, if we use equation (6.32), we can compute the third integral in the last expression of (6.47) as follows:

$$
\begin{aligned}
&\int_{m_0}^{\infty} p(m')\varrho(m')\Big(\Phi(t+\cdot) \otimes \big[D_+(0;\cdot,m_0,m') - D_+(0;\cdot,m_c,m')\big]\Big)(\tau)dm' \\
&= \int_{m_0}^{\infty} dm'p(m')\varrho(m') \int_0^{\tau} dx\, \Phi(t+\tau-x) \int_{m_0}^{m_c} G(0;x|0,m'')p(m''|m')dm'' \\
&= \int_0^{\tau} dx\, \Phi(t+\tau-x) \int_{m_0}^{\infty} \int_{m_0}^{m_c} p(m')\varrho(m')G(0;x|0,m'')p(m''|m')dm''dm'
\end{aligned}
$$

$$= \int_0^\tau dx\, \Phi(t+\tau-x) \int_{m_0}^{m_c} dm'' G(0;x|0,m'') \int_{m_0}^{\infty} p(m')\varrho(m')p(m''|m')dm'$$
$$= \int_0^\tau dx\, \Phi(t+\tau-x) \int_{m_0}^{m_c} G(0;x|0,m'')I_A(m'')dm''.$$

Then, the thesis of the theorem can be easily derived by substituting in equation (6.47) the results obtained. □

Before proceeding with the study of the approximation of the second function $\overline{N}_-(t)$ in (6.38), we focus for a moment on the approximation obtained for $\overline{N}(t,\tau)$.

Corollary 1. *For small τ, it holds*

$$\int_0^\infty \overline{N}(t,\tau)dt \approx \frac{\eta}{1-\eta}(a(\cdot) \otimes \overline{N}_-(\cdot))(\tau), \tag{6.48}$$

where $a(\cdot)$ is the survivor function relative to the Omori-Utsu law (equation (6.15)*) and η is the branching ratio (equation* (2.18)*).*

Proof. Let's impose in (6.45) the crucial hypothesis $I_A(m'') = \eta p(m'')$, as previously assumed before in the case $m_c = m_0$ (see the proof of Theorem 12). We have already explained that this condition guarantees the validity of the Gutenberg-Richter law when averaging on all triggering events' magnitudes and corresponds to the fact that $p(m'')$ is the eigenfunction corresponding to the eigenvalue η of the operator $\mathcal{K}_2$ defined in (4.6). We get

$$\begin{aligned}
\overline{N}(t,\tau) \approx\, & \eta b(t+\tau) - \eta \int_0^t dx\, \Phi(x) \int_{m_0}^{\infty} G(0;\tau|t-x,m'')p(m'')dm'' \\
& - \eta \int_0^\tau dx\, \Phi(t+\tau-x) \int_{m_0}^{m_c} G(0;x|0,m'')p(m'')dm'' \\
=\, & \eta[b(t+\tau) - b(t)] \\
& + \eta\Big\{\int_0^t dx\, \Phi(x)\Big[1 - \int_{m_0}^{\infty} G(0;\tau|t-x,m'')p(m'')dm''\Big]\Big\} \\
& + \eta\Big\{\int_0^\tau dx\, \Phi(t+\tau-x)\Big[-\int_{m_0}^{m_c} G(0;x|0,m'')p(m'')dm''\Big]\Big\} \\
=\, & \eta \int_t^{t+\tau} \Phi(y)dy + \eta \int_0^t \Phi(x)\overline{N}(t-x,\tau)dx
\end{aligned}$$

$$
\begin{aligned}
&+\eta\Big\{\int_t^{t+\tau} dy\,\Phi(y)\Big[-\int_{m_0}^{m_c} G(0;t+\tau-y|0,m'')p(m'')dm''\Big]\Big\}\\
=&\eta(\Phi(\cdot)\otimes\overline{N}(\cdot,\tau))(t)+\eta\int_t^{t+\tau} dy\,\Phi(y)\Big[1\\
&-\int_{m_0}^{m_c} G(0;t+\tau-y|0,m'')p(m'')dm''\Big]\\
=&\eta(\Phi(\cdot)\otimes\overline{N}(\cdot,\tau))(t)+\eta\int_t^{t+\tau}\Phi(y)\overline{N}_-(t+\tau-y)dy\\
=&\eta(\Phi(\cdot)\otimes\overline{N}(\cdot,\tau))(t)+\eta\int_0^{\tau}\Phi(t+\tau-x)\overline{N}_-(x)dx,
\end{aligned}
$$

where we have used equations (6.7), (6.41) and (6.42) and in addition, in third and last equality, we have made the change of variable $y = t+\tau-x$. It then follows that

$$\overline{N}(t,\tau) \approx \eta(\Phi(\cdot)\otimes\overline{N}(\cdot,\tau))(t)+\eta(\Phi(t+\cdot)\otimes\overline{N}_-(\cdot))(\tau).$$

Integrating both the members with respect to time, between zero and infinity, we get

$$
\begin{aligned}
\int_0^\infty \overline{N}(t,\tau)dt \approx&\eta\int_0^\infty\int_0^t \Phi(t-x)\overline{N}(x,\tau)dxdt\\
&+\eta\int_0^\infty\int_0^\tau \Phi(t+\tau-x)\overline{N}_-(x)dxdt\\
=&\eta\int_0^\infty\int_x^\infty \Phi(t-x)\overline{N}(x,\tau)dtdx\\
&+\eta\int_0^\tau\int_0^\infty \Phi(t+\tau-x)\overline{N}_-(x)dtdx\\
=&\eta\int_0^\infty dx\overline{N}(x,\tau)\int_x^\infty \Phi(t-x)dt\\
&+\eta\int_0^\tau dx\overline{N}_-(x)\int_0^\infty \Phi(t+\tau-x)dt\\
=&\eta\int_0^\infty \overline{N}(x,\tau)dx+\eta\int_0^\tau dx\overline{N}_-(x)\int_{\tau-x}^\infty \Phi(y)dy\\
=&\eta\int_0^\infty \overline{N}(x,\tau)dx+\eta\int_0^\tau \overline{N}_-(x)a(\tau-x)dx\\
=&\eta\int_0^\infty \overline{N}(x,\tau)dx+\eta(a(\cdot)\otimes\overline{N}_-(\cdot))(\tau),
\end{aligned}
$$

where we have used that

$$\int_x^\infty \Phi(t-x)dt = 1$$

and (6.15). We have then obtained equation (6.48). □

As a consequence of Corollary 1, the function $L(0;\cdot)$ in (6.38) for small τ can be written only in terms of the function $\overline{N}_-(\cdot)$. We have indeed

$$\begin{aligned} L(0;\tau) &= \int_0^\infty \overline{N}(t,\tau)dt + \int_0^\tau \overline{N}_-(t)dt \\ &\approx \frac{\eta}{1-\eta}(a(\cdot)\otimes \overline{N}_-(\cdot))(\tau) + \int_0^\tau \overline{N}_-(t)dt. \end{aligned} \tag{6.49}$$

We are ready now to proceed with the study of the approximation of $\overline{N}_-(t)$.

Lemma 6. *For $t \in [0,\tau]$ and τ small, the function $\overline{N}_-(t)$ in* (6.38) *can be approximated as*

$$\begin{aligned} \overline{N}_-(t) \approx & Q + \eta b(t)\left[1 - e^{-(\beta-a)(m_c-m_0)}\right] \\ & - \int_0^t dx\, \Phi(t-x) \int_{m_0}^{m_c} G(0;x|0,m'')[I_A(m'') - I_B(m'')]dm'', \end{aligned} \tag{6.50}$$

where

$$Q := e^{-\beta(m_c-m_0)}, \tag{6.51}$$

that is $\mathbb{P}\{M > m_c\}$ if M is distributed according to the Gutenberg-Richter law (equation (2.2)*); $\Phi(\cdot)$ is the Omori-Utsu law (equation* (6.3)*); η is the branching ratio (equation* (2.18)*); $b(\cdot)$ is the cumulative function relative to the Omori-Utsu law (equation* (6.7)*); $I_A(m'')$ is the integral defined in* (6.46) *and*

$$I_B(m'') := \int_{m_c}^\infty p(m')\varrho(m')p(m''|m')dm'. \tag{6.52}$$

Remark 15. Recalling (6.2), we can write the average rate of the total observable events as

$$\bar{\lambda} = \frac{\varpi Q}{1-\eta}.$$

Proof. Using (6.27), (6.40) and (6.44), we get

$$\begin{aligned}
&\overline{N}_-(t) = 1 - \Psi[s(0;t,m_0,m'),m_0] + \Psi[s(0;t,m_0,m'),m_c] \\
&= 1 - \int_{m_0}^{m_c} p(m')e^{-\varrho(m')s(0;t,m_0,m')}dm' \\
&\approx 1 - \left[1 - \int_{m_c}^{\infty} p(m')dm' - \int_{m_0}^{m_c} p(m')\varrho(m')s(0;t,m_0,m')dm'\right] \\
&= Q + \int_{m_0}^{m_c} p(m')\varrho(m')s(0;t,m_0,m')dm' \\
&= Q + \int_{m_0}^{m_c} p(m')\varrho(m')dm' b(t) \\
&\quad - \int_{m_0}^{m_c} p(m')\varrho(m')\Big(\Phi(\cdot) \otimes \big[D_+(0;\cdot,m_0,m') - D_+(0;\cdot,m_c,m')\big]\Big)(t)dm',
\end{aligned} \tag{6.53}$$

where

$$Q = \int_{m_c}^{\infty} p(m')dm' = e^{-\beta(m_c - m_0)}.$$

As previously done, we study separately the integrals in the final expression of (6.53). Recalling again equations (2.1), (2.2), (2.18), (6.32) and (6.46), we obtain the following results:

$$\begin{aligned}
b(t)\int_{m_0}^{m_c} p(m')\varrho(m')dm' &= b(t)\left[\eta - \beta\kappa\int_{m_c}^{\infty} e^{-(\beta-a)(m'-m_0)}dm'\right] \\
&= b(t)\left[\eta - \frac{\beta\kappa e^{-(\beta-a)(m_c-m_0)}}{\beta - a}\right] \\
&= \eta b(t)\left[1 - e^{-(\beta-a)(m_c-m_0)}\right];
\end{aligned}$$

$$\begin{aligned}
&\int_{m_0}^{m_c} p(m')\varrho(m')\Big(\Phi(\cdot) \otimes \big[D_+(0;\cdot,m_0,m') - D_+(0;\cdot,m_c,m')\big]\Big)(t)dm' \\
&= \int_{m_0}^{m_c} dm' p(m')\varrho(m') \int_0^t dx\, \Phi(t-x) \int_{m_0}^{m_c} G(0;x|0,m'')p(m''|m')dm'' \\
&= \int_0^t dx\, \Phi(t-x) \int_{m_0}^{m_c} dm'' G(0;x|0,m'') \int_{m_0}^{m_c} p(m')\varrho(m')p(m''|m')dm' \\
&= \int_0^t dx\, \Phi(t-x) \int_{m_0}^{m_c} G(0;x|0,m'')[I_A(m'') - I_B(m'')]dm'',
\end{aligned}$$

where we have set

$$I_B(m'') := \int_{m_c}^{\infty} p(m')\varrho(m')p(m''|m')dm'.$$

Let's notice that this latter integral is exactly $\mathcal{K}_2 p(m)\mathbb{1}_{[m_c,\infty)}(m)$, where $\mathcal{K}_2$ is the operator defined in the proof of Lemma 1.

Substituting the obtained results in (6.53), we are then able to find the thesis of this theorem, that is the approximated expression (6.50) of the function $\overline{N}_-(t)$. $\square$

At this point we can work on the approximation of the function $\overline{N}_-(t)$ as we made for $\overline{N}(t,\tau)$ (see Corollary 1), again by imposing some conditions. In particular, the first condition to impose is once more the crucial one for our analysis $I_A(m'') = \eta p(m'')$, already used in the case $m_c = m_0$ (see the proof of Theorem 12) and in Corollary 1. The second condition regards instead the integral $I_B(m'')$, for which one can notice that its only difference with $I_A(m'')$ is in the lower extreme of integration. In order to find the suitable condition that the integral $I_B(m'')$ must satisfy, we have to use an explicit form of the probability $p(m''|m')$. For this purpose, we have used the function (4.37) found in Chapter 4, i.e.,

$$\begin{aligned} p(m''|m') &= p(m'')\big[1 + f(m', m'')\big] \\ &= \beta e^{-\beta(m''-m_0)}\left[1 + C_1\big(1 - 2e^{-(\beta-a)(m'-m_0)}\big)\big(1 - 2e^{-\beta(m''-m_0)}\big)\right], \end{aligned}$$

for $0 \leq C_1 \leq 1$. By substituting this explicit form in (6.52), recalling that $I_B(m'') = \mathcal{K}_2 p(m'')\mathbb{1}_{[m_c,\infty)}(m'')$, where $\mathcal{K}_2$ is the operator defined in the proof of Lemma 1 and repeating the same computations as in the latter Lemma (see (B.15)), we can deduce that $I_B(m'')$ is a linear combination of $p(m)$ and $p^2(m)$. More precisely, we get

$$\begin{aligned} I_B(m'') =& p(m'')\eta H + p(m'')C_1\left[1 - 2\frac{p(m'')}{\beta}\right]\eta(H - H^2) \\ =& p(m'')\eta L_1 + p^2(m'')\eta L_2, \end{aligned} \tag{6.54}$$

where

$$H := e^{-(\beta-a)(m_c-m_0)} \tag{6.55}$$

and

$$L_1 := H[1 + C_1 - C_1 H], \tag{6.56}$$

$$L_2 := \frac{2C_1 H}{\beta}(H - 1). \tag{6.57}$$

The expression (6.54) obtained for the integral $I_B(m'')$ depends then on both the Gutenberg-Richter function $p(m'')$ and its square. This complicates a lot our study. However, we can observe that

$$\left|1 - 2e^{-\beta(m''-m_0)}\right| = \left|1 - 2\frac{p(m'')}{\beta}\right| < 1,$$

$C_1 \leq 1$ and, for the typical values of the parameters, $H \approx 1$ (then $(\beta - a)(m_c - m_0)$ is close to zero) and the difference $H - H^2$ is generally small. It follows that we can approximate $I_B(m'')$ with a function of kind $p(m'')\eta L$, where the constant L is very close to H. More precisely, the value L can be estimated by minimizing the difference between $I_B(m'')$ and $p(m'')\eta L$. There are various methods which can be used to this aim. Setting Setting

$$z(m'') := p(m'')L_1 + p^2(m'')L_2,$$

we can use for example the 2-norm and then compute

$$\underset{L}{\operatorname{argmin}} \int_{m_0}^{m_c} [z(m'') - p(m'')L]^2 dm''. \tag{6.58}$$

Since

$$\int_{m_0}^{m_c} p^k(m'')dm'' = \frac{\beta^{k-1}}{k}[1 - Q^k],$$

as one can easily obtain recalling (2.2), we have that the above equation (6.58) becomes

$$\begin{aligned} &\underset{L}{\operatorname{argmin}} \int_{m_0}^{m_c} [z(m'') - p(m'')L]^2 dm'' \\ &= \underset{L}{\operatorname{argmin}} \int_{m_0}^{m_c} \left[(L_1 - L)p(m'') + L_2 p^2(m'')\right]^2 dm'' \end{aligned}$$

Recalling the definitions of L_1, L_2 (see (6.56), (6.57)), we get that the argument of the minimum is obtained in

$$L_{min} = L_1 + L_2\beta\frac{2(1-Q^3)}{3(1-Q^2)} = H + C_1 H(1-H)\left[1 - \frac{4(1-Q^3)}{3(1-Q^2)}\right]$$

and values

$$L_2^2\beta^3\left[-\frac{2}{9}\frac{(1-Q^3)^2}{1-Q^2} + \frac{1-Q^4}{4}\right] = \beta\left[\frac{C_1 H(1-H)}{3}\right]^2 \frac{(1-Q)^3(Q^2+4Q+1)}{1+Q}.$$

Another approach could be a numerical one, by means for example of the MATLAB function *"nlinfit"*, which estimates the best coefficient L minimizing the difference $z(m'') - p(m'')L$ by least squares. This function allows to obtain the optimal curve as close as possible to the data.

As an example, let's consider the following parameters dataset:

$$\beta = 2.0493, \quad a = 1.832, \quad C_1 = 0.8, \quad m_0 = 1, \quad m_c = 1.8.$$

If we use the 2-norm, we get that the minimum is obtained in $L_{min} = 0.8001$ and values $2.08 * 10^{-3}$. We also observe that L_{min} is close to the constant $H = 0.8404$, as expected. If instead we proceed by means of a numerical algorithm consisting of the above estimation through the MATLAB function *"nlinfit"*, we get the constant $L_{min} = 0.8032$ that is close to the value obtained with the 2-norm and then to $H = 0.8404$. A graphical check for the numerical case is given in Fig. 6.1.

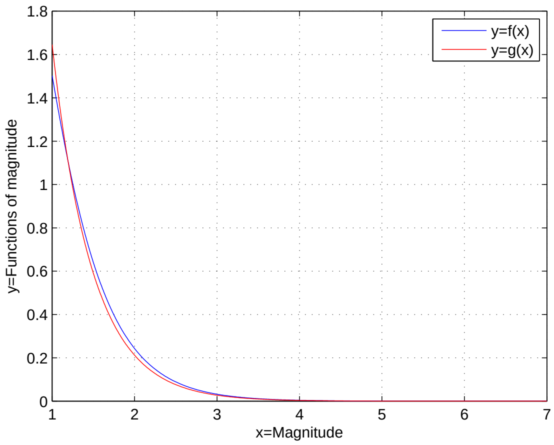


Figure 6.1: Plot of the functions $f(x) = Hp(x) + C_1 p(x)(1 - 2\frac{p(x)}{\beta})H(1 - H)$ and $g(x) = Lp(x)$, where $p(x) = \beta \exp\{-\beta(x - m_0)\}$ and $(\beta, a, C_1, m_0, m_c) = (2.0493, 1.832, 0.8, 1, 1.8)$.

We can then proceed with the substitution in equation (6.50) of the two conditions

$$I_A(m'') = \eta p(m'')$$

and

$$I_B(m'') \approx \eta L p(m''), \tag{6.59}$$

for a certain constant L to be found numerically. For small τ, we have

$$\begin{aligned}
\overline{N}_-(t) \approx & Q + \eta b(t)[1-H] \\
& - \eta[1-L] \int_0^t dx\, \Phi(t-x) \int_{m_0}^{m_c} G(0; x|0, m'')p(m'')dm'' \\
= & Q + \eta b(t)[1-H] - \eta[1-L]b(t) + \eta[1-L]b(t) \\
& - \eta[1-L] \int_0^t dx\, \Phi(x) \int_{m_0}^{m_c} G(0; t-x|0, m'')p(m'')dm'' \\
= & Q - \eta b(t)[H-L] \\
& + \eta[1-L] \int_0^t dx\, \Phi(x) \Big[1 - \int_{m_0}^{m_c} G(0; t-x|0, m'')p(m'')dm'' \\
= & Q - \eta b(t)[H-L] + \eta\big[1-L\big] \int_0^t \Phi(x)\overline{N}_-(t-x)dx,
\end{aligned} \tag{6.60}$$

where we have used equation (6.42). The error that we commit is

$$\begin{aligned}
& \min \left| \int_0^t dx \Phi(t-x) \int_{m_0}^{m_c} G(0; x|0, m)[I_B(m) - \eta L p(m)]dm \right| \\
& \leq \min \int_0^t dx \Phi(t-x) \int_{m_0}^{m_c} 1 \cdot \eta \left| z(m) - L p(m) \right| dm \\
& \leq \min \eta b(t) \left(\int_{m_0}^{m_c} 1^2 dm \right)^{\frac{1}{2}} \left(\int_{m_0}^{m_c} \left| z(m) - L p(m) \right|^2 dm \right)^{\frac{1}{2}}.
\end{aligned} \tag{6.61}$$

When $L = L_{min}$ and since

$$\begin{aligned}
1 - Q = & 1 - e^{-\beta(m_c - m_0)} \leq \beta(m_c - m_0) \\
1 - H = & 1 - e^{-(\beta - a)(m_c - m_0)} \leq (\beta - a)(m_c - m_0),
\end{aligned}$$

the last member in equation (6.61) becomes

$$\begin{aligned}
& \eta b(t)\sqrt{m_c - m_0} \frac{C_1 H(1-H)}{3} \sqrt{\beta \frac{\big(1-Q\big)^3 \big(Q^2 + 4Q + 1\big)}{1+Q}} \\
& \leq \eta b(t) \frac{C_1 H}{3} \sqrt{\frac{Q^2 + 4Q + 1}{1+Q}} \beta^2 (\beta - a)(m_c - m_0)^3 \\
& \leq \eta b(t) C_1 H \beta^2 (\beta - a)(m_c - m_0)^3
\end{aligned}$$

and then, if $(\beta - a)(m_c - m_0)$ is small, so is the error.

From (6.60) we then obtain

$$\overline{N}_-(t) \approx Q - \eta b(t)[H-L] + \delta(\Phi(\cdot) \otimes \overline{N}_-(\cdot))(t), \tag{6.62}$$

where

$$\delta = \eta\big[1-L\big]. \tag{6.63}$$

Before proceeding with the research of the function $\overline{N}_-(t)$, let's state the following Lemma.

Lemma 7. *When $H \approx 1$, it holds*

$$\int_0^\tau \overline{N}_-(t)dt \approx \frac{1}{1-\delta}\Big\{\tau\Big[Q-\eta[H-L]\Big] + \eta[H-L]A(\tau) - \delta(a(\cdot)\otimes\overline{N}_-(\cdot))(\tau)\Big\} \tag{6.64}$$

and

$$(a(\cdot)\otimes\overline{N}_-(\cdot))(\tau) \approx \frac{\tau}{\delta}\Big[Q-\eta[H-L]\Big] + \frac{\eta[H-L]}{\delta}A(\tau) + \Big[1-\frac{1}{\delta}\Big]\int_0^\tau \overline{N}_-(t)dt, \tag{6.65}$$

where $a(\cdot)$, $A(\cdot)$, (δ, Q, H) and η are defined in (6.15), (6.14), (6.63), (6.51), (6.55), (2.18)*, respectively, and L is the parameter deriving from the approximation of $I_B(m'')$ (see* (6.59)*).*

Proof. If we integrate both the members of (6.62) with respect to time between zero and τ, recalling equations (6.14) and (6.15) we get

$$\begin{aligned}
\int_0^\tau \overline{N}_-(t)dt &\approx Q\tau - \eta[H-L]\int_0^\tau b(t)dt + \delta\int_0^\tau\int_0^t \Phi(t-x)\overline{N}_-(x)dxdt \\
&= Q\tau - \eta[H-L]\int_0^\tau\Big[1-a(t)\Big]dt + \delta\int_0^\tau\int_x^\tau \Phi(t-x)\overline{N}_-(x)dtdx \\
&= Q\tau - \eta[H-L]\left[\tau - \int_0^\tau a(t)dt\right] + \delta\int_0^\tau dx\overline{N}_-(x)\int_0^{\tau-x}\Phi(y)dy \\
&= Q\tau - \eta[H-L]\Big[\tau - A(\tau)\Big] + \delta\int_0^\tau \overline{N}_-(x)[1-a(\tau-x)]dx \\
&= Q\tau - \eta[H-L]\Big[\tau - A(\tau)\Big] + \delta\int_0^\tau \overline{N}_-(x)dx - \delta(a(\cdot)\otimes\overline{N}_-(\cdot))(\tau).
\end{aligned}$$

Hence, it is possible to find the convolution $(a(\cdot)\otimes\overline{N}_-(\cdot))(\tau)$ as a function of the integral $\int_0^\tau \overline{N}_-(t)dt$ and viceversa, as stated in the Lemma. $\square$

Using equation (6.65), we can rewrite the function $L(0;\tau)$ (equation (6.49)) only in terms of $\int_0^\tau \overline{N}_-(t)dt$:

$$\begin{aligned}
-\frac{\ln \mathbf{P}\{\tau\}}{\varpi} =& L(0;\tau) \\
\approx& \frac{\eta}{1-\eta}\left[\tau\frac{Q-\eta[H-L]}{\delta}+\frac{\eta[H-L]}{\delta}A(\tau)+\frac{\delta-1}{\delta}\int_0^\tau \overline{N}_-(t)dt\right] \\
&+\int_0^\tau \overline{N}_-(t)dt \\
=& \int_0^\tau \overline{N}_-(t)dt\Big[\frac{\eta}{1-\eta}\frac{\delta-1}{\delta}+1\Big]+\tau\eta\frac{Q-\eta[H-L]}{\delta(1-\eta)}+\frac{\eta^2[H-L]}{\delta(1-\eta)}A(\tau) \\
=& \frac{\delta-1}{\delta}\Delta\int_0^\tau \overline{N}_-(t)dt+\tau\frac{\eta}{1-\eta}\frac{\widetilde{Q}}{\delta}+\frac{\eta^2[H-L]}{\delta(1-\eta)}A(\tau), \qquad (6.66)
\end{aligned}$$

where $H \approx 1$ and we have set

$$\Delta = \frac{\eta}{1-\eta}-\frac{\delta}{1-\delta} \qquad (6.67)$$

and

$$\widetilde{Q} = Q-\eta[H-L]. \qquad (6.68)$$

Remark 16. The probability of having zero observable events in $[0,\tau]$, given in (6.66), can be compared with the one obtained in [Saichev and Sornette, 2007] for the classical time-magnitude ETAS model, i.e., when $C_1 = 0$ and therefore $H = L = 1$. In this case, equation (6.66) becomes

$$-\frac{\ln \mathbf{P}\{\tau\}}{\varpi} \approx \frac{\delta-1}{\delta}\Delta\int_0^\tau \overline{N}_-(t)dt+\tau\frac{\eta Q}{(1-\eta)\delta},$$

which coincides with formula (C20) of page 24 in [Saichev and Sornette, 2007].

If we could have an expression for $\int_0^\tau \overline{N}_-(t)dt$, using equation (6.66) (and then (6.37)), it would be possible to find the probability of having zero observable events in $[0,\tau]$.

To the aim of obtaining an explicit expression of the function $\overline{N}_-(t)$, we use the standard technique of Laplace transform.

Theorem 13. *When $H \approx 1$, it holds*

$$\begin{aligned}
\overline{N}_-(t) \approx& \frac{2e^{\sigma t}}{\pi}\int_0^\infty \Re\left(\frac{Q-\eta[H-L]\theta[(\sigma+i\xi)c]^\theta e^{(\sigma+i\xi)c}\Gamma(-\theta,(\sigma+i\xi)c)}{(\sigma+i\xi)[1-\delta\theta[(\sigma+i\xi)c]^\theta e^{(\sigma+i\xi)c}\Gamma(-\theta,(\sigma+i\xi)c)]}\right) \\
&\cdot \cos\xi t d\xi, \qquad (6.69)
\end{aligned}$$

for $\sigma + i\xi \in \mathbb{C}$, θ, c the parameters of the Omori-Utsu law (6.3) *and with the notations of Lemma 7.*

Proof. Let's compute the Laplace transform $\mathfrak{L}[\overline{N}_-(t)](s)$ of the function $\overline{N}_-(t)$. Recalling equation (6.62) we get, for $s \in \mathbb{C}$,

$$\begin{aligned}
&\mathfrak{L}[\overline{N}_-(t)](s) = \int_0^\infty e^{-st}\overline{N}_-(t)dt \\
&\approx \int_0^\infty e^{-st}Qdt - \eta[H-L]\int_0^\infty e^{-st}b(t)dt + \delta\int_0^\infty e^{-st}(\Phi(\cdot)\otimes\overline{N}_-(\cdot))(t)dt \\
&=\frac{Q}{s} - \eta[H-L]\int_0^\infty e^{-st}\Big[1-a(t)\Big]dt + \delta\int_0^\infty e^{-st}\Phi(t)dt\int_0^\infty e^{-st}\overline{N}_-(t)dt \\
&=\frac{Q}{s} - \frac{\eta[H-L]}{s} + \eta[H-L]\int_0^\infty e^{-st}a(t)dt + \delta\int_0^\infty e^{-st}\Phi(t)dt\cdot\mathfrak{L}[\overline{N}_-(t)](s).
\end{aligned}$$

We then have

$$\mathfrak{L}[\overline{N}_-(t)](s) \approx \frac{1}{1-\delta\int_0^\infty e^{-st}\Phi(t)dt}\left[\frac{Q-\eta[H-L]}{s} + \eta[H-L]\int_0^\infty e^{-st}a(t)dt\right]. \tag{6.70}$$

Recalling (6.7) and (6.15), we get

$$\begin{aligned}
\int_0^\infty e^{-st}a(t)dt &= \int_0^\infty e^{-st}\int_t^\infty \Phi(x)dxdt \\
&= \int_0^\infty e^{-st}\left[1-\int_0^t\Phi(x)dx\right]dt \\
&= \int_0^\infty e^{-st}dt - \int_0^\infty e^{-st}\int_0^t\Phi(x)dxdt \\
&= \frac{1}{s} - \int_0^\infty \Phi(x)\int_x^\infty e^{-st}dtdx \\
&= \frac{1}{s}\left[1-\int_0^\infty e^{-sx}\Phi(x)dx\right].
\end{aligned}$$

Consequently,

$$\int_0^\infty e^{-st}\overline{N}_-(t)dt \approx \frac{1}{s[1-\delta\int_0^\infty e^{-st}\Phi(t)dt]}\left[Q-\eta[H-L]\int_0^\infty e^{-st}\Phi(t)dt\right]. \tag{6.71}$$

Let's notice that the above function of the complex variable s has one pole in $s=0$. Furthermore, there are no poles with real part larger than zero. In fact, the other factor in the denominator is never zero for $\Re(s) > 0$, since $\delta < 1$ and

$$\begin{aligned}\left|\int_0^\infty e^{-st}\Phi(t)dt\right| \leq \int_0^\infty |e^{-st}\Phi(t)|dt \\ &< \int_0^\infty |\Phi(t)|dt \\ &= 1.\end{aligned}$$

Let's then compute the Laplace transform of the function $\Phi(\cdot)$ appearing in (6.71). Recalling equation (6.3), we have

$$\begin{aligned}\int_0^\infty e^{-st}\Phi(t)dt &= \theta c^\theta \int_0^\infty e^{-st}(t+c)^{-1-\theta}dt \\ &= \theta c^\theta \int_0^\infty e^{-x}\left(\frac{x}{s}+c\right)^{-1-\theta}\frac{1}{s}dx \\ &= \theta(sc)^\theta e^{sc}\int_0^\infty e^{-(x+sc)}(x+sc)^{-\theta-1}dx \\ &= \theta(sc)^\theta e^{sc}\Gamma(-\theta, sc),\end{aligned} \tag{6.72}$$

where

$$\Gamma(\ell, z) = \int_z^\infty e^{-t}t^{\ell-1}dt = \int_0^\infty e^{-(x+z)}(x+z)^{\ell-1}dx, \quad |\arg z| < \pi,$$

is the incomplete Gamma function (see [Bateman, 1953; Temme, 1996]). Let's notice that the integral in the third line of (6.72) is the explicit form of the complex integral along the horizontal half-line starting from the complex point sc.

Substituting the solution (6.72) in equation (6.71), we obtain

$$\mathfrak{L}[\overline{N}_-(t)](s) \approx \frac{1}{s[1-\delta\theta(sc)^\theta e^{sc}\Gamma(-\theta, sc)]}\Big[Q - \eta[H-L]\theta(sc)^\theta e^{sc}\Gamma(-\theta, sc)\Big]. \tag{6.73}$$

Now, one possible way to find the function $\overline{N}_-(t)$ consists of computing the inverse Laplace transform through the Bromwich integral:

$$\overline{N}_-(t) = \frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} e^{st}\mathfrak{L}[\overline{N}_-(t)](s)ds,$$

where σ is any positive number. We are going to proceed without contour integration (see [Berberan-Santos, 2005]). We get

$$
\begin{aligned}
\overline{N}_-(t) =& \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{(\sigma+i\xi)t}\mathfrak{L}[\overline{N}_-(t)](\sigma+i\xi)d\xi \\
=& \frac{e^{\sigma t}}{2\pi}\int_{-\infty}^{+\infty} e^{i\xi t}\mathfrak{L}[\overline{N}_-(t)](\sigma+i\xi)d\xi \\
=& \frac{e^{\sigma t}}{2\pi}\int_{-\infty}^{+\infty} \Big[\mathfrak{L}[\overline{N}_-(t)](\sigma+i\xi)\cos\xi t + i\mathfrak{L}[\overline{N}_-(t)](\sigma+i\xi)\sin\xi t\Big]d\xi \\
=& \frac{e^{\sigma t}}{2\pi}\Bigg\{\int_{-\infty}^{+\infty} \Big[\Re\big(\mathfrak{L}[\overline{N}_-(t)](\sigma+i\xi)\big) + i\Im\big(\mathfrak{L}[\overline{N}_-(t)](\sigma+i\xi)\big)\Big]\cos\xi t d\xi \\
& + i\int_{-\infty}^{+\infty} \Big[\Re\big(\mathfrak{L}[\overline{N}_-(t)](\sigma+i\xi)\big) + i\Im\big(\mathfrak{L}[\overline{N}_-(t)](\sigma+i\xi)\big)\Big]\sin\xi t d\xi\Bigg\} \\
=& \frac{e^{\sigma t}}{2\pi}\Bigg\{\int_{-\infty}^{+\infty} \Big[\Re\big(\mathfrak{L}[\overline{N}_-(t)](\sigma+i\xi)\big)\cos\xi t - \Im\big(\mathfrak{L}[\overline{N}_-(t)](\sigma+i\xi)\big)\sin\xi t\Big]d\xi \\
& + i\int_{-\infty}^{+\infty} \Big[\Im\big(\mathfrak{L}[\overline{N}_-(t)](\sigma+i\xi)\big)\cos\xi t + \Re\big(\mathfrak{L}[\overline{N}_-(t)](\sigma+i\xi)\big)\sin\xi t\Big]d\xi\Bigg\},
\end{aligned}
\tag{6.74}
$$

where

$$
\Re\big(\mathfrak{L}[\overline{N}_-(t)](\sigma+i\xi)\big) = \Re\left(\int_0^\infty \overline{N}_-(t)e^{-(\sigma+i\xi)t}dt\right) = \int_0^\infty e^{-\sigma t}\overline{N}_-(t)\cos\xi t dt
\tag{6.75}
$$

and

$$
\Im\big(\mathfrak{L}[\overline{N}_-(t)](\sigma+i\xi)\big) = \Im\left(\int_0^\infty \overline{N}_-(t)e^{-(\sigma+i\xi)t}dt\right) = -\int_0^\infty e^{-\sigma t}\overline{N}_-(t)\sin\xi t dt.
\tag{6.76}
$$

Since $\overline{N}_-(t)$ is a real function, we have

$$
\int_{-\infty}^{+\infty} \Big[\Im\big(\mathfrak{L}[\overline{N}_-(t)](\sigma+i\xi)\big)\cos\xi t + \Re\big(\mathfrak{L}[\overline{N}_-(t)](\sigma+i\xi)\big)\sin\xi t\Big]d\xi = 0;
$$

actually, this also follows easily from the odd-parity of the integrand. Moreover, observing that the function

$$
W(\xi) = \Re\big(\mathfrak{L}[\overline{N}_-(t)](\sigma+i\xi)\big)\cos\xi t - \Im\big(\mathfrak{L}[\overline{N}_-(t)](\sigma+i\xi)\big)\sin\xi t
$$

is of even-parity, the last expression in (6.74) becomes:

$$\frac{e^{\sigma t}}{\pi}\int_0^{\infty}\Big[\Re\big(\mathfrak{L}[\overline{N}_-(t)](\sigma+i\xi)\big)\cos\xi t - \Im\big(\mathfrak{L}[\overline{N}_-(t)](\sigma+i\xi)\big)\sin\xi t\Big]d\xi.$$

Finally, recalling (6.75), (6.76) and that $\overline{N}_-(t) = 0$ for $t < 0$, we have that

$$\begin{aligned}
&\int_{-\infty}^{+\infty}\Big[\Re\big(\mathfrak{L}[\overline{N}_-(t)](\sigma+i\xi)\big)\cos\xi t + \Im\big(\mathfrak{L}[\overline{N}_-(t)](\sigma+i\xi)\big)\sin\xi t\Big]d\xi\\
&= \int_{-\infty}^{+\infty}\int_0^{\infty} e^{-\sigma x}\overline{N}_-(x)\cos\xi x dx\ \cos\xi t d\xi\\
&\quad + \int_{-\infty}^{+\infty} - \int_0^{\infty} e^{-\sigma x}\overline{N}_-(x)\sin\xi x dx\ \sin\xi t d\xi\\
&= \int_0^{\infty} e^{-\sigma x}\overline{N}_-(x)\int_{-\infty}^{+\infty}[\cos\xi x\cos\xi t - \sin\xi x\sin\xi t]d\xi dx\\
&= \int_0^{\infty} e^{-\sigma x}\overline{N}_-(x)\int_{-\infty}^{+\infty}\cos[\xi(x+t)]d\xi dx\\
&= \int_0^{\infty} e^{-\sigma x}\overline{N}_-(x)\delta(x+t)dx\\
&= e^{\sigma t}\overline{N}_-(-t)\\
&= 0,
\end{aligned}$$

where $\delta(y)$ is the Dirac function having value one when $y = 0$, zero otherwise. Then, the thesis (6.69) of the theorem is obtained:

$$\begin{aligned}
\overline{N}_-(t) =& \frac{2e^{\sigma t}}{\pi}\int_0^{\infty}\Re\big(\mathfrak{L}[\overline{N}_-(t)](\sigma+i\xi)\big)\cos\xi t d\xi\\
\approx& \frac{2e^{\sigma t}}{\pi}\int_0^{\infty}\Re\left(\frac{Q-\eta[H-L]\theta[(\sigma+i\xi)c]^{\theta}e^{(\sigma+i\xi)c}\Gamma(-\theta,(\sigma+i\xi)c)}{(\sigma+i\xi)[1-\delta\theta[(\sigma+i\xi)c]^{\theta}e^{(\sigma+i\xi)c}\Gamma(-\theta,(\sigma+i\xi)c)]}\right)\\
&\cdot\cos\xi t d\xi.
\end{aligned}$$

□

One may use a numerical approach to compute the last integral and then to find the required function $\overline{N}_-(t)$. Once done this, one must integrate it with respect to time. Then, the obtained result must be substituted in equation (6.66). In conclusion, we have proved the following theorem.

Theorem 14. *In the case $m_c \geq m_0$, approximatively for small τ and for $H \approx 1$, the probability of having zero observable events in $[0, \tau]$ is*

$$
\begin{aligned}
-\frac{\ln \boldsymbol{P}\{\tau\}}{\varpi} \approx & \frac{\delta - 1}{\delta} \Delta \int_0^\tau \Bigg[\frac{2e^{\sigma t}}{\pi} \\
& \cdot \int_0^\infty \Re \left(\frac{Q - \eta [H-L] \theta [(\sigma + i\xi)c]^\theta e^{(\sigma+i\xi)c} \Gamma(-\theta, (\sigma+i\xi)c)}{(\sigma+i\xi)[1 - \delta\theta[(\sigma+i\xi)c]^\theta e^{(\sigma+i\xi)c}\Gamma(-\theta,(\sigma+i\xi)c)]} \right) \\
& \cdot \cos \xi t d\xi \Bigg] dt + \tau \frac{\eta}{1-\eta} \frac{\widetilde{Q}}{\delta} + \frac{\eta^2 [H-L]}{\delta(1-\eta)} A(\tau),
\end{aligned}
\tag{6.77}
$$

with $\widetilde{Q}$ and Δ defined in (6.68) *and* (6.67)*, respectively, and the other notations as in Lemma 7.*

Furthermore, we can repeat the same heuristic reasoning as in the final part of Section 6.2, according to which, if

$$
\hat{L}(0;\tau) = \frac{\delta - 1}{\delta} \Delta \int_0^\tau \overline{N}_-(t) dt + \tau \frac{\eta}{1-\eta} \frac{\widetilde{Q}}{\delta} + \frac{\eta^2 [H-L]}{\delta(1-\eta)} A(\tau),
$$

that is the approximation obtained for $L(0;\tau)$ in the case $m_c \geq m_0$, it holds also

$$
\frac{d^k L(0;\tau)}{d\tau^k} \approx \frac{d^k \hat{L}(0;\tau)}{d\tau^k},
$$

for $k = 1, 2$. Consequently, we can derive the interevent time density by computing the second derivative of the probability in (6.77), in fact

$$
F_{inter}(\tau) = \frac{1}{\bar{\lambda}} \frac{d^2}{d\tau^2} \mathbf{P}\{\tau\}. \tag{6.78}
$$

Let's notice that in the case $m_c > m_0$, differently from the one $m_c = m_0$, the transition probability density function plays a role in $\mathbf{P}\{\tau\}$ and then in $F_{inter}(\tau)$. More precisely, the parameters of the above transition probability distribution appears in the expression (6.77) through some of the constants, as for example δ. The reason is that, when $m_c = m_0$, all the events that we observe may trigger its own offspring; on the other hand, when $m_c > m_0$, we do not observe all the triggering events. For example, if we consider at first the process obtained with the condition $m_c = m_0$ and then we increase m_c, then we get a process with a lower number of both spontaneous and triggered events. In general, we have a different number of shocks when $m_c > m_0$ and then a different process associated with the events. The mean number of events depends on the magnitudes and then, when considering a given time interval, the higher (lower) is the number of shocks, the lower (higher) are the distances between the events. Because of this, together with the fact that the proportion of triggered and triggering events changes and observing that our hypothesis of magnitude correlation is based on the proportion between these two kinds of shocks, it is intuitive to understand that, when $m_c > m_0$, the interevent time density explicitly depends on our hypothesis.

6.3.1 Another approximation for small τ

Because of the difficulty of the computation of both $\overline{N}_-(t)$ and the interevent time density, we simplify the case we are analyzing with a further approximation.

Theorem 15. *By considering another approximation for small τ, the results in Theorem 14 can be obtained explicitly as*

$$-\frac{\ln \boldsymbol{P}\{\tau\}}{\varpi} \approx \Big[\tau\widetilde{Q} + \eta[H-L]A(\tau)\Big]\Big[\frac{1-\eta+\eta a(\tau)}{(1-\eta)[1-\delta+\delta a(\tau)]}\Big], \tag{6.79}$$

with the notations as in Theorem 14.

Proof. If we consider again τ sufficiently small, we can set $a(\tau - t) \approx a(\tau)$ in (6.64), writing the extended integral of the convolution. We get

$$(1-\delta)\int_0^\tau \overline{N}_-(t)dt \approx \tau\Big[Q - \eta[H-L]\Big] + \eta[H-L]A(\tau) - \delta\int_0^\tau a(\tau)\overline{N}_-(t)dt,$$

from which we have

$$\int_0^\tau \overline{N}_-(t)dt \approx \frac{1}{1-\delta+\delta a(\tau)}\Big\{\tau\Big[Q-\eta[H-L]\Big] + \eta[H-L]A(\tau)\Big\}. \tag{6.80}$$

The above-mentioned approximation makes sense since, if $t \in [0, \tau]$ and τ is small, so is t.

We can then conveniently substitute the expression (6.80) of $\int_0^\tau \overline{N}_-(t)dt$ in (6.66):

$$\begin{aligned}-\frac{\ln \mathbf{P}\{\tau\}}{\varpi} \approx & \frac{\delta-1}{\delta}\Delta\frac{1}{1-\delta+\delta a(\tau)}\Big\{\tau\Big[Q-\eta[H-L]\Big] + \eta[H-L]A(\tau)\Big\} \\ & + \tau\frac{\widetilde{Q}}{\delta}\frac{\eta}{1-\eta} + \frac{\eta^2}{\delta(1-\eta)}[H-L]A(\tau)\end{aligned}$$

$$
\begin{aligned}
=&\frac{\delta-1}{\delta}\frac{\Delta\widetilde{Q}\tau}{1-\delta+\delta a(\tau)}+\frac{\delta-1}{\delta}\frac{\Delta}{1-\delta+\delta a(\tau)}\eta[H-L]A(\tau)\\
&+\tau\frac{\widetilde{Q}}{\delta}\frac{\eta}{1-\eta}+\frac{\eta^2}{\delta(1-\eta)}[H-L]A(\tau)\\
=&\frac{\tau\widetilde{Q}}{\delta}\left[\frac{(\delta-1)\Delta}{1-\delta+\delta a(\tau)}+\frac{\eta}{1-\eta}\right]\\
&+\frac{\eta[H-L]A(\tau)}{\delta}\left[\frac{(\delta-1)\Delta}{1-\delta+\delta a(\tau)}+\frac{\eta}{1-\eta}\right]\\
=&\left[\tau\widetilde{Q}+\eta[H-L]A(\tau)\right]\left[\frac{\frac{\delta-1}{\delta}\Delta}{1-\delta+\delta a(\tau)}+\frac{\eta}{(1-\eta)\delta}\right]\\
=&\left[\tau\widetilde{Q}+\eta[H-L]A(\tau)\right]\left[\frac{1-\eta+\eta a(\tau)}{(1-\eta)[1-\delta+\delta a(\tau)]}\right],
\end{aligned}
$$

where we have used equation (6.68) and the last equality follows because

$$
\begin{aligned}
\frac{\frac{\delta-1}{\delta}\Delta}{1-\delta+\delta a(\tau)}+\frac{\eta}{(1-\eta)\delta}=&\frac{\delta-1}{\delta}\left[\frac{\eta}{1-\eta}+\frac{\delta}{\delta-1}\right]\frac{1}{1-\delta+\delta a(\tau)}+\frac{\eta}{(1-\eta)\delta}\\
=&\left[\frac{\eta(\delta-1)}{(1-\eta)\delta}+1\right]\frac{1}{1-\delta+\delta a(\tau)}+\frac{\eta}{(1-\eta)\delta}\\
=&\frac{\eta(\delta-1)+(1-\eta)\delta}{(1-\eta)\delta[1-\delta+\delta a(\tau)]}+\frac{\eta}{(1-\eta)\delta}\\
=&\frac{\eta(\delta-1)+(1-\eta)\delta+\eta(1-\delta)+\eta\delta a(\tau)}{(1-\eta)\delta[1-\delta+\delta a(\tau)]}\\
=&\frac{1-\eta+\eta a(\tau)}{(1-\eta)[1-\delta+\delta a(\tau)]}.
\end{aligned}
$$

In the latter calculus we have used equation (6.67). We have then obtained (6.79) and the thesis of the theorem is proved. □

Once again, we can heuristically derive the probability density function $F_{inter}(\tau)$ relative to the interevent time. More precisely, setting

$$
\hat{L}(0;\tau)=\left[\tau\widetilde{Q}+\eta[H-L]A(\tau)\right]\left[\frac{1-\eta+\eta a(\tau)}{(1-\eta)[1-\delta+\delta a(\tau)]}\right],
$$

that is the approximation obtained for $L(0;\tau)$ in the case $m_c \geq m_0$ when considering the further approximation for small τ as we are supposing in the current subsection, we can suppose that it holds also

$$\frac{d^k L(0;\tau)}{d\tau^k} \approx \frac{d^k \hat{L}(0;\tau)}{d\tau^k},$$

for $k = 1, 2$. Then, using (6.78) and the calculations to obtain the second derivative of $\mathbf{P}\{\tau\} = e^{-\varpi L(0;\tau)}$ (equation (C.1)), shown in Appendix C, we get the following result.

$$\begin{aligned}
F_{inter}(\tau) \approx \frac{1}{\overline{\lambda}} \exp&\left\{-\left[\tau\widetilde{Q} + \varpi\eta[H-L]A(\tau)\right]\left[\frac{1-\eta+\eta a(\tau)}{(1-\eta)[1-\delta+\delta a(\tau)]}\right]\right\} \\
&\cdot \left\{\left[\left(\widetilde{Q} + \varpi\eta[H-L]a(\tau)\right)\left(\frac{1-\eta+\eta a(\tau)}{(1-\eta)[1-\delta+\delta a(\tau)]}\right)\right.\right. \\
&+ \left.\left(\tau\widetilde{Q} + \varpi\eta[H-L]A(\tau)\right)\left(\frac{\Phi(\tau)(\delta-\eta)}{(1-\eta)[1-\delta+\delta a(\tau)]^2}\right)\right]^2 \\
&- \left[-\varpi\eta[H-L]\Phi(\tau)\left(\frac{1-\eta+\eta a(\tau)}{(1-\eta)[1-\delta+\delta a(\tau)]}\right)\right. \\
&+ 2\left(\widetilde{Q} + \varpi\eta[H-L]a(\tau)\right)\left(\frac{\Phi(\tau)(\delta-\eta)}{(1-\eta)[1-\delta+\delta a(\tau)]^2}\right) \\
&+ \left.\left.\left(\tau\widetilde{Q} + \varpi\eta[H-L]A(\tau)\right)\frac{(\delta-\eta)\Phi(\tau)[(\theta+1)(\delta-1)+\delta a(\tau)(\theta-1)]}{(1-\eta)(\tau+c)[1-\delta+\delta a(\tau)]^3}\right]\right\},
\end{aligned} \tag{6.81}$$

Remark 17. In the proof of Theorem 15, we have obtained that

$$\int_0^\tau \overline{N}_-(t)dt \approx \frac{1}{1-\delta+\delta a(\tau)}\left\{\tau\left[Q-\eta[H-L]\right] + \eta[H-L]A(\tau)\right\}. \tag{6.82}$$

Now, we can compare this approximation with the one obtained in [Saichev and Sornette, 2007] for the classical time-magnitude ETAS model, that is when $C_1 = 0$ and therefore $H = L = 1$. In this case, the above result (6.82) becomes

$$\int_0^\tau \overline{N}_-(t)dt \approx \frac{\tau Q}{1-\delta+\delta a(\tau)},$$

that is the approximation $C(25)$ in Appendix C by Saichev and Sornette. Consequently, their equation (46) of page 8 coincides with the probability (6.79), obtained from (6.66) with the further approximation for small τ, when $C_1 = 0$ and then $H = L = 1$.

Let's conclude saying that, as previously discussed at the end of Section 6.3, we notice again that in the case $m_c > m_0$ our hypothesis of a transition probability density function for the magnitudes of triggered events appears in the probability of having zero observable events in $[0, \tau]$ and then in the interevent time density $F_{inter}(\tau)$.

6.4 Conclusions

In this chapter we theoretically analyzed our new version of the ETAS model, where triggered event's magnitudes are probabilistically dependent on triggering event's ones. Thanks to the tool of the probability generating function and the Palm theory, we obtained a closed-form approximation of density $F_{inter}(\tau)$ of interevent time for small values of τ. More precisely, the closed-form is found by a linear approximation around zero of an exponential function. This is the only approximation needed for the case $m_c = m_0$. In the general case $m_c \geq m_0$, we have seen that it can be used the additional approximation $a(\tau - t) \approx a(\tau)$ to get the result in a closed-form. However, without using this latter approximation, we were able to find the studied density function in terms of the inverse Laplace transform of a suitable function. The results obtained show that our hypothesis of magnitude correlation plays a role when we do not observe all the triggering events, as it happens in practice. In fact, if we consider the particular case $m_c = m_0$, that is if we observe all the events that are able to trigger, the density obtained for the interevent time doesn't depend on the new transition probability proposed for triggered events' magnitude. On the other hand, in the case $m > m_0$, in which not all the triggering events are observable, one can see the influence of the above-mentioned proposal on some of the constants appearing in $F_{inter}(\tau)$. The interevent time density then depends on the Omori-Utsu law, on its temporal integrals and on the new transition probability density function for triggered events' magnitude.

Overall conclusions and future work

In this thesis we have proposed and theoretically analyzed a new version of the pure temporal Epidemic Type Aftershock Sequence model for earthquakes: the ETAS model with correlated triggered events' magnitudes. The version we proposed is based on a new transition probability density function for the magnitudes of triggered events, depending on the magnitude of the triggering shocks. More precisely, we take into account past seismicity in terms of mother/daughter relations between events for modeling the above magnitudes.

The process we have considered is a marked Hawkes process such that:

- the immigrants are a homogeneous Poisson process with independent marks consisting of the magnitudes;
- each event gives birth to a non-homogeneous Poisson process, independently of the other shocks and of the process of the immigrants, with marks again consisting of the magnitudes, but now depending on the magnitudes of the relative triggering events.

The conditional intensity function, completely characterizing our new model, is given by

$$\lambda(t, m|\mathcal{H}_t) = p(m)\varpi + \sum_{\{i|t_i<t\}} \varrho(m_i)\Phi(t-t_i)p(m|m_i),$$

where ϖ is the average rate corresponding to the events belonging to the spontaneous component of the process; $p(\cdot)$ is the Gutenberg-Richter law (equation (2.2)); $\varrho(\cdot)$ is the productivity law (equation (2.1)); $\Phi(\cdot)$ is the Omori law (equation (2.5)) and $p(\cdot|\cdot)$ is a transition probability density function for the magnitudes of the triggered shocks. The model we propose is very interesting, since it allows to describe seismicity in a more realistic way: the magnitudes of the aftershocks are supposed to be correlated with the ones of the respective triggering events. It is important to notice that the above hypothesis of correlation doesn't complicate the statistical and theoretical analysis; furthermore, as explained in Section 4.1, the conditions for the non-explosion of the process remain the same as for the classical temporal ETAS model: they are $p > 1$, $\beta > a$ and $\kappa < \frac{\beta - a}{\beta}$, for (p, β, a, κ) the parameters of the Gutenberg-Richter, the Omori and the productivity laws. The Laplace functional of the process modeled by the new ETAS model with correlated magnitudes have been derived in Section 4.2, in order to obtain a useful tool for the analysis of several characteristics of the model itself. To our knowledge, in the literature there is only one work concerning a similar analysis on the Laplace functional: independently of us, Roueff *et al.* repeat the analysis for a locally stationary Hawkes processes [Roueff et al., 2015], but they present a condition for the non-explosion of their process not satisfied by our magnitude correlated model.

In order to support the new hypothesis of magnitudes correlation and then to justify the proposal of our new model, we firstly have performed two different kinds of analysis of four seismic real catalogs: three Italian datasets and a Californian one. The results obtained for all the catalogs, with the exception of the one related to the whole Italy, show that our hypothesis is well supported by the behavior of the real data. By looking at the kernel density estimation of the aftershocks' magnitudes, one can see that when triggering events' magnitudes increase, the probability of having higher triggered events' magnitudes increases, too; it instead decreases for low values of the mother shocks' magnitude. This behavior is not evident for the whole Italian catalog (Fig. 3.10) because in this case the area considered is very wide and the analysis puts in causal relation events that are temporally close, but spatially separated. The plots obtained by the two types of analysis for the other three catalogs, show that the density of the aftershocks' magnitudes changes in shape when the triggering events' magnitude increases and in some cases, for high values of the latter, it has a relative maximum (Figures 3.11, 3.13, 3.15). The means of the aftershocks' densities are also increasing, as one can see from the plots obtained with the two kinds of analysis for all the catalogs, again with the exception of the whole Italian one (Figures 3.17, 3.18). Even if there are small differences between the figures of the catalogs relative to L'Aquila till 2012, L'Aquila till April the 5^{th}, 2009 and the Southern California, caused for example by the different number of events due to the presence or not of a strong shock in the data, the qualitative behavior is the same. This ensures that the magnitudes dependence we are stating is not connected to the area under examination; furthermore, it is not induced by the occurrence of shocks with high magnitude or by some problems of incompleteness.

By looking at the variations of the estimated magnitude densities of triggered shocks, with respect to the triggering events' magnitudes, we can conclude that there is correlation between the magnitudes of events in mother/daughter causal relation. However, we have to recall that the Gutenberg-Richter law is a well-validated model for the magnitudes when not considering past seismicity. Taking this into account and considering the experimental results obtained, we have proposed the following explicit form of the transition probability density function $p(m''|m')$ for triggered events' magnitudes:

$$p(m''|m') = \beta e^{-\beta(m''-m_0)}\left[1 + C_1\big(1 - 2e^{-(\beta-a)(m'-m_0)}\big)\big(1 - 2e^{-\beta(m''-m_0)}\big)\right],$$

for all $0 \leq C_1 \leq 1$. It is such that, averaging over all the possible mother events' magnitudes, we obtain again the Gutenberg-Richter law. This means that the latter, weighted by the productivity law $\varrho(m')$, remains invariant with respect to $p(m''|m')$. Furthermore, again in agreement with the results obtained for the real catalogs, we have that:

- when $m' > \frac{1}{\beta-a}\ln 2+m_0$ ($m' < \frac{1}{\beta-a}\ln 2+m_0$), the transition probability density $p(m''|m')$ is below (above) $p(m'')$ for magnitudes values smaller than $m'' = \frac{1}{\beta}\ln 2 + m_0$, above for higher values;

- when the triggering event's magnitude increases, the transition probability density function increases (decreases) at a point $m'' > \frac{1}{\beta}\ln 2+m_0$ ($m'' < \frac{1}{\beta}\ln 2 + m_0$);

- when $m' \leq \frac{1}{\beta-a}\ln 2 + m_0$, $p(m''|m')$ is always decreasing in m''; when instead $m' > \frac{1}{\beta-a}\ln 2+m_0$, $p(m''|m')$ increases in m'' till a certain maximum, obtained in $m'' = \frac{1}{\beta}\ln\frac{4q(m')}{1+q(m')} + m_0$ and after that it decreases.

The explicit form obtained for $p(m''|m')$ is used for the explicit derivation of the conditional intensity function characterizing the new ETAS model with correlated triggered events' magnitudes, i.e.,

$$\begin{aligned}\lambda(t,m|\mathcal{H}_t) =\,& p(m)\Bigg\{\varpi + \sum_{\{i|t_i<t\}} \kappa e^{a(m_i-m_0)}\frac{p-1}{c}\Big(1+\frac{t-t_i}{c}\Big)^{-p}\Bigg[1 \\ & + C_1\big(1 - 2e^{-(\beta-a)(m_i-m_0)}\big)\big(1 - 2e^{-\beta(m-m_0)}\big)\Bigg]\Bigg\},\end{aligned}$$

where $0 \leq C_1 \leq 1$.

In order to assess the validity of the explicit form proposed for $p(m''|m')$ and to give a stronger support to our hypothesis of magnitude correlation, we have then performed the same two types of analysis of above for some simulated catalogs. We have presented here the analysis concerning only two of the latter, simulated in two different ways. The first catalog is simulated with the Ogata's FORTRAN program [**etasim.f**] for the classical ETAS model. The second catalog is instead simulated using an algorithm obtained by modifying the program [**etasim.f**] in the magnitudes simulation: the background events' magnitudes have been simulated through the Gutenberg-Richter law, instead the aftershocks' ones have been simulated with our transition probability density $p(m''|m')$. We have obtained that the results for the first synthetic catalog are similar to those concerning the whole Italian real catalog (Fig. 5.3). This can be explained obviously by the fact that the classical Ogata's program takes into account only the Gutenberg-Richter law to simulate the magnitudes of all the events, both spontaneous and triggered. On the other hand, the results for the catalog simulated with our new model strongly support our hypothesis of magnitude correlation. In fact, they are very similar to the ones obtained for the three remaining real catalogs (Fig. 5.5). Also the mean agrees with the one relative to these three real datasets (Fig. 5.7). Since we have used our transition density $p(m''|m')$ to simulate the second synthetic catalog, but then we have performed the analyses for this catalog using the set of parameters obtained with the classical Ogata's FORTRAN program for estimation [**etas.f**], the accordance between these results are fundamental to justify the validity of our hypothesis of magnitude correlation. Let's notice that the results obtained for the other simulated catalogs, not reported here, are very similar to the ones for the two catalogs considered.

After the validation of the transition probability density $p(m''|m')$, we have mathematically studied the probability generating function formalism of our new version of the temporal ETAS model. This theoretical analysis allowed us to derive an expression for the interevent time density $F_{inter}(\tau)$, for small τ, thanks to the Palm theory. We recall that this random variable has much importance in terms of seismic hazard. We have considered two cases. In the first one, the completeness magnitude m_c and the reference one m_0 are equal, as often happens in practice; this is a particular case. The second case is instead the general one, in which $m_c \geq m_0$. In both cases we have considered the linear approximation around zero of an exponential function. Thanks to this approximation, we have found a closed-form for the interevent density in the case $m_c = m_0$ (equation (6.26)). The function obtained doesn't depend on the new hypothesis of magnitude correlation. This can be explained because, in this case, all the observable events may generate aftershocks and the process is not influenced by external sources.

Instead, the above exponential approximation is not enough to obtain a closed-form for the general case. In fact, when $m_c \geq m_0$, the density $F_{inter}(\tau)$

has been obtained in terms of the inverse Laplace transform of a suitable function. However, by considering an additional approximation, also in the case $m_c \geq m_0$ we have been able to obtained a closed-form approximation for the interevent time density (equation (6.81)). In this function one can see the "presence" of our hypothesis of magnitude correlation through some constants. The fact that in the general case the magnitude correlations are evident in $F_{inter}(\tau)$ is due to the non-observability of all the triggering events. More precisely, when $m_c > m_0$ we get a process with a different number of events and a different proportion of triggering and triggered shocks. For example, let's consider the process obtained for $m_c = m_0$. It has a certain number of events. Now, for example, by increasing m_c the new process will have a lower number of events. In general, the number of events changes, consequently so does the proportion of background and aftershock events, the proportion of triggered and triggering events and furthermore the interevent time distances. Then, the final process is different. Since the transition probability we propose concerns the mother/daughter relations, the evidence of this probability in $F_{inter}(\tau)$ for the general case is justified. Let's notice that the case in which $m_c = m_0$ is the one considered in the experimental studies, since one doesn't know which value assign to m_0. Nevertheless, this is not the real case. By considering these threshold magnitudes different, we can model the phenomenon in a more precise and realistic way. This is effectively the situation in which one can observe how our hypothesis of magnitude correlation intervenes.

We conclude with the possible problems to work on. First of all, it could be interesting to implement the new model and then to perform a statistical inference study. One could use the maximum likelihood analysis to estimate the parameters of this new version, or could try with some other technique. Through the Akaike's Information Criterion, one could also compare the goodness of fit of our new model with respect to other models. Still an open question is the asymptotic study of the interevent time density for long times, in the case of transition magnitude probability density function. This could be interesting in terms of extinction of the seismic sequence. Furthermore, it would be fundamental to include the spatial component of the events. In fact, it is a very important aspect to take into account for the mathematical analysis of the seismic phenomenon. From a theoretical point of view, it could be also of interest the study of other characteristics of the model proposed, such that the correlations between events' magnitudes in some fixed subintervals, or events' occurrence times in some fixed time windows. This could be done with the tool of the Laplace functional.

Appendix A

A brief review on the probability generating function and the statistical tests for seismic residual analysis

In this Appendix, we briefly recall some notions about the probability generating functions and the statistical tests of Kolmogorov-Smirnov and Runs. We will present only the relative properties useful for the analysis in this thesis.

A.1 Probability generating function and some of its fundamental properties

The probability generating function is a very useful tool in probability theory, in fact it completely describes the law of a random variable (see [Grimmett and Stirzaker, 2001]). Given a non-negative random variable X, it is defined as $G_x(z) = \mathbb{E}[z^X]$ and one can easily prove that

$$\frac{1}{n!}\frac{d^n}{dz^n}G_X(z)\bigg|_{z=0} = \mathbb{P}\{X = n\}.$$

Among its several properties, we state here the following two, used for the mathematical characterization of the ETAS model in this thesis.

Proposition 6. *Let $\{X_1, X_2, \dots\}$ be a sequence of independent and identically distributed random variables, with generating function G_X. Let also N*

be a discrete positive random variable independent of $X_i \quad \forall i$*, with generating function* G_N*. Then, given the random variable* S *defined as*

$$S = \sum_{i=1}^{N} X_i,$$

one has

$$G_S(z) = G_N[G_X(z)]. \tag{A.1}$$

Proposition 7. *Given two independent random variables* X *and* Y*, it holds*

$$G_{X+Y}(z) = G_X(z)G_Y(z). \tag{A.2}$$

A.2 The Kolmogorov-Smirnov test

The *Kolmogorov-Smirnov test* is a non-parametric statistical test used for the residual analysis of the ETAS model, in order to assess if the interevent times of the process are exponentially distributed. More precisely, it is used the *one-sample Kolmogorov-Smirnov test* and this is the case presented here.

Starting from a sample of i.i.d. observations $(X_1, \dots, X_n)$, we want to test whether the $X's$ have been drawn from a specified distribution F_0. Then, supposing that the unknown true distribution of the sample is F_X, the null hypothesis to test is

$$H_0 : F_X = F_0.$$

Since the *Glivenko-Cantelli theorem* states that the empirical distribution function

$$\hat{F}_n(x) = \frac{\# of X_i \leq x}{n},$$

for all $x \in \mathbb{R}$, approaches uniformly the true distribution F_X as n increases, it follows that the deviation between the true function and the empirical distribution function should be small for all values of x, when n is large. Then, a natural statistic for the accuracy of the estimate is

$$D_n = \sup_x |\hat{F}_n(x) - F_X(x)|,$$

for any n. This is the *Kolmogorov-Smirnov one-sample statistic*. For the Kolmogorov-Smirnov one-sample goodness-of-fit test, when the null hypothesis is true, it is then expected that $|\hat{F}_n(x) - F_0(x)| < \varepsilon$, for all x and for large n. The null hypothesis is instead rejected in favour of

$$H_1 : F_X(x) \neq F_0(x), \text{ for some } x \in \mathbb{R},$$

for large absolute values of the above difference, that is when $\sup_x |\hat{F}_n(x) - F_X(x)| > D_{n,\alpha}$, for a probability (significance level) α [Gibbons and Chakraborti, 2003].

A.3 The Runs test

The *Runs test* is a non-parametric statistical test used for the residual analysis of the ETAS model as the Kolmogorov-Smirnov test, but it is used to assess whether the interevent times of the process are independent. It is a test of randomness based on the total number or *runs*, that are sequences of one or more types of elements, preceded and succeeded by a different element or no element at all [Gibbons and Chakraborti, 2003]. If the sequence exhibits some pattern of tendency of some type of element, then randomness is violated. One way of testing randomness is to look at the number of runs in one sequence. In particular, we can consider the randomness of runs about the mean of the sample: we take positive sign for the data larger than the mean of the sample, negative sign otherwise. Then, we have an ordered sequence of n elements, of which n_1 are of type one and n_2 are of type two, with $n_1 + n_2 = n$. For the test based on the total number of runs R, one needs the probability distribution of R under the null hypothesis of randomness holds. This probability distribution can be found as

$$\mathbb{P}\{R \leq r\} = \begin{cases} 2\binom{n_1-1}{r/2-1}\binom{n_2-1}{r/2-1} \Big/ \binom{n_1+n_2}{n_1}, & \text{if } r \text{ is even} \\ \binom{n_1-1}{(r-1)/2}\binom{n_2-1}{(r-3)/2} + \binom{n_1-1}{(r-3)/2}\binom{n_2-1}{(r-1)/2} \Big/ \binom{n_1+n_2}{n_1}, & \text{if } r \text{ is odd.} \end{cases}$$

Appendix B

Proof of Lemma 1

In this appendix, we prove Lemma 1 of Section 4.1 in Chapter 4. It is important to observe that, when $p(m|m')$ reduces to $p(m)$, that is $C_1 = 0$, we are in the case of the classical ETAS model and in fact we obtain only one eigenvalue, that is exactly η, as we have seen in Section 2.2.1 of Chapter 2. This is expressed also in condition (4.29) of Section 4.3. Lemma (1) guarantees that the eigenvalue $\eta = \frac{\beta\kappa}{\beta - a}$ is effectively the maximum one.

Let's then proceed with the proof. In what follows, we will consider two sections: in the first one, we derive the eigenvalues and the eigenvectors obtained for the rate of the triggered shocks in the ETAS model with correlated magnitudes; instead, in the second section we will obtain an expression for the expected value of the descendants that will be proved to be finite.

B.1 Eigenvalues and eigenfunctions of the operators $\mathcal{K}_1$ and $\mathcal{K}_2$

Following the reasoning in Subsection 2.2.1 of Chapter 2, let's consider $\ell_1(m')$ and $\ell_2(m)$, which are respectively the left and right eigenfunctions of the rate for the triggered component of the sequence, corresponding to the maximum eigenvalue $\tilde{\eta}$, i.e.,

$$\tilde{\eta}\ell_1(m') = \int_{m_0}^{\infty} \ell_1(m)\varrho(m')p(m|m')dm, \tag{B.1}$$

$$\tilde{\eta}\ell_2(m) = \int_{m_0}^{\infty} \varrho(m')p(m|m')\ell_2(m')dm'. \tag{B.2}$$

The aim is to show that the maximum eigenvalue is strictly less than one.

Let's notice that the above equations have been obtained considering (2.13) and (2.14) and computing the integrals with respect to space and time. As for the classical case, one can easily verify that the condition for the temporal integral to be finite is $p > 1$. Then, the first condition for the non-explosion of the process in (4.4) is obtained.

In order to find the critical parameter for the case of the ETAS model with correlated magnitudes, we can focus either on equation (B.1) or (B.2). In the case of the first equation (B.1), we can define the operator for functions $\mathcal{K}_1$, i.e.,

$$\begin{aligned}\mathcal{K}_1\ell(m') &= \int_{m_0}^{\infty} \ell(m)\varrho(m')p(m|m')dm \\ &= \int_{m_0}^{\infty} \ell(m)\kappa e^{a(m'-m_0)}\beta e^{-\beta(m-m_0)}\Big[1 + C_1\big(1 - 2e^{-(\beta-a)(m'-m_0)}\big) \\ &\quad \cdot \big(1 - 2e^{-\beta(m-m_0)}\big)\Big]dm. \end{aligned} \tag{B.3}$$

On the other hand, by considering the second equation (B.2), we can define the operator that maps a measure (not necessarily a probability measure) with density $\ell(m)$ to a measure with density

$$\begin{aligned}\mathcal{K}_2\ell(m) &= \int_{m_0}^{\infty} \varrho(m')p(m|m')\ell(m')dm' \\ &= \int_{m_0}^{\infty} \kappa e^{a(m'-m_0)}\beta e^{-\beta(m-m_0)}\Big[1 + C_1\big(1 - 2e^{-(\beta-a)(m'-m_0)}\big) \\ &\quad \cdot \big(1 - 2e^{-\beta(m-m_0)}\big)\Big]\ell(m')dm'. \end{aligned} \tag{B.4}$$

Let's focus on the above first case, that is equation (B.3). Under suitable hypothesis of integrability and setting

$$\mathcal{I}(\ell) = \int_{m_0}^{\infty} \ell(m)e^{-\beta(m-m_0)}dm, \tag{B.5}$$

$$\widetilde{\mathcal{I}}(\ell) = \int_{m_0}^{\infty} \ell(m)e^{-2\beta(m-m_0)}dm, \tag{B.6}$$

equation (B.3) can be rewritten as

$$\begin{aligned}\mathcal{K}_1\ell(m') =& \kappa\beta e^{a(m'-m_0)}\Big[1 + C_1\big(1 - 2e^{-(\beta-a)(m'-m_0)}\big)\Big]\mathcal{I}(\ell) \\ &- 2C_1\kappa\beta e^{a(m'-m_0)}\big(1 - 2e^{-(\beta-a)(m'-m_0)}\big)\widetilde{\mathcal{I}}(\ell) \\ =& \kappa\beta e^{a(m'-m_0)}\Big[\mathcal{I}(\ell) + C_1\big[\mathcal{I}(\ell) - 2\widetilde{\mathcal{I}}(\ell)\big]\Big] \\ &- 2C_1\kappa\beta e^{-(\beta-2a)(m'-m_0)}\big[\mathcal{I}(\ell) - 2\widetilde{\mathcal{I}}(\ell)\big]. \end{aligned}$$

The above expression shows that the eigenfunctions are necessarily of the form

$$\ell_1(m') = \alpha_1 \kappa\, e^{a(m'-m_0)} + \alpha_2 \kappa\, e^{(2a-\beta)(m'-m_0)} = \alpha_1 \varrho_a(m') + \alpha_2 \varrho_{2a-\beta}(m'), \tag{B.7}$$

where $\varrho_\vartheta(m') = \kappa\, e^{\vartheta(m'-m_0)}$ is the productivity law of exponent parameter ϑ.

When we take

$$\ell(m) = \varrho_\vartheta(m) = \kappa\, e^{\vartheta(m-m_0)}, \qquad \text{with } \beta > \vartheta, \tag{B.8}$$

then

$$\mathcal{I}(\ell) = \int_{m_0}^{\infty} \kappa e^{-(\beta-\vartheta)(m-m_0)}\, dm = \frac{\kappa}{\beta - \vartheta} \tag{B.9}$$

and

$$\widetilde{\mathcal{I}}(\ell) = \int_{m_0}^{\infty} \kappa e^{-(2\beta-\vartheta)(m-m_0)}\, dm = \frac{\kappa}{2\beta - \vartheta}, \tag{B.10}$$

so that the operator $\mathcal{K}_1$ can be easily computed:

$$\begin{aligned}
\mathcal{K}_1 \ell(m') =& \kappa\beta e^{a(m'-m_0)} \left[\frac{\kappa}{\beta-\vartheta} + C_1 \Big(\frac{\kappa}{\beta-\vartheta} - \frac{2\kappa}{2\beta-\vartheta} \Big) \right] \\
& - 2C_1 \kappa\beta e^{(2a-\beta)(m'-m_0)} \Big(\frac{\kappa}{\beta-\vartheta} - \frac{2\kappa}{2\beta-\vartheta} \Big) \\
=& \beta \left[\frac{\kappa}{\beta-\vartheta} + C_1 \frac{\vartheta\kappa}{(\beta-\vartheta)(2\beta-\vartheta)} \right] \varrho_a(m') \\
& - 2\beta\, C_1\, \frac{\vartheta\kappa}{(\beta-\vartheta)(2\beta-\vartheta)}\, \varrho_{2a-\beta}(m').
\end{aligned}$$

Let's notice that, since ϑ is equal to a or $2a-\beta$, the condition $\beta > \vartheta$ reduces to $\beta > a$. We have then obtained the second condition for the non-explosion of the process in (4.4).

Now, it follows that a function as in (B.7) is an eigenfunction of $\mathcal{K}_1$, with eigenvalue $\tilde{\eta}$, if and only if

$$\begin{aligned}
\tilde{\eta}\big(\alpha_1 \varrho_a + \alpha_2 \varrho_{2a-\beta}\big) =& \mathcal{K}_1(\alpha_1 \varrho_a + \alpha_2 \varrho_{2a-\beta})(m') \\
=& \alpha_1 \beta \left[\frac{\kappa}{\beta - a} + C_1 \frac{a\kappa}{(\beta-a)(2\beta-a)} \right] \varrho_a(m') \\
& - \alpha_1 2\beta\, C_1\, \frac{a\kappa}{(\beta-a)(2\beta-a)}\, \varrho_{2a-\beta}(m') \\
& + \alpha_2 \beta \left[\frac{\kappa}{2(\beta-a)} + C_1 \frac{(2a-\beta)\kappa}{2(\beta-a)(3\beta-2a)} \right] \varrho_a(m') \\
& - \alpha_2 2\beta\, C_1\, \frac{(2a-\beta)\kappa}{2(\beta-a)(3\beta-2a)}\, \varrho_{2a-\beta}(m').
\end{aligned} \tag{B.11}$$

Setting

$$\eta_a(\beta) = \kappa\, \frac{\beta}{\beta - a}$$

in (B.11), we can then write

$$\begin{aligned}\tilde\eta\big(\alpha_1\varrho_a + \alpha_2\varrho_{2a-\beta}\big) = &\left[\alpha_1\eta_a(\beta)\left[1 + C_1\frac{a}{2\beta - a}\right] + \alpha_2\frac{\eta_a(\beta)}{2}\big[1 + C_1\,\frac{2a-\beta}{3\beta - 2a}\big]\right]\\ &\cdot\varrho_a(m') - \left[2\alpha_1\eta_a(\beta)C_1\frac{a}{2\beta - a} + \alpha_2\,\eta_a(\beta)\,C_1\,\frac{2a-\beta}{3\beta-2a}\right]\,\varrho_{2a-\beta}(m').\end{aligned}$$

From the above computation it follows that

$$\begin{cases}\tilde\eta\alpha_1 = \eta_a(\beta)\left[1 + C_1\frac{a}{2\beta-a}\right]\alpha_1 + \frac{\eta_a(\beta)}{2}\left[1 + C_1\,\frac{2a-\beta}{3\beta-2a}\right]\alpha_2\\[2ex] \tilde\eta\alpha_2 = -\eta_a(\beta)2C_1\frac{a}{2\beta-a}\alpha_1 - \eta_a(\beta)C_1\,\frac{2a-\beta}{3\beta-2a}\alpha_2.\end{cases}$$

Setting

$$x_1 = C_1\frac{a}{2\beta - a}, \qquad y_1 = C_1\frac{2a-\beta}{3\beta-2a}, \tag{B.12}$$

we have to find the solution of $det(A - \tilde\eta I) = 0$, where I is the identity matrix and

$$A = \begin{bmatrix} \eta_a(\beta)\,[1 + x_1] & \frac{\eta_a(\beta)}{2}\,[1 + y_1] \\ \\ -\eta_a(\beta)2x_1 & -\eta_a(\beta)y_1\end{bmatrix}.$$

Setting also $\tilde\eta = \eta_a(\beta)\overline{\eta}$, it holds

$$det(A - \tilde\eta I) = 0 \qquad \Leftrightarrow \qquad det(\overline{A} - \overline{\eta} I) = 0$$

and then we will work with the matrix

$$\overline{A} = \begin{bmatrix} [1 + x_1] & \frac{1}{2}[1 + y_1] \\ \\ -2x_1 & -y_1\end{bmatrix}.$$

We have

$$\begin{aligned}det(\overline{A} - \overline{\eta} I) &= -(1 + x_1 - \overline{\eta})(y_1 + \overline{\eta}) + x_1(1 + y_1)\\ &= (\overline{\eta} - (x_1 - y_1))(\overline{\eta} - 1),\end{aligned}$$

from which it follows that

$$\begin{aligned}\overline{\eta}^{(1)} &= x_1 - y_1 \\ \overline{\eta}^{(2)} &= 1.\end{aligned}$$

As a consequence, recalling (B.12), the eigenvalues $\tilde{\eta}^{(1)}, \tilde{\eta}^{(2)} \in \mathbb{R}$ are

$$\begin{cases}\tilde{\eta}^{(1)} & = \eta_a(\beta) \cdot [x_1 - y_1] = \eta_a(\beta)\, C_1\, \frac{2\beta(\beta-a)}{(2\beta-a)(3\beta-2a)} \\ \\ \tilde{\eta}^{(2)} & = \eta_a(\beta) \cdot 1 = \eta_a(\beta) = \kappa\frac{\beta}{\beta-a}.\end{cases} \tag{B.13}$$

In order to have $\tilde{\eta}^{(1)} < \tilde{\eta}^{(2)} < 1$, we have to impose the last condition for the non-explosion of the process in (4.4), that is $\kappa\frac{\beta}{\beta-a} < 1$. Then, the conditions for the non-explosion remain the same as for the classical ETAS, in fact we recall that $\eta = \eta_a(\beta) = \kappa\frac{\beta}{\beta-a}$.

The respective row eigenvectors (α_1, α_2) can be derived from the system

$$\begin{cases}\overline{\eta}\alpha_1 = [1 + x_1]\alpha_1 + \frac{1}{2}[1 + y_1]\alpha_2 \\ \\ \overline{\eta}\alpha_2 = -2x_1\alpha_1 - y_1\alpha_2.\end{cases}$$

In particular, for the eigenvalue $\tilde{\eta} = \eta_a(\beta)(x_1 - y_1)$, that is $\overline{\eta} = (x_1 - y_1)$, we have

$$\begin{cases}(x_1 - y_1)\alpha_1 = (1 + x_1)\alpha_1 + \frac{1}{2}(1 + y_1)\alpha_2 \\ \\ (x_1 - y_1)\alpha_2 = -2x_1\alpha_1 - y_1\alpha_2,\end{cases}$$

from which

$$\alpha_2 = -2\alpha_1.$$

Instead, for the eigenvalue $\tilde{\eta} = \eta_a(\beta)$, that is $\overline{\eta} = 1$, we have

$$\begin{cases}\alpha_1 = (1 + x_1)\alpha_1 + \frac{1}{2}(1 + y_1)\alpha_2 \\ \\ \alpha_2 = -2x_1\alpha_1 - y_1\alpha_2,\end{cases}$$

from which

$$\alpha_2 = -2\frac{x_1}{1 + y_1}\alpha_1,$$

where we recall that x_1 and y_1 are defined in (B.12).

Summarizing, the eigenfunctions are proportional to

$$\begin{cases}\ell_1^{(1)}(m') & \propto \varrho_a(m') - 2\varrho_{2a-\beta}(m') \\ \\ \ell_1^{(2)}(m') & \propto \varrho_a(m') - 2\frac{C_1 a(3\beta-a)}{(2\beta-a)[3\beta-2a+C_1(2a-\beta)]}\varrho_{2a-\beta}(m').\end{cases}$$

Exactly as before, one can proceed starting from equation (B.2) and then considering the operator $\mathcal{K}_2$ defined in (B.4), i.e.,

$$\begin{aligned}\mathcal{K}_2\ell(m) &= \int_{m_0}^{\infty} \varrho(m')p(m|m')\ell(m')dm' \\ &= \int_{m_0}^{\infty} \kappa e^{a(m'-m_0)}\beta e^{-\beta(m-m_0)}\Big[1 + C_1\big(1-2e^{-(\beta-a)(m'-m_0)}\big) \\ &\quad \cdot \big(1-2e^{-\beta(m-m_0)}\big)\Big]\ell(m')dm'.\end{aligned}$$

In this case one obtains that the integrals (B.5) and (B.6) are

$$\begin{aligned}\mathcal{I}(\ell) &= \int_{m_0}^{\infty} e^{a(m'-m_0)}\,\ell(m')dm', \\ \widetilde{\mathcal{I}}(\ell) &= \int_{m_0}^{\infty} e^{-(\beta-2a)(m'-m_0)}\,\ell(m')dm'\end{aligned}$$

and that

$$\begin{aligned}\mathcal{K}_2\ell(m) =& \kappa\beta e^{-\beta(m-m_0)}\Big[\mathcal{I}(\ell) + C_1\big[\mathcal{I}(\ell) - 2\,\widetilde{\mathcal{I}}(\ell)\big]\Big] \\ &- \kappa 2\beta e^{-2\beta(m-m_0)}C_1\big[\mathcal{I}(\ell) - 2\widetilde{\mathcal{I}}(\ell)\big].\end{aligned}$$

Then, the eigenfunctions are necessarily of the form

$$\ell_2(m) = \alpha_1\beta\, e^{-\beta(m-m_0)} + \alpha_2 2\beta\, e^{-2\beta(m-m_0)} = \alpha_1 p_\beta(m) + \alpha_2 p_{2\beta}(m), \quad \text{(B.14)}$$

where $p_\vartheta(m) = \vartheta\, e^{-\vartheta(m-m_0)}$ is a Gutenberg-Richter density of decay parameter ϑ. Taking

$$\ell(m) = p_\vartheta(m) = \vartheta\, e^{-\vartheta(m-m_0)}, \qquad \text{with } \vartheta > a \text{ (and therefore } \beta+\vartheta > 2a),$$

the integrals in (B.9) and (B.10) become respectively

$$\begin{aligned}\mathcal{J}(\ell) &= \int_{m_0}^{\infty} \vartheta e^{-(\vartheta-a)(m'-m_0)}\,dm' = \frac{\vartheta}{\vartheta-a}, \\ \widetilde{\mathcal{J}}(\ell) &= \int_{m_0}^{\infty} \vartheta e^{-(\beta+\vartheta-2a)(m'-m_0)}\,dm' = \frac{\vartheta}{\beta+\vartheta-2a}\end{aligned}$$

and we obtain that

$$\begin{aligned}\mathcal{K}_2 p_\vartheta(m) =& \kappa\left[\frac{\vartheta}{\vartheta-a} + C_1\frac{\vartheta(\beta-\vartheta)}{(\vartheta-a)(\beta+\vartheta-2a)}\right]p_\beta(m) \\ &- \kappa\, C_1\,\frac{\vartheta(\beta-\vartheta)}{(\vartheta-a)(\beta+\vartheta-2a)}\,p_{2\beta}(m).\end{aligned} \quad \text{(B.15)}$$

Let's notice that, since ϑ is equal to β or 2β, the condition $\vartheta > a$ reduces to $\beta > a$.

Now, by similar computations as for $\mathcal{K}_1$, one can find that the eigenvalues are

$$\begin{cases} \tilde{\eta}^{(1)} & = C_1\, \eta_a(\beta)\frac{2(\beta-a)}{2\beta-a}\,\frac{\beta}{3\beta-2a} \\ \\ \tilde{\eta}^{(2)} & = \eta_a(\beta) = \kappa\frac{\beta}{\beta-a}. \end{cases} \tag{B.16}$$

We observe that the above values are exactly the same as for the operator $\mathcal{K}_2$, as expected. We have already explained for the latter operator that, in order to have $\tilde{\eta}^{(1)} < \tilde{\eta}^{(2)} < 1$, we have to impose the third condition for the non-explosion in (4.4). We recall again that $\eta = \eta_a(\beta) = \frac{\kappa\beta}{\beta-a}$.

Finally, the eigen(density)functions in the case of equation (B.2) are proportional to

$$\begin{cases} \ell_2^{(1)}(m) & \propto p_\beta(m) - [1 - C_1\,\frac{\beta}{3\beta-2a}]\eta_a(\beta)\frac{2(\beta-a)}{2\beta-a}p_{2\beta}(m) \\ \\ & = p_\beta(m) - [1 - C_1\,\frac{\beta}{3\beta-2a}]\kappa\,\frac{2\beta}{2\beta-a}\,p_{2\beta}(m) \\ \\ \ell_2^{(2)}(m) & \propto p_\beta(m). \end{cases} \tag{B.17}$$

B.2 Finiteness of the mean cluster size

In the case of the operator $\mathcal{K}_1$, one can analyze the eigenfunctions coefficients when the process evolves, as in [Zhuang, 2002]. More precisely, starting from the function $\varrho_a(\cdot)$, we can consider the evolution of the equation

$$\ell(m) = \vartheta_1\varrho_a(m) + \vartheta_2\varrho_{2a-\beta}(m)$$

after n steps, i.e.,

$$\mathcal{K}_1^{(n)}\ell(m) = \vartheta_1^n\varrho_a(m) + \vartheta_2^n\varrho_{2a-\beta}(m)$$

and then compute $(\vartheta_1^n, \vartheta_2^n)$. To do that, we observe that we can write the expected number of triggered events per triggering shock in terms of the operator $\mathcal{K}_1$ as follows. Using the notations as in Chapter 4, we can write that the expected value of the number $N_i^{tr}(\mathbb{R}^2)$ of all the triggered events generated by the spontaneous shock $(S_i = s, M_i = m)$, with $N_i^{tr,n}(\mathbb{R}^2)$ indicating its n^{th} generation, is

$$\mathbb{E}\left[N_i^{tr}(\mathbb{R}^2)\Big|S_i = s, M_i = m\right] = \mathbb{E}\left[\sum_{n=1}^{\infty} N_i^{tr,n}(\mathbb{R}^2)\Big|S_i = s, M_i = m\right]. \tag{B.18}$$

Now, for $\mathbf{1}(m) = \mathbb{1}_{[m_0,\infty)}(m)$, it holds

$$\begin{aligned}\mathbb{E}\Big[N_i^{tr,1}(\mathbb{R}^2)\Big|S_i = s, M_i = m\Big] &= \int_s^\infty dt \int_{m_0}^\infty \Phi(t-s)\varrho(m)p(m'|m)dm' \\ &= \int_{m_0}^\infty \varrho(m)p(m'|m)dm' \\ &= \varrho(m) \\ &= \mathcal{K}_1\mathbf{1}(m),\end{aligned}$$

$$\begin{aligned}&\mathbb{E}\Big[N_i^{tr,2}(\mathbb{R}^2)\Big|S_i = s, M_i = m\Big] = \int_s^\infty dt_1 \int_{m_0}^\infty dm_1 \int_{t_1}^\infty dt_2 \int_{m_0}^\infty \Phi(t_1-s) \\ &\quad \cdot \Phi(t_2-t_1)\varrho(m)p(m_1|m)\varrho(m_1)p(m_2|m_1)dm_2 \\ &= \int_{m_0}^\infty dm_1 \varrho(m)p(m_1|m)\int_{m_0}^\infty \varrho(m_1)p(m_2|m_1)dm_2 \\ &= \int_{m_0}^\infty dm_1 \varrho(m)p(m_1|m)\mathcal{K}_1\mathbf{1}(m_1) \\ &= \int_{m_0}^\infty dm_1 \varrho(m)p(m_1|m)\varrho(m_1) \\ &=\mathcal{K}_1\left(\mathcal{K}_1\mathbf{1}\right)(m) \\ &=\mathcal{K}_1^{(2)}\mathbf{1}(m)\end{aligned}$$

and so on. Then,

$$\mathbb{E}\Big[N_i^{tr,n}(\mathbb{R}^2)\Big|S_i = s, M_i = m\Big] = \mathcal{K}_1^{(n)}\mathbf{1}(m) = \mathcal{K}_1^{(n-1)}\varrho(m).$$

Given the matrix U such that

$$\overline{A} = UDU^{-1}$$

with

$$D = \begin{bmatrix} (x_1-y_1) & 0 \\ 0 & 1 \end{bmatrix},$$

where x_1 and y_1 are defined above in (B.12), we consequently have that

$$\mathcal{K}_1^{(n-1)}\varrho(m)\ell(m) = \eta^{n-1}[UD^{n-1}U^{-1}]\begin{pmatrix}\vartheta_1 \\ \vartheta_2\end{pmatrix} = \begin{pmatrix}\vartheta_1^n \\ \vartheta_2^n\end{pmatrix}.$$

We recall that η is the branching ratio and is equal to $\eta_a(\beta) = \frac{\kappa\beta}{\beta-a}$.

To the aim of finding explicitly

$$\begin{pmatrix} \vartheta_1^n \\ \\ \vartheta_2^n \end{pmatrix},$$

let's start saying that U is the matrix whose columns are the eigenvectors of the two eigenvalues $(\overline{\eta}^{(1)}, \overline{\eta}^{(2)}) = (x_1 - y_1, 1)$ relative to $\overline{A}$, i.e.,

$$U = \begin{bmatrix} 1 & 1 \\ \\ -2 & -\frac{2x_1}{1+y_1} \end{bmatrix}$$

and then

$$U^{-1} = \frac{1}{1+y_1-x_1} \begin{bmatrix} -x_1 & -\frac{1+y_1}{2} \\ \\ 1+y_1 & \frac{1+y_1}{2} \end{bmatrix}.$$

As a consequence, setting

$$z_1 = x_1 - y_1 = C_1 \frac{2\beta(\beta - a)}{(2\beta - a)(3\beta - 2a))},$$

one obtains

$$UD^{n-1}U^{-1} = \frac{1}{1-z_1} \begin{bmatrix} 1+y_1-x_1 z_1^{n-1} & -\frac{1+y_1}{2}(1-z_1^{n-1}) \\ \\ -2x_1(1-z_1^{n-1}) & (1+y_1)z_1^{n-1} - x_1 \end{bmatrix}.$$

We observe that $|z_1\eta| = z_1\eta = \tilde{\eta}^{(1)} < 1$.

Finally, since ϱ_a corresponds to the vector

$$\begin{pmatrix} \vartheta_1^1 \\ \\ \vartheta_2^1 \end{pmatrix} = \begin{pmatrix} 1 \\ \\ 0 \end{pmatrix}.$$

we obtain that

$$\begin{pmatrix} \vartheta_1^n \\ \\ \vartheta_2^n \end{pmatrix} = \eta^{n-1}(UD^{n-1}U^{-1}) \begin{pmatrix} 1 \\ \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1+y_1-x_1 z_1^{n-1}}{1-z_1}\eta^{n-1} \\ \\ \frac{-2x_1(1-z_1^{n-1})}{1-z_1}\eta^{n-1} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1+y_1}{1-z_1}\left[\tilde{\eta}^{(2)}\right]^{n-1} - \frac{x_1}{1-z_1}\left[\tilde{\eta}^{(1)}\right]^{n-1} \\ \\ \frac{-2x_1}{1-z_1}\left[\tilde{\eta}^{(2)}\right]^{n-1} + \frac{2x_1}{1-z_1}\left[\tilde{\eta}^{(1)}\right]^{n-1} \end{pmatrix}.$$

Now, due to the fact that both the eigenvalues $\tilde{\eta}^{(1)}, \tilde{\eta}^{(2)}$ are less than one, we can conclude that the series in (B.18) is finite, i.e., setting

$$\gamma_1 = \sum_{n=1}^{\infty}\left(\frac{1+y_1-x_1 z_1^{n-1}}{1-z_1}\eta^{n-1}\right) = \sum_{n=1}^{\infty}\left(\frac{1+y_1}{1-z_1}\left[\tilde{\eta}^{(2)}\right]^{n-1} - \frac{x_1}{1-z_1}\left[\tilde{\eta}^{(1)}\right]^{n-1}\right), \tag{B.19}$$

$$\gamma_2 = \sum_{n=1}^{\infty}\left(\frac{-2x_1(1-z_1^{n-1})}{1-z_1}\eta^{n-1}\right) = \sum_{n=1}^{\infty}\left(\frac{-2x_1}{1-z_1}\left[\tilde{\eta}^{(2)}\right]^{n-1} + \frac{2x_1}{1-z_1}\left[\tilde{\eta}^{(1)}\right]^{n-1}\right), \tag{B.20}$$

it holds

$$\begin{aligned}\mathbb{E}\Big[\sum_{n=1}^{\infty} N_i^{tr,n}(\mathbb{R}^2)\Big|S_i = s, M_i = m\Big] &= \sum_{n=1}^{\infty}\left[\varrho_a(m)\vartheta_1^n + \varrho_{2a-\beta}(m)\vartheta_2^n\right] \\ &= \varrho_a(m)\sum_{n=1}^{\infty}\vartheta_1^n + \varrho_{2a-\beta}(m)\sum_{n=1}^{\infty}\vartheta_2^n \\ &= \varrho_a(m)\gamma_1 + \varrho_{2a-\beta}(m)\gamma_2 < \infty.\end{aligned}$$

Consequently,

$$\begin{aligned}&\mathbb{E}\Big[N_i^{tr}(\mathbb{R}^2)\Big|S_i = s\Big] = \int_{m_0}^{\infty}\left[\gamma_1\varrho_a(m) + \gamma_2\varrho_{2a-\beta}(m)\right]p_\beta(m)dm \\ &= \gamma_1\int_{m_0}^{\infty}\varrho_a(m)p_\beta(m)dm + \gamma_2\int_{m_0}^{\infty}\varrho_{2a-\beta}(m)p_\beta(m)dm \\ &= \eta\left[\gamma_1 + \frac{\gamma_2}{2}\right] \\ &= \sum_{n=1}^{\infty}\eta^n\left[\frac{1+y_1-x_1 z_1^{n-1}}{1-z_1} - \frac{x_1(1-z_1^{n-1})}{1-z_1}\right] \\ &= \sum_{n=0}^{\infty}\eta^n - 1 \\ &= \frac{\eta}{1-\eta} < \infty,\end{aligned}$$

where the last three equalities follows from (B.19), (B.20) and the fact that $z_1 = x_1 - y_1$.

Recalling Remark 3 in Chapter 1, we can then conclude that the resulting process does not explode, in fact the mean cluster size is finite.

Appendix C

Time derivatives of the probability P$\{\tau\}$ of having zero observable events in $[0,\tau]$

In this appendix we include the calculations we made to obtain the second time derivative of the probability $\mathbf{P}\{\tau\} = e^{-\varpi L(0;\tau)}$ in the case $m_c \geq m_0$. This derivative is indeed useful in order to get the interevent time density $F_{inter}(\tau)$ by means of equation (6.24). We notice that in this appendix we are heuristically assuming that, if

$$\hat{L}(0;\tau) = \left[\tau\widetilde{Q} + \eta[H-L]A(\tau)\right]\left[\frac{1-\eta+\eta a(\tau)}{(1-\eta)[1-\delta+\delta a(\tau)]}\right],$$

that is the approximation obtained for $L(0;\tau)$ in the case $m_c \geq m_0$ when considering the further approximation for small τ, it holds also

$$\frac{d^k L(0;\tau)}{d\tau^k} \approx \frac{d^k \hat{L}(0;\tau)}{d\tau^k},$$

for $k = 1, 2$.

Let's then proceed with the computations. Using equations (6.25) and (6.79) we have

$$\begin{aligned}
\frac{d}{d\tau}\left(-\frac{\ln \mathbf{P}\{\tau\}}{\varpi}\right) =& \frac{d}{d\tau}L(0;\tau) \\
\approx& \left[\widetilde{Q} + \eta[H-L]a(\tau)\right]\left[\frac{1-\eta+\eta a(\tau)}{(1-\eta)[1-\delta+\delta a(\tau)]}\right] \\
&+ \left[\tau\widetilde{Q} + \eta[H-L]A(\tau)\right]\frac{d}{d\tau}\left[\frac{1-\eta+\eta a(\tau)}{(1-\eta)[1-\delta+\delta a(\tau)]}\right].
\end{aligned}$$

It follows that

$$\frac{d\mathbf{P}\{\tau\}}{d\tau} \approx -\, e^{-\varpi L(0;\tau)} \Bigg\{ \left[\widetilde{Q} + \varpi\eta[H-L]a(\tau)\right] \left[\frac{1-\eta+\eta a(\tau)}{(1-\eta)[1-\delta+\delta a(\tau)]}\right] + \left[\tau\widetilde{Q} + \varpi\eta[H-L]A(\tau)\right] \frac{d}{d\tau}\left[\frac{1-\eta+\eta a(\tau)}{(1-\eta)[1-\delta+\delta a(\tau)]}\right] \Bigg\}.$$

The second derivative of $\mathbf{P}\{\tau\}$ with respect to τ is then

$$\frac{d^2\mathbf{P}\{\tau\}}{d\tau^2} = \frac{d^2 e^{-\varpi L(0;\tau)}}{d\tau^2} \tag{C.1}$$

$$\begin{aligned}
&\approx e^{-\varpi L(0;\tau)} \frac{d}{d\tau}\varpi L(0;\tau) \Bigg\{ \left[\widetilde{Q} + \varpi\eta[H-L]a(\tau)\right] \left[\frac{1-\eta+\eta a(\tau)}{(1-\eta)[1-\delta+\delta a(\tau)]}\right] \\
&\quad + \left[\tau\widetilde{Q} + \varpi\eta[H-L]A(\tau)\right] \frac{d}{d\tau}\left[\frac{1-\eta+\eta a(\tau)}{(1-\eta)[1-\delta+\delta a(\tau)]}\right] \Bigg\} \\
&\quad - e^{-\varpi L(0;\tau)} \Bigg\{ -\varpi\eta[H-L]\Phi(\tau) \left[\frac{1-\eta+\eta a(\tau)}{(1-\eta)[1-\delta+\delta a(\tau)]}\right] \\
&\quad + \left[\widetilde{Q} + \varpi\eta[H-L]a(\tau)\right] \frac{d}{d\tau}\left[\frac{1-\eta+\eta a(\tau)}{(1-\eta)[1-\delta+\delta a(\tau)]}\right] \\
&\quad + \left[\widetilde{Q} + \varpi\eta[H-L]a(\tau)\right] \frac{d}{d\tau}\left[\frac{1-\eta+\eta a(\tau)}{(1-\eta)[1-\delta+\delta a(\tau)]}\right] \\
&\quad + \left[\tau\widetilde{Q} + \varpi\eta[H-L]A(\tau)\right] \frac{d^2}{d\tau^2}\left[\frac{1-\eta+\eta a(\tau)}{(1-\eta)[1-\delta+\delta a(\tau)]}\right] \Bigg\} \\
&= e^{-\varpi L(0;\tau)} \Bigg\{ \Bigg[\left(\widetilde{Q} + \varpi\eta[H-L]a(\tau)\right) \left(\frac{1-\eta+\eta a(\tau)}{(1-\eta)[1-\delta+\delta a(\tau)]}\right) \\
&\quad + \left(\tau\widetilde{Q} + \varpi\eta[H-L]A(\tau)\right) \left(\frac{\Phi(\tau)(\delta-\eta)}{(1-\eta)[1-\delta+\delta a(\tau)]^2}\right) \Bigg]^2 \\
&\quad - \Bigg[-\varpi\eta[H-L]\Phi(\tau) \left(\frac{1-\eta+\eta a(\tau)}{(1-\eta)[1-\delta+\delta a(\tau)]}\right) \\
&\quad + 2\left(\widetilde{Q} + \varpi\eta[H-L]a(\tau)\right) \left(\frac{\Phi(\tau)(\delta-\eta)}{(1-\eta)[1-\delta+\delta a(\tau)]^2}\right) \\
&\quad + \left(\tau\widetilde{Q} + \varpi\eta[H-L]A(\tau)\right) \frac{(\delta-\eta)\Phi(\tau)[(\theta+1)(\delta-1)+\delta a(\tau)(\theta-1)]}{(1-\eta)(\tau+c)[1-\delta+\delta a(\tau)]^3} \Bigg] \Bigg\}.
\end{aligned} \tag{C.2}$$

The previous formula has been obtained using the following results:

$$\begin{aligned}\frac{da(\tau)}{d\tau} &= \frac{d}{d\tau}\left[1-\int_0^\tau \Phi(x)dx\right]\\ &= -\Phi(\tau);\end{aligned}$$

$$\begin{aligned}\frac{d}{d\tau}\left[\frac{1-\eta+\eta a(\tau)}{(1-\eta)[1-\delta+\delta a(\tau)]}\right] =& \frac{d}{d\tau}\left[\frac{\delta-\eta+\eta[1-\delta+\delta a(\tau)]}{\delta(1-\eta)[1-\delta+\delta a(\tau)]}\right]\\ =& \frac{\delta-\eta}{\delta(1-\eta)}\frac{d}{d\tau}\left[\frac{1}{1-\delta+\delta a(\tau)}\right]\\ =& \frac{\Phi(\tau)(\delta-\eta)}{(1-\eta)[1-\delta+\delta a(\tau)]^2};\end{aligned}$$

in the end

$$\begin{aligned}&\frac{d^2}{d\tau^2}\left[\frac{1-\eta+\eta a(\tau)}{(1-\eta)[1-\delta+\delta a(\tau)]}\right] = \frac{\delta-\eta}{1-\eta}\frac{d}{d\tau}\left[\frac{\Phi(\tau)}{[1-\delta+\delta a(\tau)]^2}\right]\\ =&\frac{\delta-\eta}{(1-\eta)[1-\delta+\delta a(\tau)]^4}\left\{-\Phi(\tau)\frac{\theta+1}{\tau+c}[1-\delta+\delta a(\tau)]^2\right.\\ &\left.+2\Phi(\tau)[1-\delta+\delta a(\tau)]\delta\Phi(\tau)\right\}\\ =&\frac{(\delta-\eta)\Phi(\tau)}{(1-\eta)[1-\delta+\delta a(\tau)]^3}\left\{\frac{\theta+1}{\tau+c}[-1+\delta-\delta a(\tau)]+2\delta\Phi(\tau)\right\}\\ =&\frac{(\delta-\eta)\Phi(\tau)}{(1-\eta)(\tau+c)[1-\delta+\delta a(\tau)]^3}\left[(\theta+1)(\delta-1)-\delta\theta a(\tau)-\delta a(\tau)+2\delta\theta a(\tau)\right]\\ =&\frac{(\delta-\eta)\Phi(\tau)[(\theta+1)(\delta-1)+\delta a(\tau)(\theta-1)]}{(1-\eta)(\tau+c)[1-\delta+\delta a(\tau)]^3}.\end{aligned}$$

The last computation has been obtained since, recalling equation (6.3) and (6.15), we have

$$\begin{aligned}\frac{d\Phi(\tau)}{d\tau} &= \theta c^\theta\frac{d}{d\tau}(\tau+c)^{-\theta-1}\\ &= -\theta c^\theta(\theta+1)(\tau+c)^{-\theta-2}\\ &= -\Phi(\tau)\frac{\theta+1}{\tau+c}\end{aligned}$$

and

$$\Phi(\tau) = \frac{\theta a(\tau)}{\tau+c}.$$

Appendix D

Algorithms

The codes of the programs of the first and the second kinds of analysis, the choice of the time amplitude δ^* and the new simulation according to the hypothesis of correlated magnitudes are available at `https://www.dropbox.com/s/cazsfi7kbf196bl/Algorithms.pdf?dl=0`. We recall that the first three codes have been written in MATLAB language; the latter has instead been written in FORTRAN language.

Bibliography

H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, December 1974.

P. Bak and K. Chen. Self-organized criticality. *Scientific American (United States)*, 264(1), January 1991.

P. Bak and C. Tang. Earthquakes as a self-organized critical phenomenon. *Journal of Geophysical Research*, 94(B11), November 1989.

P. Bak, K. Christensen, L. Danon, and T. Scanlon. Unified scaling law for earthquakes. *Physical Review Letters*, 88(17):1–4, April 2002.

H. Bateman. *Higher transcendental functions.* Mc Graw-Hill Book Company, Inc., 1953.

M. N. Berberan-Santos. Analytical inversion of the Laplace transform without contour integration: application to luminescence decay laws and other relaxation functions. *Journal of Mathematical Chemistry*, 38(2):165–173, August 2005.

P. Bormann. *New manual of seismological observatory practice (NMSOP), Chapter 4.* Deutsches GeoForschungsZentrum GFZ, p. 1-34, 2009.

L. Brémaud, P. Massoulié. Stability of nonlinear Hawkes processes. *The Annals of Probability*, 24(3):1563–1588, July 1996.

P. Brémaud. *A course on point processes.* 2013. Unpublished manuscript. Lectures given at the Scuola Matematica Interuniversitaria of the Scuola Normale Superiore of Pisa.

P. Brémaud, G. Nappo, and G. L. Torrisi. Rate of convergence to equilibrium of marked Hawkes processes. *Journal of Applied Probability*, 39(1):123–136, March 2002.

N. R. Campbell. The study of discontinuous phenomena. *Mathematical Proceedings of the Cambridge Philosophical Society*, 15:117–136, 1909.

M. Caputo and G. Sebastiani. Time and space analysis of two earthquakes in the Apennines (Italy). *Natural science*, 3(9):768–774, September 2011.

I. T. Cezar. *The mechanism of induced seismicity*. Birkhuser, 2002.

A. Corral. Local distributions and rate fluctuations in a unified scaling law for earthquakes. *Physical Review E*, 68(3):035102/1–4, September 2003.

A. Corral. Long-term clustering, scaling, and universality in the temporal occurrence of earthquakes. *Physical Review Letters*, 92(10):108501/1–4, March 2004.

A. Corral. Dependence of earthquake recurrence times and independence of magnitudes on seismicity history. *Tectonophysics*, 424(3-4):177193, 2006.

D. R. Cox and V. Isham. *Point processes*. Chapman & Hall/CRC, Monographs on Statistics & Applied Probability, 2000.

D. J. Daley and D. Vere-Jones. *An introduction to the theory of the point processes, Volume I: elementary theory and methods, 2nd Edition*. Springer, New York, 2003.

D. J. Daley and D. Vere-Jones. *An introduction to the theory of the point processes, Volume II: general theory and structure, 2nd Edition*. Springer, New York, 2008.

J. Davidsen and C. Goltz. Are seismic waiting time distributions universal? *Geophysical Research Letters*, 31(21), November 2004.

C.M.R. Fowler. *The Solid Earth: An Introduction to Global Geophysics*. Cambridge University Press, 1990.

J. D. Gibbons and S. Chakraborti. *Nonparametric statistical inference*. Marcel Dekker, Inc., New York, 2003.

G. Grimmett and D. Stirzaker. *Probability and random processes - Third edition*. Oxford University Press, New York, 2001.

B. Gutenberg and C. F. Richter. Frequency of earthquakes in California. *Bulletin of the Seismological Society of America*, 34(8):185–188, 1944.

E. Hauksson, W. Yang, and P. M. Shearer. Waveform relocated earthquake catalog for Southern California (1981 to 2011). *Bulletin of the Seismological Society of America*, 102(5):2239–2244, October 2012.

A. G. Hawkes. Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 33(3): 438–443, 1971a.

A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, April 1971b.

A. G. Hawkes and D. Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11(3):493–503, September 1974.

A. Helmstetter and D. Sornette. Subcritical and supercritical regimes in epidemic models of earthquake aftershocks. *Journal of Geophysical Research*, 107(B10, month = October), 2002a.

A. Helmstetter and D. Sornette. Subcritical and supercritical regimes in epidemic models of earthquake aftershocks. *Journal of Geophysical Research*, 107(B10), October 2002b.

A. Helmstetter, Y. Y. Kagan, and D. D. Jackson. Comparison of short-term and time-independent earthquake forecast models for Southern California. *Bulletin of the Seismological Society of America*, 96(1):90–106, February 2006.

F. Kishinouye and H. Kawasumi. An application of the theory of fluctuation to problems in statistical seismology. *Jishin Kenkyusho Ihoo*, 4:75–83, 1928.

P. A. W. Lewis and G. S. Shedler. Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics*, 26:403–413, 1979.

G. Lin, P. M. Shearer, and E. Hauksson. Applying a three-dimensional velocity model, waveform cross correlation, and cluster analysis to locate Southern California seismicity from 1981 to 2005. *Journal of Geophysical Research*, 112(B12), December 2007.

E. Lippiello, C. Godano, and L. de Arcangelis. Dynamical scaling in branching models for seismicity. *Physical Review Letters*, 98(9):098501/1–4, February 2007a.

E. Lippiello, M. Bottiglieri, C. Godano, and L. de Arcangelis. Dynamical scaling and generalized Omori law. *Geophysical Research Letters*, 34(23), 2007b.

E. Lippiello, L. de Arcangelis, and C. Godano. Influence of time and space correlations on earthquake magnitude. *Physical Review Letters*, 100(3): 038501/1–4, January 2008.

E. Lippiello, C. Godano, and L. de Arcangelis. The earthquake magnitude is influenced by previous seismicity. *Geophysical Research Letters*, 39(5), 2012.

R. S. Liptser and A. N. Shiryaev. *Statistics of random processes*. Springer, New York, 1978.

A. M. Lombardi. Estimation of the parameters of ETAS models by simulated annealing. *Scientific Reports*, 5, February 2015.

A. M. Lombardi, M. Cocco, and W. Marzocchi. On the increase of background seismicity rate during the 1997 - 1998 Umbria-Marche, central Italy, sequence. apparent variation or fluid-driven triggering? *Bulletin of the Seismological Society of America*, 100(3):11381152, June 2010.

W. Marzocchi and A. M. Lombardi. Real-time forecasting following a damaging earthquake. *Geophysical Research Letters*, 36(L21302), November 2009.

W. Marzocchi and L. Sandri. A review and a new insights on the estimation of the b-value and its uncertainty. *Annals of Geophysics*, 46(6):1271–1282, December 2003.

G. Molchan. Interevent time distribution in seismicity: a theoretical approach. *Pure and Applied Geophysics*, 162(6-7):1135–1150, June 2005.

O. Molchanov. About climate-seismicity coupling from correlation analysis. *Natural Hazards and Earth System Sciences*, 10(2):299–304, February 2010.

J. Møller and J. G. Rasmussen. Perfect simulation of Hawkes processes. *Advances in Applied Probability*, 37(3):629–646, June 2005.

J. Møller and J. G. Rasmussen. Approximate simulation of Hawkes processes. *Methodology and Computing in Applied Probability*, 8(1):53–64, May 2006.

F. Musmeci and D. Vere-Jones. A space-time clustering model for historical earthquakes. *Annals of the Institute of Statistical Mathematics*, 44(1):1–11, March 1992.

Y. Ogata. On Lewis simulation method for point processes. *IEEE Transactions on Information Theory*, IT-27(1):23–31, January 1981.

Y. Ogata. Likelihood analysis of point processes and its application to seismological data. *Bulletin of the International Statistical Institute*, 50(2): 943–961, 1983.

Y. Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83 (401):9–27, March 1988.

Y. Ogata. Statistical model for standard seismicity and detection of anomalies by residual analysis. *Tectonophysics*, 169(1-3):159–174, 1989.

Y. Ogata. Space-time point process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2):379–402, 1998.

Y. Ogata. Seismicity analysis through point-process modeling: a review. *Pure and Applied Geophysics*, 155(2-4):471–507, 1999.

Y. Ogata. Statistical analysis of seismicity: updated version (SASeis2006). *Computer Science Monographs*, 33, April 2006.

Y. Ogata and D. Vere-Jones. Inference for earthquake models: a self-correcting model. *Stochastic Processes and their Applications*, 17(2): 337347, July 1984.

S. Ohmi and K. Hirahara. Temporal variation of the decay rate of the autocorrelation function of the ambient seismic noise. *American Geophysical Union Fall Meeting Abstracts*, page A1701, December 2009.

F. Omori. On aftershocks of earthquakes. *Journal of the College of Science, Imperial University of Tokyo*, 7:111–200, 1894.

G. \`Ozel and C. İnal. The probability function of the compound Poisson process and an application to aftershock sequence in Turkey. *Environmetrics*, 19(1):79–85, February 2008.

F. Papangelou. Integrability of expected increments of point processes and a related random change of scale. *Transactions of the American Mathematical Society*, 165:483–506, 1972.

E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, August 1962.

M. B. Priestley. *Spectral analysis and time series, Volume I.* Academic Press Inc., London, New York, 1982.

M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832, 1956.

F. Roueff, R. von Sachs, and L. Sansonnet. Time-frequency analysis of locally stationary Hawkes processes. *ISBA - Institut de Statistique, Biostatistique et Sciences Actuarielles - Discussion Paper*, pages 1–32, 2015. Available at http://hdl.handle.net/2078.1/160933.

A. Saichev and D. Sornette. Anomalous power law distribution of total lifetimes of branching processes: application to earthquake aftershock sequences. *Physical Review E*, 70(4):046123/1–8, October 2004.

A. Saichev and D. Sornette. Vere-Jones' self-similar branching model. *Physical Review E*, 72(5):056122/1–13, November 2005.

A. Saichev and D. Sornette. "Universal" distribution of interearthquake times explained. *Physical Review Letters*, 97(7):078501/1–4, August 2006a.

A. Saichev and D. Sornette. Power law distribution of seismic rates: theory and data analysis. *The European Physical Journal B*, 49(3):377–401, February 2006b.

A. Saichev and D. Sornette. Theory of earthquake recurrence times. *Journal of Geophysical Research*, 112(B4), April 2007.

A. Saichev and D. Sornette. Fertility heterogeneity as a mechanism for power law distributions of recurrence times. *Physical Review E*, 87(2):022815/1–11, February 2013.

A. Saichev, A. Helmstetter, and D. Sornette. Power law distributions of offspring and generation numbers in branching models of earthquake triggering. *Pure and Applied Geophysics*, 162(6-7):1113–1134, June 2005.

N. V. Sarlis, E. S. Skordas, and P. A. Varotsos. Multiplicative cascades and seismicity in natural time. *Physical Review E*, 80(2):022102/1–4, August 2009.

N. V. Sarlis, E. S. Skordas, and P. A. Varotsos. Nonextensivity and natural time: the case of seismicity. *Physical Review E*, 82(2):021110/1–9, August 2010.

D. W. Scott. *Multivariate Density Estimation: Theory, Practice and Visualization*. John Wiley and Sons, New York, 1992.

Y. Shi and B. A. Bolt. The standard error of the magnitude-frequency b-value. *Bulletin of the Seismological Society of America*, 72(5):1677–1687, 1982.

A. Sornette and D. Sornette. Self-organized criticality and earthquakes. *Europhysics Letters*, 9(3), 1989.

A. Sornette and D. Sornette. Renormalization of earthquake aftershocks. *Geophysical Research Letters*, 26(13):1981–1984, July 1999.

D. Sornette and M. J. Werner. Apparent clustering and apparent background earthquakes biased by undetected seismicity. *Journal of Geophysical Research*, 110(B9), September 2005.

A. E. Taylor and D. Lay. *Introduction to functional analysis*. John Wiley and Sons Inc., New York, 1980.

N. M. Temme. *Special functions. An introduction to the classical functions of mathematical physics*. John Wiley and Sons Inc., New York, 1996.

T. Utsu. Magnitudes of earthquakes and occurrence of their aftershocks. *Zisin (Journal of the Seismological Society of Japan), Series 2*, 10:35–45, 1957. In Japanese with English summary.

T. Utsu. Aftershocks and earthquake statistics (ii). Further investigation of aftershocks and other earthquake sequences based on a new classification of earthquake sequences. *Journal of the Faculty of Science, Hokkaido University, Series VII*, 3:197–266, 1970.

T. Utsu and A. Seki. A relation between the area of aftershock region and the energy of mainshock. *Zisin (Journal of the Seismological Society of Japan), Series 2*, 7:233–240, 1954. In Japanese with English summary.

D. Vere-Jones. Earthquake prediction - a statistician's view. *Journal of Physics of the Earth*, 26(2):129–146, 1978.

D. Vere-Jones. A class of self-similar random measure. *Advances in Applied Probability*, 37(4):908–914, 2005.

D. Vere-Jones and Y. Ogata. On the moments of a self-correcting process. *Journal of Applied Probability*, 21(2):335342, June 1984.

S. Wiemer. A software package to analyze seismicity: ZMAP. *Seismological research letters*, 72(2):374–383, March-April 2001.

J. Zhuang. *Some Applications of Point Processes in Seismicity Modelling and Prediction*. 2002. Ph. D. thesis, the Graduate Unversity for Advanced Studies, Kanagawa, Japan.

J. Zhuang and Y. Ogata. Properties of the probability distribution associated with the largest event in an earthquake cluster and their implications to foreshocks. *Physical Review E*, 73(4):046134/1–12, April 2006.

J. Zhuang, Y. Ogata, and D. Vere-Jones. Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, 97(458):369–380, 2002.

J. Zhuang, D. Harte, M. J. Werner, S. Hainzl, and S. Zhou. Basic models of seismicity: temporal models. *Community Online Resource for Statistical Seismicity Analysis*, August 2012.