UNIVERSITÀ COMMERCIALE "LUIGI BOCCONI" – Milano
Facoltà di Economia
Dottorato di Ricerca in Statistica
Ciclo XVI

Sequential estimation methods in continuous-time state-space models

Coordinatore: Chiar.mo Prof. Pietro Muliere

Tesi di
Elisa Varini

UNIVERSITÀ COMMERCIALE "LUIGI BOCCONI"
ISTITUTO DI METODI QUANTITAVI

The thesis "**Sequential estimation methods in continuous-time state-space models**" by **Elisa Varini** is recommended for acceptance by the members of the delegated committee, as stated by the enclosed reports, in partial fulfillment of the requirements for the degree of **Dottore di Ricerca in Statistica**.

Dated: 10 January 2005

Research Supervisor: **Renata Rotondi**
Research Supervisor: **Nicolas Chopin**
External Examiner: **Eugénio Melilli**

UNIVERSITÀ COMMERCIALE "LUIGI BOCCONI"

Date: **10 January 2005**

Author :	**Elisa Varini**
Title:	**Sequential estimation methods in continuous-time state-space models**
Department:	**Istituto di Metodi Quantitativi**

Permission is herewith granted to Università Commerciale "Luigi Bocconi" to circulate and to have copied for non-commercial purposes, at its discretion, the above title upon the request of individuals or institutions.

Elisa Varini

Signature of Author

Acknowledgments

I would first of all like to thank all members from the Institute of Quantitative Methods at Bocconi University in Milan and particularly Prof. Pietro Muliere for giving me the chance to attend a fruitful PhD course.
I would also give a warm thank to my research supervisors Dr. Renata Rotondi (IMATI-CNR), who has introduced me into the world of statistics and has constantly supported me with infectious excitement and interest, and Prof. Nicolas Chopin (University of Bristol) who supervised this work with great knowledge, careful edits and encouragement.
I am very grateful to Prof. Eugenio Melilli for his punctual and patient revision of my thesis and for his useful advice. I want to reserve a special word of gratitude to Prof. Anna Gerardi (University of L'Aquila) who promptly gave me a decisive and reliable help with regard to the solution of the filtering equations.
Finally, I wish to thank the Institute of Applied Mathematics and Information Technology at the CNR (National Research Council) in Milan and the Department of Mathematics at the University of Bristol for having given me hospitality and the chance of an interesting experience of research.

Contents

Table of notations

$\mathcal{N}$	the set of the integer numbers
$\mathcal{R}$	the set of the real numbers
$\mathcal{R}^+$	the set $[0, +\infty)$
$\bar{\mathcal{R}}$	the set $[-\infty, +\infty]$
$\mathcal{B}(E)$	the Borel σ-algebra on the set $E \subseteq \bar{\mathcal{R}}$
$(E, \mathcal{E})$	a measurable space
$card(E)$	the cardinality of the set E
$(\Omega, \mathcal{F}, P)$	a probability space
a.s.	almost surely
$\xrightarrow{d}$	the convergence in distribution
$I_a(y) = I(y = a) = \begin{cases} 1 & \text{if } y = a \\ 0 & \text{otherwise} \end{cases}$	the indicator function
$[x]^+ = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$	
SRM	Stress Release Model
ETAS	Epidemic-Type Aftershock-Sequence

Introduction

This thesis represents an important stage of my educational and research path, begun in 2000 when I had the chance to collaborate in a research project conducted by my supervisor Dr. Renata Rotondi.
My initial contribution was essentially mathematical and consisted in the implementation of Fortran software programs for the analysis, based on the classical maximum likelihood methods, of statistical models for seismic sequences.
The opportunity to extend my knowledge in statistics during my PhD course allowed me to better understand and carry on the research.
First, led by my supervisor Rotondi, I explored the literature of the so called *statistical seismology*: this expression, coined by the Japanese seismologist Keiiti Aki (1930-), points out a fusion of physical and statistical ideas to produce models that can be fitted to catalogues and other types of seismic data.
I focused on the wide literature concerning the point process models, of which the statistician and mathematician David Vere-Jones is one of the main scholars. In the 1970s, Vere-Jones detected the point processes among the most suitable stochastic models to describe a sequence of earthquakes.
Chapter 1 of this thesis presents three important families of point processes for seismic sequences, each of them having a physical foundation: the Poisson process postulates that the rate of earthquake occurrence is constant; the stress release model is based on the elastic rebound theory, formulated by the famous seismologist Harry Fielding Reid at the beginning of the 20^{th} century; while the Epidemic-Type Aftershock-Sequence model is conceived to model clusters of shocks that typically occur after strong earthquakes.
I described and, togheter with my supervisor, studied these models and further their versions in order to familiarize myself with them. This work led us to devise the subject of my PhD thesis, which develops a new dynamic stochastic model for forecasting future earthquakes. It is based on the idea that the physical system is successively in different states, and in each of them a different stochastic model describes the seismic activity. This is the typical structure of a continuous-time state-space or state-observation model.
Chapter 2 tackles the problem of the inference of such a complex and rich model: the estimation of the generally unknown model parameters and the estimation of the filtering distribution.
A standard Markov Chain Monte Carlo method for Bayesian inference seems very hard to apply, because of the large dimension of the parameter space. Moreover,

when a new earthquake occurs, it is necessary to repeat the analysis on the global data set.

The collection of the data needs to be carried out in a sequential way because of the nature of the phenomenon. This fact suggested exploiting a new Bayesian sequential Monte Carlo method for the inference. This method, called also particle filter, can provide a solution to both the above mentioned estimation problems and allows to update the present estimates as new information comes in.

During my stay at the Department of Mathematics of the University of Bristol, under the supervision of Prof. Nicolas Chopin, I reviewed the literature of the particle filtering theory and checked that no standard methods are available for continuous-time state-space models. Then I examined some of the sequential Monte Carlo methodologies and I proposed a new version of them suitable for solving my problem.

In chapter 3, the filtering problem is solved assuming that the parameter values are known; the solution is obtained by exploiting the martingale representation of a point process so that it is possible to apply the innovation method, following the procedure used for example in [B81], and moreover taking into account the time dependence in the expressions of the conditional intensity function of the involved point processes. Then the solution of the filtering problem can be approximated using some Monte Carlo integrations.

This approximated solution can be oriented to improve the estimate of the filtering distribution obtained through the sequential Monte Carlo method proposed in chapter 2.

Finally, chapter 4 concerns the application of the proposed methods.

The phenomenon we are analysing evolves in extremely long time - millenia or even geological eras. Consequently, historical catalogues constitute the more commonly used data bases in statistical analyses. The worldwide longest historical catalogues are the Chinese, Iranian, Japanese and Italian ones, which date back up to 6000 years; of course, they are affected by incompleteness. In order not to add issues related to the data quality to the already highly structured model, I have preferred to examine a simulated data set focusing my attention on the properties of the estimation methods proposed and on their sensitivity to the assessment of some particular parameters.

First, I checked that the filtering solution in chapter 3 is satisfactory when the model parameters are considered as known.

Then I assumed unknown model parameters and applied the sequential Monte Carlo method proposed in chapter 2 proceeding in two steps. In the first step I applied

the particle filter jointly with a technique, due to Künsch and Hürzeler [DFG01], aimed to better explore the parameter space. In the second step the particle filter procedure is repeated without moving the particles, using the estimates obtained in the first step as input.

Finally, I substituted the plug-in estimate of the likelihood function of the different models in the procedure of chapter 3 in order to improve the filtering distribution approximation.

The results obtained are quite satisfactory, especially as regards the filtering problem in which the model parameters are known and the probability distribution of the states is dynamically estimated. Including the model parameters as components of the state to estimate raises remarkably the complexity of the problem; in the light of this fact, the results related to the estimation of all unknown quantities in the proposed state-space model can be considered satisfactory even if I think they can be further improved through a better calibration of the importance distribution and the combination of estimation techniques.

Chapter 1

The problem

This chapter is devoted to the presentation of the problem that I intend to tackle. First, basic and essential definitions of a marked point process and of a pure jump Markov process are provided. Then, some popular marked point processes for earthquake analysis are described and discussed.
The above mentioned processes are some of the constituent elements of the state-space model that I analyse in the present thesis.

1.1 Point processes

Point processes are stochastic processes modelling points randomly distributed in some space. A point corresponds to an event concerning some observable phenomenon and records occurrence time, location or some feature of the event. Point processes reveal to be suitable stochastic processes in many fields. Examples are the price modelling and the theory of credit risk in economy, the study of neural activity in biology, the analysis of forest fires and of earthquakes in environmental science, network theory and communications in engineering, image reconstruction applied in cartography, forest cover study, medical image analysis, etc...

In order to define mathematically a point process, I recall some concepts on measurability, referring to [B81], [K91] and[LS01.I].
Let $\mathbf{X} = \{X_t\}_{t\geq 0}$ be a random process on a probability space $(\Omega, \mathcal{F}, P)$ in a measurable space $(E, \mathcal{E})$.

A sequence of sub-σ-fields $\{\mathcal{F}_t\}_{t\geq 0}$ of $\mathcal{F}$ is defined a **filtration** when

$$\mathcal{F}_s \subseteq \mathcal{F}_t \quad \text{for all } s \leq t$$

A filtration $\{\mathcal{H}_t^X\}_{t\geq 0}$ is the **internal history** of $\mathbf{X}$ if each $\mathcal{H}_t^X$ is the σ-field generated by $\{X_s, s \in [0,t]\}$:

$$\mathcal{H}_t^X = \sigma(X_s, s \in [0,t])$$

A **history** of $\mathbf{X}$ is every filtration $\{\mathcal{H}_t\}_{t\geq 0}$ such that

$$\mathcal{H}_t \supseteq \mathcal{H}_t^X \quad \text{for all } t \geq 0$$

When a filtration $\{\mathcal{H}_t\}_{t\geq 0}$ on $(\Omega, \mathcal{F}, P)$ is a history of $\mathbf{X}$, then $\mathbf{X}$ is said to be **adapted** to $\{\mathcal{H}_t\}_{t\geq 0}$. In such a case, X_t is $\mathcal{H}_t$-measurable, for every $t \geq 0$ (that is, $\{\omega : X_t(\omega) \in B\} \in \mathcal{H}_t$ for all $B \in \mathcal{E}$).

The process $\mathbf{X}$ is said to be (**right continuous, left continuous**) **continuous** if the trajectories $t \longrightarrow X_t(\omega)$ are (P-a.s. right continuous, P-a.s. left continuous) P-a.s. continuous.

An adapted process $\mathbf{X}$ to the filtration $\{\mathcal{H}_t\}_{t\geq 0}$ is said to be **increasing** if $X_0 = 0$ and $X_s \leq X_t$, P-a.s., for all $0 \leq s \leq t$.

Let $\mathbf{X}$ and $\{\mathcal{H}_t\}_{t\geq 0}$ be respectively a random process and a filtration on $(\Omega, \mathcal{F}, P)$; $\mathbf{X}$ is called to be $\mathcal{H}_t$-**measurable** if and only if $\mathbf{X}$ is adapted to $\{\mathcal{H}_t\}_{t\geq 0}$ and

$$\{(\omega, t) : X_t(\omega) \in B\} \in \mathcal{F} \times \mathcal{B}(\Re^+) \quad \text{for all } B \in \mathcal{E}$$

If the random process $\mathbf{X}$ is $\mathcal{H}_t$-measurable, then by Fubini's theorem its trajectories are $\mathcal{H}_t$-measurable ($\{t : X_t(\omega) \in B\} \in \mathcal{B}(\Re^+)$ for all $B \in \mathcal{E}$).

I set $E = \Re^+$ and $\mathcal{E} = \mathcal{B}(\Re^+)$.

The progressive measurability is now introduced; it is a stronger definition of measurability, since it allows that the measurability of the process $\mathbf{X}$ implies the measurability of the process $\mathbf{Y} = \{Y_t = \int_0^t X_s ds\}_{t\geq 0}$.

By definition, $\mathbf{X}$ is $\mathcal{H}_t$-**progressive** if, for all $t \in \Re^+$, the mapping

$$[0,t] \times \Omega \longrightarrow \Re^+$$
$$(s, \omega) \longrightarrow X_s(\omega)$$

is $\mathcal{B}([0,t]) \times \mathcal{H}_t$-measurable.

Note that every (left, right) continuous process is $\mathcal{H}_t$-progressive.

A further extension of progressiveness is represented by the concept of predictability. By definition, the random process $\mathbf{X}$ is $\mathcal{H}_t$-**predictable** if X_0 is $\mathcal{H}_0$-measurable and the mapping

$$(0,+\infty) \times \Omega \longrightarrow \Re^+$$
$$(t,\omega) \longrightarrow X_t(\omega)$$

is $\mathcal{P}(\mathcal{H}_t)$-measurable, where $\mathcal{P}(\mathcal{H}_t)$ is the σ-field over $(0,+\infty) \times \Omega$ generated by the rectangles of the form $(s,t] \times A$, for $0 \leq s \leq t$ and $A \in \mathcal{H}_s$.
The σ-field $\mathcal{P}(\mathcal{H}_t)$ is called the $\mathcal{H}_t$-**predictable** σ-**field** over $(0,+\infty) \times \Omega$.
It can be shown that the family of all $\mathcal{H}_t$-adapted and left continuous random processes is a generator of the σ-field $\mathcal{P}(\mathcal{H}_t)$.
Every $\mathcal{H}_t$-predictable process is also $\mathcal{H}_t$-progressive and every $\mathcal{H}_t$-adapted, left continuous random process is $\mathcal{H}_t$-predictable.
Moreover, an integrable increasing process $\mathbf{A}$ is $\mathcal{H}_t$-predictable if and only if for every positive $\mathcal{H}_t$-martingale process $\mathbf{M}$ and each $t \geq 0$

$$E\left[\int_0^t M_s \, dA_s\right] = E\left[\int_0^t M_{s-} \, dA_s\right]$$

I recall that a $\Re^+$-valued random variable T on $(\Omega, \mathcal{F})$ is a $\mathcal{H}_t$-**stopping time** if $\{\omega : T(\omega) \leq t\} \in \mathcal{H}_t$ for every $t \in \Re^+$; denoting $\mathcal{H}_\infty = \lim_{t \to \infty} \mathcal{H}_t$, the following σ-fields are defined:

$$\begin{aligned} \mathcal{H}_T &= \{A \in \mathcal{H}_\infty : A \cap \{T \leq t\} \in \mathcal{H}_t \text{ for all } t\} \\ \mathcal{H}_{T-} &= \sigma(A \cap \{T > t\} : t \geq 0, A \in \mathcal{H}_t) \vee \mathcal{H}_0 \end{aligned}$$

They are called respectively the past and the strict past of the process, since $\mathcal{H}_{T-} \subseteq \mathcal{H}_T$. Justification of the term *predictability* is that it can be shown that, if $\mathbf{X}$ is $\mathcal{H}_t$-predictable and T is a finite $\mathcal{H}_t$-stopping time, then X_T is $\mathcal{H}_{T-}$-measurable; while, if $\mathbf{X}$ is only $\mathcal{H}_t$-progressive, X_T is only $\mathcal{H}_T$-measurable.

1.1.1 Definition of point process

In this section, I present some basic concepts in point process theory.
For further in-depth studies and proofs, I refer to [B81], [K91], [DVj88] and [LS01.II].

A **point process** is a sequence of random variables $\mathbf{T} = \{T_n\}_{n \in \mathcal{N}}$ from a probability space $(\Omega, \mathcal{F}, P)$ to $(\Re^+, \mathcal{B}(\Re^+))$ such that

$$T_0 = 0$$
$$T_n < +\infty \implies T_n \leq T_{n+1}$$

The **associated counting process** $\mathbf{N} = \{N_t\}_{t \geq 0}$ of a point process $\mathbf{T}$ is defined by

$$N_t = \begin{cases} n & \text{if } t \in [T_n, T_{n+1}),\ n \in \mathcal{N} \\ +\infty & \text{if } t \geq T_\infty \end{cases}$$

where $T_\infty = \lim_{n \to +\infty} T_n$.

Sometimes, N_t is equivalently denoted by $N((0,t])$; for $A \in \mathcal{B}(\Re^+)$ in general, $N(A)$ denotes $\sum_{n=1}^{+\infty} I(T_n \in A)$, where I is the indicator function.

We say that a point process $\mathbf{T}$ is

- **nonexplosive** if $T_\infty = \lim_{n \to +\infty} T_n = +\infty$ (P-a.s.)
- **simple** if $T_i \neq T_j$ for $i \neq j$ (P-a.s.)
- **regular** if $\lim_{\Delta t \to 0} \frac{P\{N(A) > 1\}}{\Delta t} = 0$, where $A = [t, t + \Delta t]$.

It follows that each regular point process is obviously simple.
Moreover, $\mathbf{N}$ is a right-continuous step function such that

$$N_0 = 0$$
$$N_{T_n} - N_{T_{n-1}} = 1 \quad \text{for all } n \in \mathcal{N}$$

and, if $\mathbf{T}$ is nonexplosive, then $N_t < +\infty$ for all $t > 0$.
Observing that

$$N_t = card\{n \geq 1 : T_n \leq t\}$$
$$T_n = inf\{t \geq 0 : N_t \geq n\}$$

a one-to-one correspondence exists between $\mathbf{T}$ and $\mathbf{N}$, that is a point process and its associated counting process carry the same information; for this reason, the associated counting process is often called point process.

The distribution of a point process $\mathbf{N}$ is completely characterized by the class of finite dimensional distributions of the vectors $(N(A_1), N(A_2), ..., N(A_k))$, for each

$k \in \mathcal{N}$ and for each $A_1, A_2, ..., A_k \in \mathcal{B}(\Re^+)$.

Let $\mathbf{N}$ be a nonexplosive and regular point process and let $\{\mathcal{H}_t\}_{t\geq 0}$ be a history of $\mathbf{N}$. Since N_t is not decreasing with respect to t, it always holds that $E(N_t \mid \mathcal{H}_s) \geq N_s$ for $t > s$; that is $\mathbf{N}$ is a submartingale.

An increasing stochastic process $\mathbf{A} = \{A_t\}_{t\geq 0}$ such that for all $0 \leq s \leq t$

$$E[N_t - N_s \mid \mathcal{H}_s] = E[A_t - A_s \mid \mathcal{H}_s] \qquad \text{P-a.s.}$$

is called a **compensator** of $\mathbf{N}$.

Hence, the compensator of a point process is not unique and $\mathcal{H}_t$-adapted generally; but, from the **Doob-Meyer decomposition** for submartingales, there exists an essentially unique decomposition

$$N_t = M_t + A_t$$

of a point process $\mathbf{N}$ such that $\{M_t\}_{t\geq 0}$ is a zero mean $\mathcal{H}_t$-martingale process and $\{A_t\}_{t\geq 0}$ is a $\mathcal{H}_t$-predictable compensator of $\mathbf{N}$.

A nonnegative $\mathcal{H}_t$-progressive process $\{\lambda_t\}_{t\geq 0}$ satisfying

$$\int_0^t \lambda_s \, ds < +\infty \quad \forall t \geq 0, \quad P - a.s.$$

and one of the following equivalent conditions:

(i) for every nonnegative $\mathcal{H}_t$-predictable process C_t

$$E\left[\int_0^\infty C_s \, dN_s\right] = E\left[\int_0^\infty C_s \lambda_s \, ds\right] \qquad \text{P-a.s.}$$

(ii) for all $0 \leq s \leq t$

$$E[N_t - N_s \mid \mathcal{H}_s] = E\left[\int_s^t \lambda_u \, du \mid \mathcal{H}_s\right] \qquad \text{P-a.s.}$$

is called the **conditional intensity function** of $\mathbf{N}$.

Given a regular point process, it can be shown that a conditional intensity function exists always; moreover, every point process $\mathbf{N}$ having conditional intensity function

λ_t is nonexplosive and λ_t is related to its $\mathcal{H}_t$-predictable compensator A_t as follows

$$A_t = \int_0^t \lambda_s \, ds$$

This means that the conditional intensity function λ_t, for all $t > 0$, uniquely characterizes the point process $\mathbf{N}$.

The definition of conditional intensity function suggests its possible interpretation as the infinitesimal expected rate of events at time t, conditionally to the history $\mathcal{H}_t$ of the point process up to time t

$$\lambda_t = \lim_{\Delta t \longrightarrow 0} \frac{E\{N[t, t+\Delta t] \mid \mathcal{H}_t\}}{\Delta t}$$

or, equivalentely, the conditional infinitesimal occurrence probability at time t

$$\lambda_t = \lim_{\Delta t \longrightarrow 0} \frac{P\{N[t, t+\Delta t] > 0 \mid \mathcal{H}_t\}}{\Delta t}$$

It is worth enouncing the following important and useful theorem for martingales with respect to the internal history of a point process; such martingales have the form of a Stieltjes integral with respect to the fundamental martingale of the point process. I refer, for the proof, to the chapter 3. of [B81].

Integral representation of a point process martingale: *let* $\mathbf{N}$ *be a nonexplosive regular point process on* $(\Omega, \mathcal{F}, P)$ *with conditional intensity function* λ_t *and internal history* $\{\mathcal{H}_t\}$. *Let* M_∞ *be some* P*-integrable random variable such that* $\{M_t = E[M_\infty \mid \mathcal{H}_t]\}_{t \geq 0}$ *is a right-continuous* $\mathcal{H}_t$*-martingale on* $(\Omega, \mathcal{F}, P)$ *; then, for all* $t \geq 0$, *there exists a* $\mathcal{H}_t$*-predictable process* $\{C_t\}$ *satisfying* $\int_0^t C_s \lambda_s ds < +\infty$ P*-a.s. such that*

$$M_t = M_0 + \int_0^t C_s (dN_s - \lambda_s ds) \quad P - a.s. \tag{1.1}$$

Finally, it is interesting to recall the relation between the conditional probability density and the conditional intensity function of a regular point process.

The **conditional probability density** of $\mathbf{N}$ represents the probability of having an observation at time t, given the history of the process up to time T_{n-1}; it is denoted

by $p_n(t \mid \mathcal{H}_{T_{n-1}})$, for all $n \in \mathcal{N}$ and has support $(T_{n-1}, +\infty)$. Let

$$S_n(t \mid \mathcal{H}_{T_{n-1}}) = 1 - \int_{T_{n-1}}^{t} p_n(s \mid \mathcal{H}_{T_{n-1}})\, ds \quad \text{for} \quad s > T_{n-1}$$

be the associated survivor function; then, when $t \in (T_{n-1}, T_n]$, the following relation holds

$$\lambda_t = \frac{p_n(t \mid \mathcal{H}_{T_{n-1}})}{S_n(t \mid \mathcal{H}_{T_{n-1}})} \tag{1.2}$$

that is, λ_t corresponds to the conditional probability density $p_n(t \mid \mathcal{H}_{T_{n-1}})$ assuming that no event has occurred in (T_{n-1}, t).
Observing that $\lambda_t = -\frac{d}{dt} \log S_n(t \mid \mathcal{H}_{T_{n-1}})$, it follows

$$S_n(t \mid \mathcal{H}_{T_{n-1}}) = exp\left(-\int_{T_{n-1}}^{t} \lambda_s\, ds\right)$$

A naive interpretation is that $\lambda(T_n)$ represents the conditional probability (or the risk) of the occurrence of an event at time T_n, given the strict past history $\mathcal{H}_{T_n-}$; so λ_t is also called **risk function**. Moreover, $S_n(T_n \mid \mathcal{H}_{T_{n-1}})$ represents the conditional probability that no event occurs in (T_{n-1}, T_n).
Then, given a realization $(t_1, t_2, ..., t_n)$ of the point process, $t_i \in [0, T]$ for $i = 1, ..., n$, the corresponding **likelihood function** $\mathcal{L}$ can be defined through the conditional intensity function:

$$\mathcal{L} = \left[\prod_{i=1}^{n} \lambda_{t_i}\right] exp\left\{-\int_0^T \lambda_s\, ds\right\} \tag{1.3}$$

1.1.2 Marked point processes

When an additional information associated with each observation of a point process is available, we speak of marked point process.

Definition : *a* **marked point process** *is composed by two processes* $(\mathbf{T}, \mathbf{m})$, *where* $\mathbf{T} = \{T_i\}_{i \in \mathcal{N}}$ *is a point process on a probability space* $(\Omega, \mathcal{F}, P)$ *and* $\mathbf{m} = \{m_i\}_{i \in \mathcal{N}}$ *is a process on a measurable space* $(E, \mathcal{E})$.
$(E, \mathcal{E})$ *is called mark space and* m_i *is called mark.*

All definitions and theorems given for a point process can be extended to a marked point process. Though these extensions are sometimes obvious, I would like to specify some of them.

Definition :

- *the* **counting process N** *associated with a marked point process* $(\mathbf{T}, \mathbf{m})$ *is defined by*

$$N_t(A) = \sum_{i \geq 1} I(m_i \in A) I(T_i \leq t) \quad \text{for every } A \in \mathcal{E}, \quad t \geq 0$$

- *the* **internal history** *of a marked point process* $(\mathbf{T}, \mathbf{m})$ *is a filtration* $\{\mathcal{F}_t\}_{t \geq 0}$ *such that*

$$\mathcal{F}_t = \sigma(N_s(A) : 0 \leq s \leq t, \ A \in \mathcal{E})$$

- *let* **N** *be a marked point process such that* $N_t(A)$ *admits the conditional intensity function* $\lambda_t(A)$ *for each* $A \in \mathcal{E}$*; then,* $\lambda_t(dm)$ *is called the* **conditional intensity function** *of* **N**.

- *let* **N** *be a marked point process with conditional intensity function* $\lambda_t(dm)$*; then for each nonnegative* $\mathcal{F}$*-predictable E-marked process* C*, it holds that*

$$E\left[\int_0^{+\infty} \int_E C(t,m) dN_t(dm)\right] = E\left[\int_0^{+\infty} \int_E C(t,m) \lambda_t(dm) dt\right]$$

- *let* **N** *be a marked point process with conditional intensity function* $\lambda_t(dm)$ *such that*

$$\lambda_t(dm) = \lambda_t \ \Phi_t(dm) \tag{1.4}$$

where λ_t *is a nonnegative* $\mathcal{F}$*-predictable process and* $\Phi_t(dm)$ *is a density. Then* $(\lambda_t, \Phi_t(dm))$ *are called* $\mathcal{F}$*-local characteristics of* **N**.

In the next pages, I always deal with marked point processes and, for simplicity, I will denote them by T neglecting the mark process.

1.2 Point processes for earthquake sequences

In [Vj95] David Vere-Jones reviews the main research lines in predicting earthquakes and evolution of earthquake risk.
He takes into account possible earthquake precursors: increase in the frequency of smaller events, fluctuation of magnetic and electric fields, anomalous animal behaviour, variation of the water-bearing stratum levels, etc...
He recognizes the importance of these studies, even if he identifies some their weak points. First, collecting reliable and numerous enough data on possible precursors is generally expensive, taking into account that seismic phenomena evolve in extremely long time - millenia or even geological eras. Moreover, some precursor events are sometimes observed before some earthquakes, but not always; and sometimes, after their observation, no seismic event occurs.
Seismological and geological measurements constitute the main and the most natural information sources on which to base an earthquake analysis; this information is collected in historical catalogues, current instrumental catalogues, reports on paleoseismologic studies and on measurements of plate tectonic motions.
Historical catalogues are affected by incompleteness, due to the lack of historical memory, especially for small earthquakes; current instrumental catalogues are in practice complete, but they cover a too short time period with respect to the seismic evolution; paleoseismic knowledge concerns ground traces of few, strong and ancient earthquakes; the evaluation of the plate motion rates is still an open problem.
Undoubtedly, the earthquake formation problem is complex and the long-term temporal extension of this phenomenon does not help to quickly and efficiently validate the proposed models.
The uncertainty on many aspects of seismic activity leads to retain that probabilistic and statistical methods are among the most suitable approaches to the analysis of earthquakes.
Vere-Jones affirms that "*the primary task of the scientist concerned with earthquake prediction should be the development of theories and models which allow such conditional probabilities* (that is, the probabilities of occurrence of an earthquake) *to be explicitely calculated from past information*" [Vj95].
Vere-Jones himself first introduced point processes for earthquake analysis ([VjD66], [Vj78]); these stochastic processes seem to offer a so reasonable and promising statistical instrument in this field, that many parametric models were proposed and studied in order to analyse trend, periodicity, clustering, fault interaction, etc...

A seismic sequence can be interpreted as a realization of a point process: for example, simple point processes can model temporal sequences, multivariate point processes or marked point processes can model space-time or time-magnitude sequences; moreover information coming from precursor phenomena can be included into point processes.

As described in previous sections, point processes have a solid and flexible mathematical structure. The conditional intensity function uniquely characterizes a point process and it is the basic element involved in the definition of its likelihood function. So, the statistical analysis can be carried out by exploiting all likelihood based methods for estimation of the model parameters and for model selection.

In the next section, I describe some of the most important point process models for earthquake analysis; I resume in the presentation of these models some of the results that I obtained by applying them to the historical catalogues of some strongly seismic Italian regions.

I point out that I neglect models aimed at checking some seismic periodicity, because numerous and different statistical studies on this subject showed null or marginal evidence of this effect (e.g. [VjY88]).

I introduce now some notation defining an earthquake data set.

Let $[0,T]$ be the observation time interval, in which only events with magnitude equal to or greater than a threshold (or cut-off) value M_0 are recorded; $[0,T]$ and M_0 are generally chosen on the basis of the catalogue completeness.

An earthquake sequence will be denoted by

$$\{(t_i, m_i, z_i)\}_{i=1,...,n}$$

where each tern (t_i, m_i, z_i) represents occurence time, magnitude and location (latitude, longitude, depth) of the i^{th} observed event in $[0,T]$ and it holds

$$\begin{aligned} & t_i \in [0,T] && \forall i = 1,...,n \\ & m_i \geq M_0 && \forall i = 1,...,n \\ & t_i < t_j && \forall i < j \end{aligned}$$

It is useful to set $t_0 = 0$, $t_{n+1} = T$ and, if no events are recorded at times 0 and T, $m_0 = m_{n+1} = 0$.

I consider marked point processes for the analysis of time-magnitude seismic sequences. The temporal component of the sequence is modelled by a point process and the magnitude is the mark characterizing each event. Then the conditional

intensity function can be written in the form (1.4), that is

$$\lambda_t(dm) = \lambda_t \Phi_t(dm)$$

where λ_t (sometimes denoted by $\lambda(t)$) is the conditional intensity function of the temporal point process and $\Phi_t(dm)$ is the density of the distribution of the magnitude. About the magnitude distribution, there exists a famous and accredited empirical formula, the **Gutenberg-Richter law**, according to which the number $F(m)$ of earthquakes with magnitude m or larger is given by

$$F(m) = 10^{a-bm} \tag{1.5}$$

where a is related to the frequency at some reference magnitude and b, called **b-value**, varies from area to area and usually takes a real value in the interval $[0.7-1.2]$. Following the Gutenberg-Richter law, a time-independent exponential distribution with mean $1/\log_{10} b$ is often assumed as the magnitude distribution. Theoretically there is no upper or lower magnitude limit, but observations suggest that the magnitude is a positive number less than 10; hence, a truncated exponential distribution for magnitude is sometimes proposed.

It is worth noting that the Gutenberg-Richter law assumes the magnitude independence of the time, but this is an open question.

In ([ZVj94]), for example, a weak temporal dependence of the magnitude seems to be statistically revealed by the analysis of Chinese data and the magnitude distribution is consequently chosen time-independent. But the contrary assumption is not in general unreasonable (e.g. [RV03]).

Of course, the magnitude-time independence hypothesis allows to simplify the quite complicate estimation problem, since in such a case the conditional intensity function $\lambda_t(dm)$ is proportional to λ_t and the magnitude part $\Phi_t(dm) = \Phi(dm)$ can be omitted from the analysis of the temporal part. So, let $\Phi_t(dm)$ be in the form $\Phi(dm)$ unless different indication, and for example modelled through the Gutenberg-Richter law.

1.2.1 Poisson model

The simplest statistical model for time occurrences is the homogeneous Poisson process; by definition, it is a point process $\mathbf{N}$ such that, for each disjoint $A_1, A_2, ..., A_k \in$

$\mathcal{B}(\Re^+)$, the random variables $N(A_1), N(A_2), ..., N(A_k)$ are independent and have a Poisson distribution, that is [GS90]:

$$P(N(A) = j) = \frac{(\mu \ell(A))^j}{j!} e^{-\mu \ell(A)} \qquad A \in \mathcal{B}(\Re^+)$$

where ℓ is the Lebesgue measure on $\Re$.

A Poisson process is characterized by a constant conditional intensity function (the constant rate) with respect to t :

$$\lambda_t = \mu \quad \text{for some} \quad \mu \in \mathcal{R}^+$$

Then, the expression of the likelihood function on $[0, T]$ is

$$\mathcal{L} = \mu^n \, e^{-\mu T}$$

Since the survivor function is $S(t) = exp\{-\mu \ t\}$, the waiting times $(T_{n+1} - T_n)$ between two consecutive events are independent and identically exponentially distributed with mean $1/\mu$.

From the previous observation, it is clear that a Poisson process is chosen to model a seismic sequence when the hypothesis of random time occurrences is assumed. In order to test the Poisson assumption for a given data set, a quick check is possible [VjY88]: since $\log S(t) = -\mu \ t$, the logarithm of the survivor function for the empirical distribution of the waiting times must have an almost linear trend.

Let us consider, for example, two Italian regions, one corresponding to the Calabrian arc and the other to whole Southern Italy. The data are collected in the NT4.1.1 Parametric Italian Catalog and belong to the seismogenic zones from 65 to 72 of the ZS4 zonation for the Calabrian arc and to the seismogenic zones from 56 to 80 of the seismic zonation ZS4 for Southern Italy (see Appendix A). Just events having magnitude equal to or greater than 6.0 and occurred between 1600 and 1992 are considered, since the NT4.1.1 Parametric Italian Catalog is generally assumed as complete in this period: the Calabrian arc counts 14 events and Southern Italy 33 events.

Assuming the Poisson model for both the regions, the average number of events per year is $\mu = 0.0357$ for the Calabrian arc and $\mu = 0.0842$ for Southern Italy.

In figure 1.1, empirical and plug-in estimated log-survivor functions for the interevent time are compared for (a) the Calabrian arc and (b) Southern Italy. Case (b) shows a greater agree with the Poisson hypothesis than case (a). Indeed, Southern Italy is so vast that its seismicity can be considered as a superposition of independent

earthquake sequences; it is known that this situation is well approximated by a Poisson process ([DVj88]).

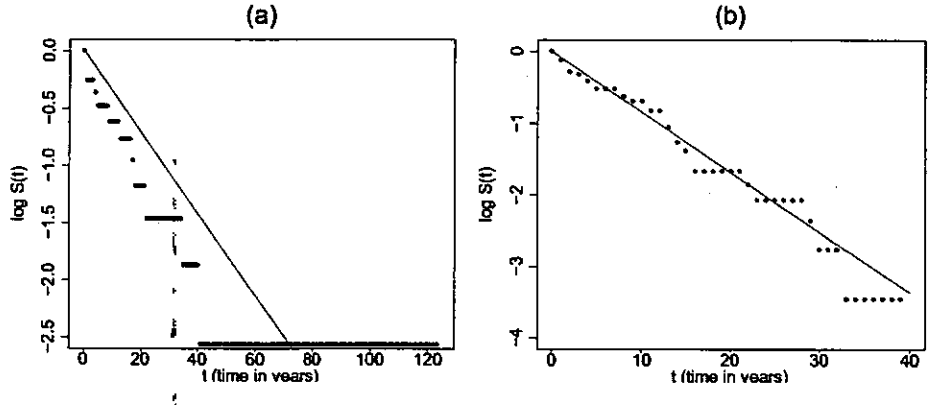

Figure 1.1: Empirical log-survivor function, $\log S(t)$ (dot), and theoretical log-survivor function, $-\mu t$ (line), versus time for (a) the Calabrian arc and (b) the South of Italy data sets.

1.2.2 Stress release models

The seismologist Harry Fielding Reid, observing the 1906 San Francisco earthquake, developed the elastic rebound theory of earthquake mechanics: some elastic stress in a seismically active region gradually accumulates and is suddenly and violently released when the stress exceeds the strength of the medium.
In the literature, different stochastic versions of Reid's theory were proposed.
If the medium strength is fixed, we have the time-predictable model [SN80], if a minimum level of stress is fixed, the slip-predictable model [KA84], while if neither upper nor lower stress bound exists, one speaks of stress release model [Vj78].
Vere-Jones proposed the stress release model (SRM) in 1978 developing the stochastic Markov model for main earthquakes by Knopoff (1971, [K71]).
The SRM belongs to the so called class of *self-correcting models*; good fits are observed for declustered catalogue, that is for sequences of strong events.
In the SRM, at time t, the conditional intensity function depends on the stress level $S(t)$ present in the considered region

$$\lambda_t = \lambda(S(t))$$

It follows that the occurrence probability is assumed as increasing with respect to the stress level $S(t)$ and abruptly decreasing when an event occurs.

In [ZVj94], it is shown that the global behaviour of the estimated conditional intensity function $\lambda(S(t))$ is not very sensitive to different choices of its functional form, provided it is a positive, convex and increasing to infinity function.
Very diffuse and computationally easy forms of $\lambda(S(t))$, for example, are the following

$$\lambda(S(t)) = [\tilde{\alpha} + \beta S(t)]^+ \tag{1.6}$$

$$\lambda(S(t)) = e^{\tilde{\alpha}+\beta S(t)} \tag{1.7}$$

where $\tilde{\alpha} \in \mathcal{R}$ and $\beta > 0$ are unknown constants, related to background and sensitivity to risk respectively (notation: $[x]^+ = x$ when $x \geq 0$ and $[x]^+ = 0$ otherwise).
In order to define $S(t)$, each earthquake with magnitude $m_i \geq M_0$ is supposed to release a stress quantity r_i and an approximation of r_i is provided by the empirical Benioff formula:

$$r_i = 10^{h(m_i - M_0)} \tag{1.8}$$

where $h = 0.75$ is the most commonly used value.
Let $R(t)$ be the accumulated stress released from all the events until time t

$$R(t) = \sum_{i:t_i<t} r_i$$

Assuming that the stress increases linearly with unknown constant rate $\rho > 0$ starting from an initial unknown stress level $S(0) > 0$, the difference between the accumulated stress $S(0) + \rho t$ and the released stress $R(t)$ up to time t provides the level of stress present at time t in the region:

$$S(t) = S(0) + \rho t - R(t) \tag{1.9}$$

Then, the conditional intensity functions of the SRM, corresponding to (1.6) and (1.7) respectively, are given by

$$\lambda_t = [\alpha + \beta(\rho t - R(t))]^+ \tag{1.10}$$

$$\lambda_t = e^{\alpha+\beta(\rho t - R(t))} \tag{1.11}$$

where α, β, ρ are parameters to estimate and $\alpha = \tilde{\alpha} + \beta S(0)$.
It can be noted that the Poisson model is a limit case of the SRM for β and ρ decreasing to zero.
The likelihood function (1.3) for the SRM in expression (1.10) is given by

$$\mathcal{L} = \left[\prod_{i=1}^{n}[\alpha + \beta(\rho t_i - R(t_i))]^+\right] exp\left\{-\sum_{i=0}^{n}(t_{i+1} - \tau_i)\left[\alpha - \beta R(t_{i+1}) + \frac{\beta\rho}{2}(t_{i+1} + \tau_i)\right]\right\}$$

where $\tau_i = \max_{t\in[t_i,t_{i+1}]}\{t : \lambda_t = 0\}$.
Instead, since the expression (1.11) for the SRM is always strictly positive, its likelihood function is simply given by

$$\mathcal{L} = \left[\prod_{i=1}^{n} e^{\alpha+\beta(\rho t_i - R(t_i))}\right] exp\left\{-\sum_{i=0}^{n}\frac{e^{\alpha-\beta R(t_{i+1})}[e^{\beta\rho t_{i+1}} - e^{\beta\rho t_i}]}{\beta\rho}\right\}$$

It is worth mentioning some of numerous developments of the SRM in order to better understand its potentiality and some possible future extensions of this work.
Many papers in the literature are devoted to model the seismic sequences of neighbouring regions, involving the spatial component.
The boundaries of the regions are generally suggested by experts on the basis of fault locations and geological information.
For example in [ZVj94], Zheng and Vere-Jones considered two subregions for Chinese data and three subregions for the Persian data and the Japanese data.
Moreover, they considered the possibility that in the same region two different processes could model respectively small and strong events (studying Chinese data, they fixed $M_0 = 6.0$ or 6.5 and events with magnitude lower than 7.6 are considered as small).
The main idea is that K subregions (K geographical neighbouring regions or $K/2$ geographical neighbouring regions, each of them driven by two point processes, one for small events and another for strong events) are each dominated by independent SRM or Poisson models; then the conditional intensity function is a combination of the probabilities π_k that an event occurs in the k^{th} region times the corresponding conditional intensity function $\lambda_t^{(k)}(m)$:

$$\lambda(t,m) = \sum_{k=1}^{K}\pi^{(k)}\ \lambda_t^{(k)}(m) \tag{1.12}$$

Assuming a time-independent magnitude distribution, each $\lambda_t^{(k)}(m)$ is proportional

to the risk function $\lambda_t^{(k)}$; so, it can be proved that

$$\pi^{(k)} = \frac{\lambda_t^{(k)}}{\sum_{i=1}^{K} \lambda_t^{(i)}}$$

and a spatial probability distribution on the K subregions can be estimated.
Jointly with Rotondi R., I applied this model to the events occurred in the zone 729 (the Calabrian arc) of the zonation ZS7, between 1600 and 1992, recorded in the Italian Catalog CPTI (see Appendix B).
I considered two possible different choices of the cutoff magnitude value: $M_0 = 4$ or $M_0 = 5$. I assumed that the weak events of the zone 729 are driven by a Poisson model, while the strong events are driven by a SRM; so, in order to define strong and weak events, a threshold magnitude M_1 is fixed equal to $M_1 = 5$ or $M_1 = 5.8$.
I recall that M_0 choice depends on the completeness of the catalogue. On the contrary, $M_1 = 5$ is a typical value used in the literature for other regions of the world and $M_1 = 5.8$ is a value advised by an Italian geologist on the basis of his experience during an informal talk.
I obtained maximum likelihood estimates of the model parameters by considering different combinations of the two threshold values M_0 and M_1 (see Table 1.1).
The Akaike Information Criterion, AIC, is used to compare models [A77].
The AIC value is defined by the maximum likelihood value $\mathcal{L}$ and the number of model parameters q as follows: $AIC = 2(q - \log \mathcal{L})$. The model having minimum AIC value is selected and a difference of $1.5 - 2$ in AIC is usually considered significant.
In Table 1.1, we can see that the best fitting is provided by $M_1 = 5.8$ in both the data sets defined by $M_0 = 4$ and $M_0 = 5$ respectively; for $M_0 = 4$, the estimated conditional intensity functions are represented in Figure 1.2: blue line corresponds to the model having $M_1 = 5.8$ and black line to that having $M_1 = 5$; analogously, for $M_0 = 5$ in Figure 1.3.

<table>
<tr><th>$4 \leq m < 5$</th><th>$5 \leq m < 5.8$</th><th>$m \geq 5.8$</th><th>AIC</th></tr>
<tr><td colspan="2">Poisson</td><td>SRM</td><td>630.63</td></tr>
<tr><td>Poisson</td><td colspan="2">SRM</td><td>666.02</td></tr>
<tr><td></td><td>Poisson</td><td>SRM</td><td>239.96</td></tr>
<tr><td></td><td colspan="2">SRM</td><td>240.48</td></tr>
</table>

Table 1.1: AIC values for different formulations of the model (1.12) applied to data drawn from the CPTI catalogue related to the zone 729 of ZS7 zonation.

In particular, black line in Figure 1.3 represents the conditional intensity function of a simple SRM like in (1.11), because $M_0 = M_1 = 5$.
This example points out the importance of basing the choice of every quantity on physical motivation.

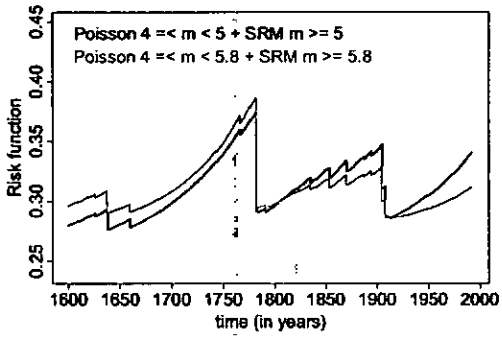

Figure 1.2: Combinations of Poisson model and SRM on the Calabrian arc data for $M_0 \geq 4$.

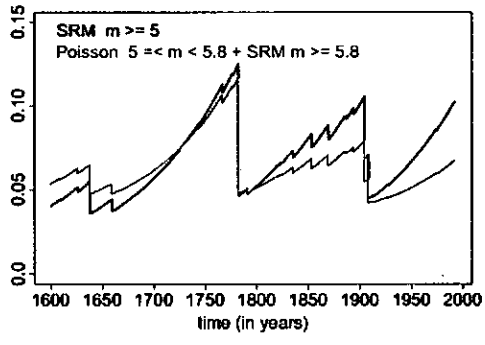

Figure 1.3: Combinations of Poisson model and SRM on the Calabrian arc data for $M_0 \geq 5$.

The **linked SRM** is another spatio-temporal extension of SRM, where possible interaction and stress transfer between neighbouring regions are allowed [LCSVj99], [LHB99], [BH03]; as in (1.12), the regions are dominated by different, but eventually dependent, processes. In order to render this dependence explicit, new interaction parameters θ_{jk} are introduced in the expression of the level of stress $S_j(t)$ of each region j :

$$S_j(t) = S_j(0) + \rho_j\, t - \sum_{k=1}^{K} \theta_{jk} R_k(t) \quad \text{for } j = 1, 2, ..., K$$

where $R_k(t)$ is the accumulated stress release due to the events occurred in region k up to time t; that is, denoting by z_i the region where the i^{th} observed event occurred, $R_k(t)$ is defined as follows

$$R_k(t) = \sum_{i:t_i<t,\ z_i=k} r_i$$

Since, for definition, each event implies a stress release corresponding to its size, we set $\theta_{jj} = 1$ in the region j where it occurs; while each $\theta_{jk} \in \Re$ for $k \neq j$ represents the proportion of stress that the region j loses or gains when an event occurs in region k.
The conditional intensity function for the whole region is the sum of the conditional

intensity functions of each subregion:

$$\lambda_t = \sum_{j=1}^{K} \lambda_t^{(j)}$$

where, following for example (1.6),

$$\lambda_t^{(j)} = [\tilde{\alpha}_j + \beta_j S_j(t)]^+ \tag{1.13}$$

For each region j, the model parameters are α_j, β_j, ρ_j, θ_{jk} for $j \neq k$, where $\alpha_j = \tilde{\alpha}_j + \beta_j S_j(0)$.
It is interesting to observe that, if θ_{jk} is positive (negative), the risk of the region j is subject to a damping (exciting) effect in correspondence of each event occurred in the region k; if $\theta_{jk} = 0$, the region j does not depend on the seismicity of the region k. In particular, if $\theta_{jk} = 0$ for all $j \neq k$, the model is called **indipendent SRM** to point out the lack of interactions among the regions.

I would like to show a Bayesian analysis of the linked stress release model, that I carried out jointly with Rotondi. In [RV03], we studied a simple SRM in the form (1.7) with magnitude distribution dependent on the level of stress; now, I present an extension of this work to the linked SRM [RV03.1].
The expression of the conditional intensity function we consider includes a time-dependent magnitude distribution:

$$\lambda_t(m) = \sum_{j=1}^{K} \lambda_t^{(j)} \Phi(m \mid S_j(t))$$

where $\lambda_t^{(j)}$ is defined as in (1.13).
Following the Gutenberg-Richter law, the magnitude distribution is chosen to be a truncated exponential distribution with the same parameter γ for each region, but different time-dependent support $[M_0, M_t^{(j)}]$.
The upper bound $M_t^{(j)}$ represents the maximum magnitude that can be observed at time t when, that is, the level of stress in the region j is $S_j(t)$; one obtains by inversion of the empirical Benioff formula (1.8):

$$M_t^{(j)} = M_0 + log_{10}\frac{S_j(t)}{h}$$

Then the magnitude distribution has the following expression

$$\Phi_t(m) = \frac{\gamma e^{-\gamma \cdot m}}{e^{-\gamma \cdot M_0} - e^{-\gamma \cdot M_t}} \, \delta_{[M_0, M_t]}(m) \tag{1.14}$$

The model parameters are γ and $\tilde{\alpha}_j$, β_j, ρ_j, $S_j(0)$, θ_{jk} for $k \neq j$, for all $j = 1, ..., K$. For each region j, we impose two physical constraints influencing the parameter space:

1. the risk cannot be null in correspondence of an earthquake, that is $\lambda_{t_i}^{(j)} > 0$ when $z_i = j$, for all $i = 1, ..., n$; it follows a constrained range $[\underline{\alpha}_j, +\infty]$ of the parameter α_j, where

$$\underline{\alpha}_j = -\beta_j \cdot \min_{i=1,...,n,\ z_i=j} \left\{ S_j(0) + \rho_j t_i - \sum_{k=1}^{K} \theta_{jk} R_k(t_i) \right\}$$

2. the level of stress is always positive, that is $S_j(t_i^+) \geq 10^{-0.75 M_0}$, where t_i^+ denote the infinitesimal instant following an earthquake and the minimum possible level of stress $10^{-0.75 M_0}$ is given by the Benioff formula (1.8) setting $m_i = 0$. It follows a constrained range $[\underline{S}_j, +\infty]$ of the parameter $S_j(0)$, where

$$\underline{S}_j = 10^{-0.75 M_0} + \max_{i=1,...,n} \left\{ 0;\ \rho_j t_i + \sum_{k=1}^{K} \theta_{jk} R_k(t_{i+1}) \right\}$$

It follows that $\tilde{\alpha}_j$, β_j, ρ_j, $S_j(0)$, θ_{jk} for $k \neq j$ must be updated jointly in the MCMC estimation procedure because they are dependent on each other; let v_j denote the vector of this parameters.

We considered the Italian region Sannio-Matese-Ofanto-Irpinia (left-hand picture in Figure1.4), divided in three subregions corresponding respectively to the zones 58, 62 and jointly 63 and 64 of the ZS4 zonation (Appendix A).

There are 31 events occurred between 1650 and 1992 and having magnitude equal to or greater than 5, in the NT4.1.1 Parametric Italian Catalog (see right-hand side of Figure 1.4).

Since the region 1 does not border on the region 3, we suppose that no interaction exists between these two regions, so we set $\theta_{13} = \theta_{31} = 0$.

In order to estimate the seventeen model parameters using such an exiguous data set, a Bayesian analysis is performed through an MCMC stochastic simulation method based on the Metropolis-Hastings algorithm.

Because of the physical model constraints, we assume the prior distribution of the

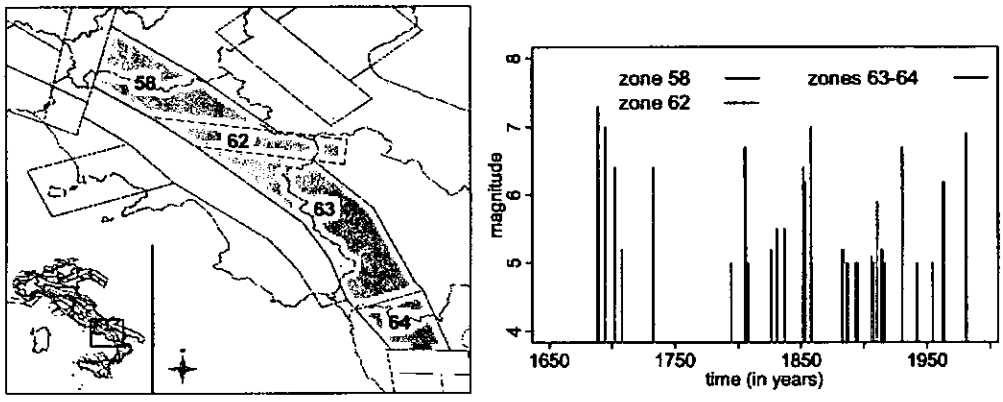

Figure 1.4: Magnitude versus time for the Italian region Sannio-Matese-Ofanto-Irpinia. Colors denote the region to which each event belongs: black, green and red respectively for regions 1 (zone 58), 2 (zone 62) and 3 (zones 63-64)

vector v_j expressed by

$$\begin{aligned}\pi(v_j) &= \pi(\beta_j)\cdot\pi(\rho_j)\cdot\prod_{j\neq k}\pi(\theta_{jk})\cdot \\ &\quad \pi(S_j(0)\mid \rho_j,\theta_{jk} \text{ for } k\neq j)\cdot \\ &\quad \pi(\tilde{\alpha}_j\mid \beta_j,S_j(0),\rho_j,\theta_{jk} \text{ for } k\neq j)\end{aligned}$$

Table 1.2 contains the chosen prior distributions; the prior parameters for $\tilde{\alpha_j}$, β_j, ρ_j are deduced from the literature (e.g. [ZVj91], [ZVj94]), while for γ are chosen according to the values generally taken by the b parameter of the Gutenberg-Richter law.

	prior distribution	mean	variance	support
γ	*Gamma*	1.	1.	
β_j	*Gamma*	0.005	0.000005	
ρ_j	*Gamma*	0.5	0.25	
θ_{jk} for $k=1..R$	*Normale*	0.	4.	
$S_j(0)$	*Uniforme*	-	-	$[\underline{S}_j, 100]$
α_j	*Normale*	-0.2	0.1	$(\underline{\alpha}_j, +\infty)$

Table 1.2: Prior distributions for the linked stress release model having time-dependent magnitude distribution

From the Bayes's theorem, the joint posterior distribution is proportional to the

product between the likelihood function and the prior distributions:

$$\pi(\gamma, v_1, v_2, ..., v_K \mid data) \propto \mathcal{L}(data \mid \gamma, v_1, v_2, ..., v_K) \cdot \pi(\gamma) \cdot \prod_{j=1}^{K} \pi(v_j) \qquad (1.15)$$

It is worth noting that the prior distributions of $S_j(0)$ and α_j, for $j = 1, ..., K$, have supports varying with the value of the other parameters; so, their normalising constants cannot be neglected in the posterior calculation (1.15).

A simulation from the posterior distribution can be obtained through a Markov Chain Monte Carlo method: it produces an ergodic Markov chain of parameter vectors whose stationary distribution is the posterior distribution. The empirical average of the values of the Markov chain converges to the posterior mean of the parameters. Among the MCMC methods, we use a Metropolis-Hastings algorithm with multivariate transition kernel.

The first vector of the chain is sampled from the prior distributions; in order to respect the physical constrained, $S_j(0)$ and $\tilde{\alpha}_j$ must be drawn last among the parameters in v_j.

Let w denote one of the parameters and let $w^{(1)}$ be its value in the first vector of the chain, while $w^{(h)}$ is its value in the h^{th} vector of the Markov chain. Then a candidate $w^{(h+1)}$ is generated according to the transition kernel $q(w^{(h+1)}, w^{(h)})$ and accepted with probability

$$\min\left\{1, \frac{\pi(w^{(h+1)} \mid data)\ q(w^{(h)}, w^{(h+1)})}{\pi(w^{(h)} \mid data)\ q(w^{(h+1)}, w^{(h)})}\right\}$$

	transition kernel	mean	variance	support
$\gamma^{(h+1)}$	$logNormal$	$\gamma^{(h)}$	1.	
$\beta^{(h+1)}$	$logNormal$	$b^{(h)}$	0.000005	
$\rho^{(h+1)}$	$logNormal$	$c^{(h)}$	0.25	
$\theta_k^{(h+1)}$	$Normal$	$\theta_k^{(h)}$	0.5	
$S(0)^{(h+1)}$	$Esponential$	10.		$[\underline{X}^{(h+1)}, 100]$
$\tilde{\alpha}^{(h+1)}$	$Normal$	$a^{(h)}$	0.05	$(\underline{a}_j^{(h+1)}, +\infty)$

Table 1.3: Transition kernels for Sannio-Matese-Ofanto-Irpinia data set.

Table 1.3 contains the transition kernels we used for each parameter component; they are centered on the current value of the parameters in the Markov chain and have fixed variance; the variances are chosen on the basis of the parameter sizes and calibrated through some pilot run of the program so as to have the acceptance rates

of the parameters between 25% and 35% as the literature advices.
The MCMC method is implemented generating a Markov chain of length 500000 and the first 100000 elements of the chain are neglected as burn-in. In order to decrease the autocorrelation between the elements of the chain, a thinning 80 is applied.
The program took 32 hours 15 minutes on a HP 9000/735 workstation; Table 1.4 contains the parameter estimates.
We check the convergence of the Markov chain to the stationary distribution using the statistical package BOA, available on the web site of the MCMC Preprint Service (*http://www.statslab.cam.ac.uk/~mcmc/*).
We applied successfully the statistical convergence diagnostics of BOA (Bayesian Output Analysis) to the Markov chain of the linked SRM: precisely, the Geweke convergence diagnostic, the Heidelberger and Welch stationarity and interval halfwidth tests, the Raftery and Lewis convergence diagnostic.

estimated parameters	$z58$	$z62$	$z63 - z64$
$\hat{\bar{\alpha}}_\cdot$	-0.1805	-0.2713	-0.0495
$\hat{\beta}_\cdot$	0.0031	0.0083	0.0017
$\hat{S}_\cdot(0)$	67.9433	31.5918	52.6111
$\hat{\rho}_\cdot$	0.4388	0.1375	0.3417
$\hat{\theta}_{\cdot 1}$	1.	-0.1282	0.
$\hat{\theta}_{\cdot 2}$	0.6467	1.	-0.1558
$\hat{\theta}_{\cdot 3}$	0.	-0.0022	1.
$\hat{\gamma}$	1.0974	1.0974	1.0974

Table 1.4: Estimated parameters for Sannio-Matese-Ofanto-Irpinia data set.

In Figure 1.5 the plug-in estimated conditional intensity function $\lambda_t^{(j)}$ of each subregion is represented separately; the vertical coloured lines indicate the earthquakes occurred in that region, and the length of the lines is proportional to their magnitude. Like in the SRM, the risk of the linked SRM is generally increasing with respect to time and decreases abruptly when an earthquake occurs. This happens, for example, in correspondence of the 1688 strong earthquake of magnitude 7.3 in zone 58.
Observing the picture, this is not the unique effect of this event: it can be noted that the risk of the zone 62 makes a jump in 1688 even if no event occurred in that area; this sudden increase of the risk in zone 62 is due to a stress transfer to zone 62 of a proportion θ_{21} of the stress released by the 1688 earthquake of the zone 58 (indeed $\hat{\theta}_{21}$ is negative). On the contrary, for example, the 1851 earthquake of magnitude 6.4 occurred in zone 62 has a damping effect on the risk of zone 58 (the estimated corresponding interaction parameter $\hat{\theta}_{12}$ is positive).

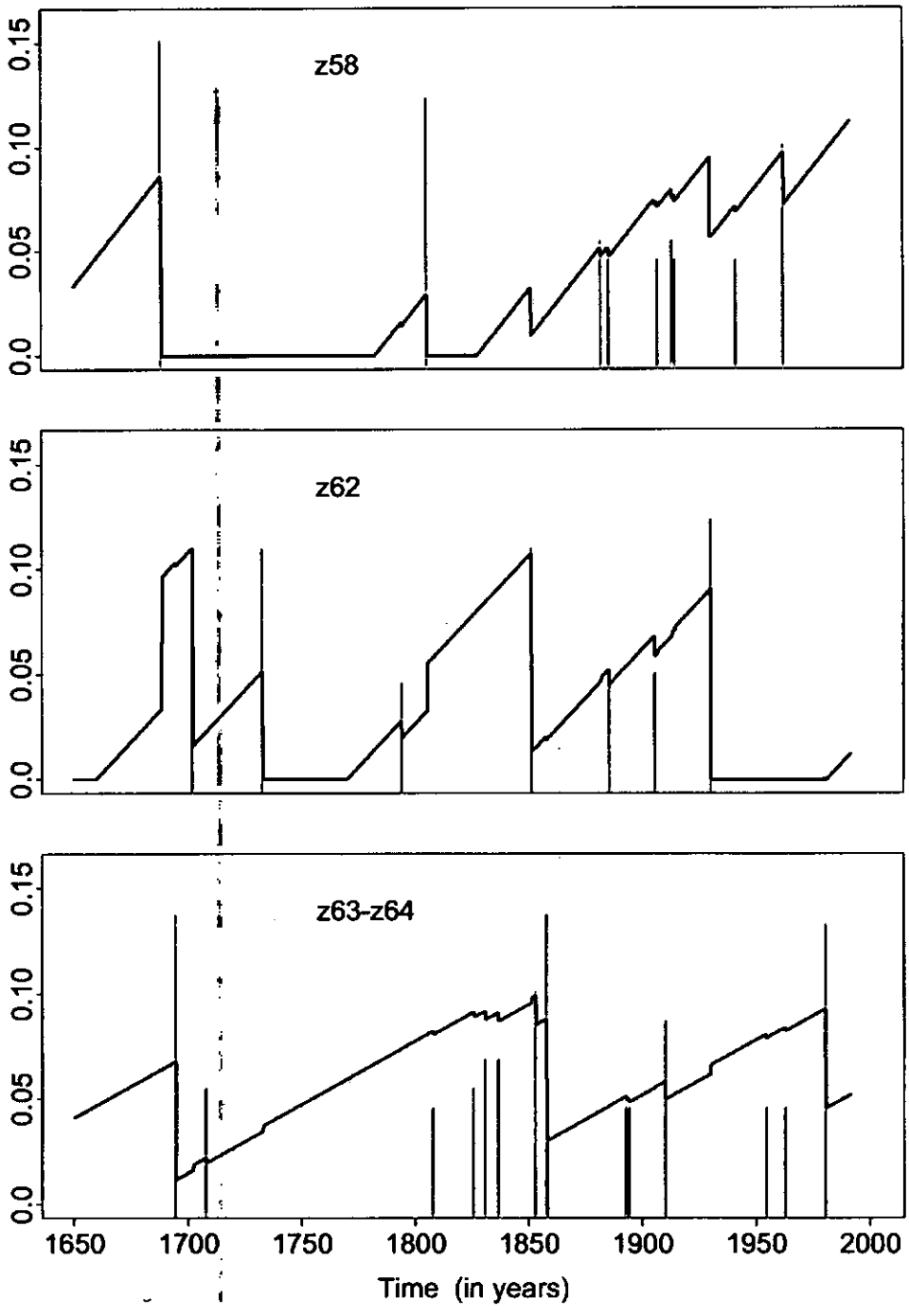

Figure 1.5: Estimated risk functions and events of each subregions of the Sannio-Matese-Ofanto-Irpinia for linked SRM having time-dependent magnitude distribution

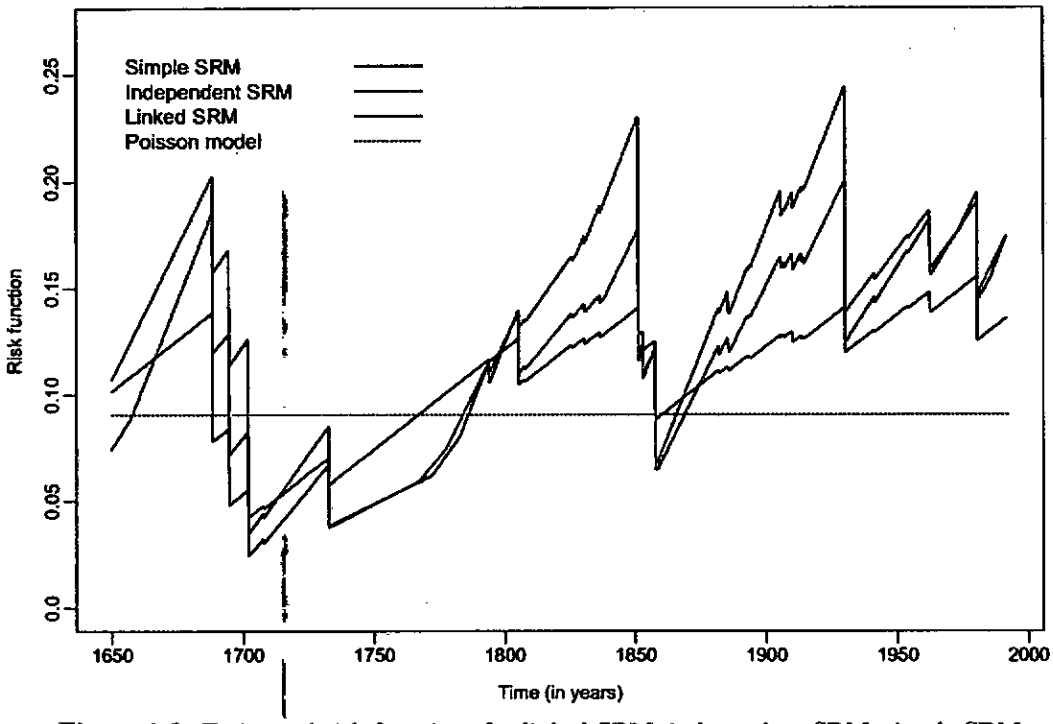

Figure 1.6: Estimated risk functions for linked SRM, independent SRM, simple SRM and Poisson model for the Sannio-Matese-Ofanto-Irpinia data set.

Figure 1.6 represents the estimated conditional intensity function for the linked SRM (red line), that is compared to the estimated conditional intensity functions of other three models.

Indeed, we have analysed in the same way the independent SRM (green line), in which all the interaction parameters θ_{jk} for $k \neq j$ are null, assuming a time-dependent magnitude distribution. Then, we have analysed the SRM (red line) again with a time-dependent magnitude distribution and the Poisson model (black dotted line), both on the global region composed by the zone 58, 62 , 63 and 64. The comparison among the values of the log-likelihood functions, obtained by substituting the parameter estimates, seems to reveal a better fitting of the models where the region is partitioned; but, since the difference between the logarithm of the likelihood for the linked SRM (-146.30) and that of the independent SRM (-146.81) is not relevant, there is no statistical evidence in favour of the linked SRM.

Moreover, this criterion does not take into account the different number of parameters in the two models; using for example the Akaike Information Criteria, the independent SRM should be preferred.

Figure 1.7 shows the conditional intensity function $\lambda_t(m)$ of the marked point process; it is possible to observe the variation of support of the magnitude distribution varying the time.

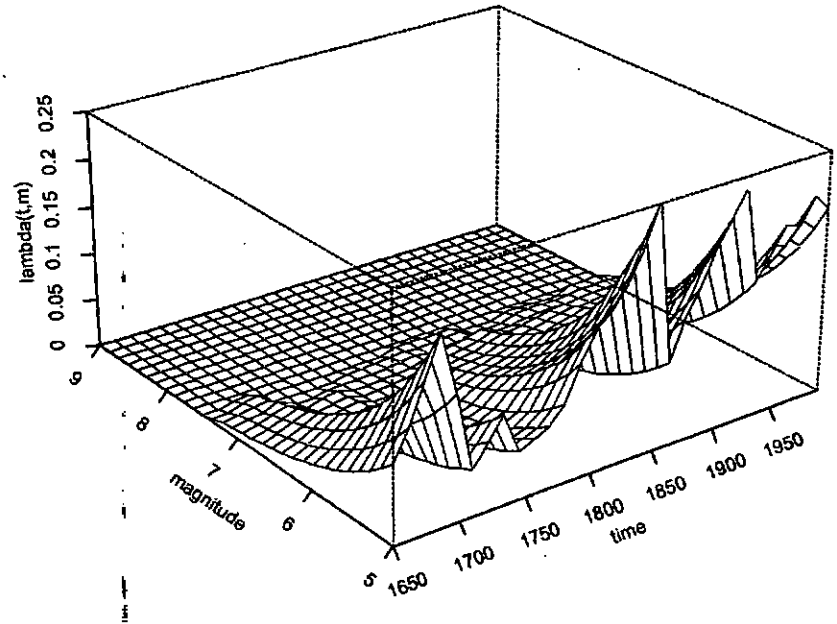

Figure 1.7: $\lambda_t(m)$ versus time and magnitude for the linked SRM applied to the Sannio-Matese-Ofanto-Irpinia data set

1.2.3 ETAS model

From the observation of seismic catalogues, a typical behaviour emerges: the earthquake tendence to occur in clusters, spatially and temporally.
Immediately after a "strong" event, a cluster of "smaller" events is observed; hence, the occurrence probability abruptly increases and then slowly decreases.
An event is called mainshock or triggering event if it is able to generate a cluster of events, otherwise it is called aftershock or triggered event.
The class of *self-exciting models* describes this behaviour in the context of point process models; the conditional intensity function generally assumes the form [HA73]:

$$\lambda(t) = \mu(t) + \int_0^t g(t-s)\, dN(s) = \mu(t) + \sum_{i:t_i<t} g(t-t_i)$$

where $\mu(t)$ represents a background rate and $g(t)$ is the triggering density.
Many expressions of $g(t)$ were proposed in the literature. For example, [L66] used

the following form, based on Boltzman's theory of elastic aftereffect,

$$g(t) = \Phi e^{-\theta\, t}$$

Many other formulations are based on the **Omori law**. In 1894, the seismologist Omori observed the decay of aftershock activity with time: aftershock frequency decreases by roughly the reciprocal of the time elapsed from the occurrence of the mainshock [UOM95], [O99]. Mathematically, assuming a mainshock occurs at time 0, the number $n(t)$ of its aftershocks is proportional to the reciprocal of the elapsed time t

$$n(t) = \frac{k}{t+c}$$

where k and c are positive parameters. The parameter k depends on the magnitudes in the cluster sequence. The parameter c is a controversial quantity, that seems strongly influenced by some eventual aftershock incompleteness in the first part of the sequence, due to overlapping on the seismogram; it is generally a "small" value, that avoids also numerical problems at the denominator.
In 1961, Utsu proposed the modified Omori formula [U61]

$$n(t) = \frac{k}{(t+c)^p}$$

introducing a new positive parameter p, that has the physical role to reflect mechanical conditions of the earth's crust and that can be interpreted as the decay rate of aftershocks; each region has a characteristic value of p, estimated by seismologists around 1.0.
In order to define a point process describing the distribution of an aftershock sequence, the basic idea is to express the conditional intensity function by the modified Omori formula:

$$\lambda(t) = \frac{k}{(t+c)^p}$$

Since pure sequences of aftershocks are rarely recognizable, while complex sequences of triggering and triggered events are often available, Vere-Jones and Davies (1966) proposed the trigger model, in which the triggering event is a hidden variable [VjD66]. To overcome the difficult detection of triggering events, Ogata proposed the **Epidemic-Type Aftershock-Sequence** (ETAS) **model**, a marked point process where each event shares in the generation of the subsequent events ([O88]). The ETAS model is very popular, because of its simplicity and its good fitting for many data sets. In the Etas model, since all the earthquakes have the double role of triggered and triggering shocks, the conditional intensity function is the sum of

contribution of all the events

$$\lambda_t(m) = \mu + \sum_{i:t_i<t} \frac{k_i}{(t - t_i + c)^p}$$

where c, p are positive parameters.
The numerator of $g(t)$ is not constant as in Omori law, but it is an explicit expression of the dependence of the frequency on the m_i magnitude; it is based on the empirical formula obtained by Utsu and Seki (1955), connecting aftershocks number and mainshock magnitude

$$k_i = ke^{\gamma(m_i - M_0)}$$

where k, γ are positive parameters [US55].
In particular, the parameter γ measures the sensitivity of the capability of generating secondary events to the magnitude of the triggering event. Earthquake swarms have generally small γ values; while sequences composed by a mainshock and few strong aftershocks have large γ values (for example, in Japan γ is typically around 2) [GO97],[O99].
For a suitable choice of the observed time interval and spatial area, it is possible to assume that there is no background seismicity and, consequently, to set $\mu = 0$; in that context, the likelihood function has two different expressions, one for $p = 1$ and one for $p \neq 1$:
if $p = 1$,

$$\mathcal{L} = \left[\prod_{i=1}^{n} \sum_{j:j<i} \frac{k\ e^{\gamma(m_j - M_0)}}{(t_i - t_j + c)^p}\right] \cdot exp\left\{-\sum_{i=0}^{n} \sum_{j<i+1} k\ e^{\gamma(m_j - M_0)}\ [\log(t_{i+1} - t_j + c) - \log(t_i - t_j + c)]\right\}$$

if $p \neq 1$,

$$\mathcal{L} = \left[\prod_{i=1}^{n} \sum_{j:j<i} \frac{k\ e^{\gamma(m_j - M_0)}}{(t_i - t_j + c)^p}\right] \cdot exp\left\{-\sum_{i=0}^{n} \sum_{j<i+1}^{i} k\ e^{\gamma(m_j - M_0)}\ \frac{(t_{i+1} - t_j + c)^{1-p} - (t_i - t_j + c)^{1-p}}{1 - p}\right\}$$

Further simplification of the previous expressions is possible by applying a cancellation of the terms in the exponential part, analogously to the trick of telescopic series [OMK93]:

if $p = 1$,

$$\mathcal{L} = \left[\prod_{i=1}^{n}\sum_{j<i}\frac{k\ e^{\gamma(m_j - M_0)}}{(t_i - t_j + c)^p}\right] \cdot exp\left\{\sum_{i=0}^{n} k\ e^{\gamma(m_i - M_0)}\ [\log c - \log(t_{n+1} - t_i + c)]\right\}$$

if $p \neq 1$,

$$\mathcal{L} = \left[\prod_{i=1}^{n}\sum_{j<i}\frac{k\ e^{\gamma(m_j - M_0)}}{(t_i - t_j + c)^p}\right] \cdot exp\left\{\sum_{i=0}^{n} k\ e^{\gamma(m_i - M_0)}\ \frac{c^{1-p} - (t_{n+1} - t_i + c)^{1-p}}{1-p}\right\}$$

The ETAS model and, in general, all the self-exciting models are used for complete catalogues starting from a small threshold magnitude, because they model earthquake clusters.

In order to show an application of the Etas model, I considered the seismic sequence occurred on the 26^{th} of September 1997 in Umbria-Marche Italian region, and started with two destructive events with magnitude respectively 5.6 and 5.8; this sequence is composed by 508 earthquakes having magnitude equal to or greater than 2.9 and occurred during about one month (*Waveforms, arrival times and locations of the 1997 Umbria-Marche aftershocks sequence* Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy; UMR Géoscience Azur, CNRS/UNSA, Valbonne, Grance; Laboratorio di Geofisica, Università di Camerino, Camerino, Italy.).

I carried out a Bayesian analysis of this data set, interpreted as a realization of an Etas model having $\mu = 0$.

Since the parameters are positive, the prior distributions, reported in Table 1.5, are supposed to be independent gamma distributions; means and variances of the prior distributions are calculated as means of the relative parameter estimates given in some studies on the Etas model presented in the literature ([GO97], [O83]).

An approximation of the posterior distributions of the parameters is obtained through a MCMC method exploiting the Metropolis-Hastings algorithm; the transition kernels (also named proposal distributions), reported in Table 1.5, are logNormal centered in the current values of the parameters in the Markov chain and having assigned variances.

A Markov chain of length 500000 is generated; the burn-in is fixed at the 20% of the Markov chain length, that is the first 100000 iterations are discarded; a thinning of 80 iterations is applied. The Bayesian analysis provides an approximation

of the posterior distribution of each parameter, represented in Figure 1.8; Table 1.5 contains the posterior means and standard deviations of the parameters.
Figure 1.9 represents the estimated risk function: it can be observed that in the Etas model the risk function suddenly increases in correspondence of an earthquake and then slowly decreases.

	prior distrib. (mean, variance)	**proposal distrib. (variance)**	**estimates (st.dev.)**
k	Gamma(1.0, 0.5)	logNormal(0.005)	**0.1237** (0.0258)
γ	Gamma(2.0, 2.0)	logNormal(0.05)	**0.8643** (0.1582)
c	Gamma(0.2, 0.1)	logNormal(0.03)	**0.1230** (0.0545)
p	Gamma(1.0, 0.2)	logNormal(0.1)	**1.5840** (0.1619)

Table 1.5: Summary of the main input and output for the Bayesian analysis of the Etas model applied to the Umbria-Marche data set.

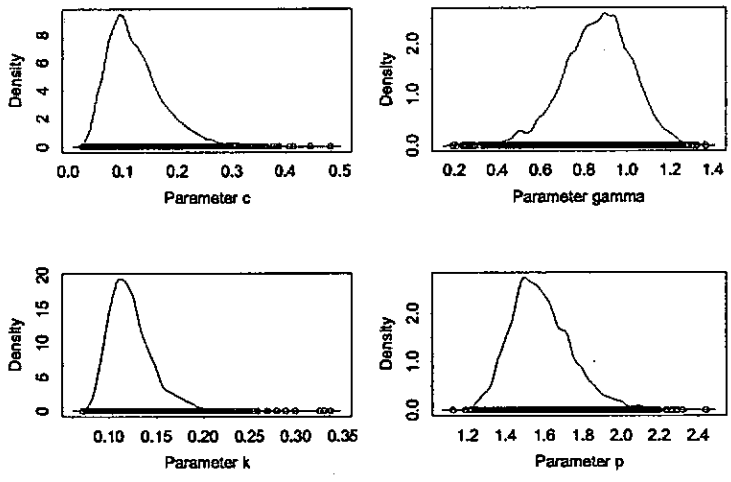

Figure 1.8: Posterior distributions of the Etas model parameters for the Umbria-Marche data set.

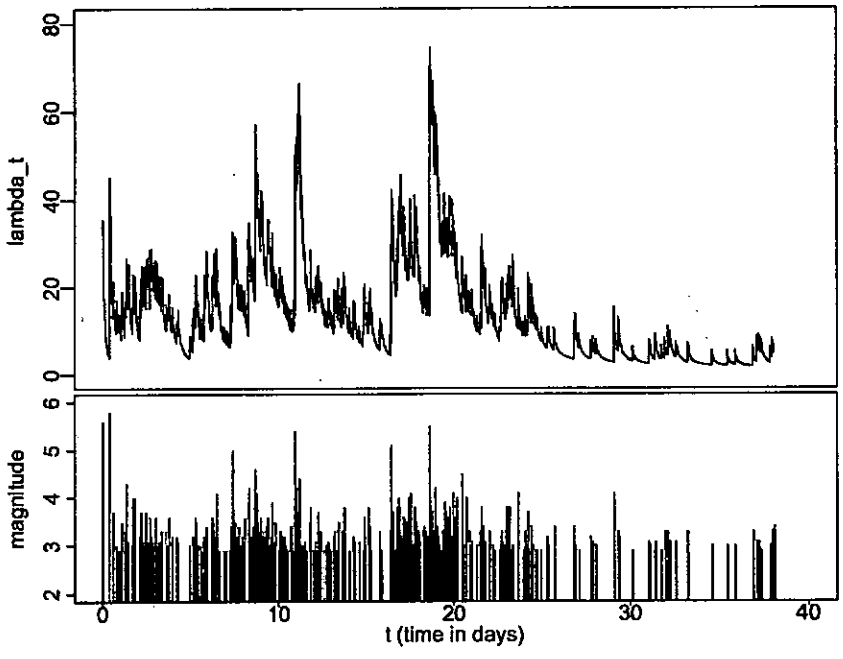

Figure 1.9: The estimated conditional intensity function of the Etas model for the Umbria-Marche data set.

Finally, I mention some other decay formulae proposed in the literature:

$$g(t) = k\, e^{-\gamma\, t}(t+c)^{-p}$$

is the product of the Omori formula and the exponential function [Ot87];

$$g(t) = \sum_{j=0}^{J} a_j t^{j-1} e^{-ct}$$

is an expansion by Laguerre-type polynomials, used to verify the hypothesis of a casual relationship between adjacent zones [O82], [DMZ88].
Self-exciting models, and especially ETAS model, have been recently extended to spatial components (e.g. [ZOVj02], [OKT03]).
For example, denoting by x_i, y_i respectively the latitude and longitude of the i^{th} mainshock event, Ogata et al. [OKT03] decomposed the aftershock rate function at time $t > t_i$ as follows:

$$g(t, m, x, y) = f(m_i) \cdot g(t - t_i \mid m_i) \cdot h(x - x_i, y - y_i \mid m_i)$$

where $f(m_i)$ is the magnitude distribution, $g(t - t_i \mid m_i)$ is defined as in the classical ETAS model and $h(x - x_i, y - y_i \mid m_i)$ represents the conditional spatial distribution.

Taking into account geophysical knowledge, it is defined as follows

$$h(x-x_i, y-y_i \mid m_i) = \left\{ \frac{\underline{z}^T S \underline{z}}{e^{2\alpha(m_i - M_0)}} + d \right\}^q$$

where $\underline{z}$ denote the vector $\begin{pmatrix} x-x_i \\ y-y_i \end{pmatrix}$, S is a suitable positive-definite symmetric matrix and α, d, q are parameters.

1.3 Pure jump Markov processes

I introduce the concept of homogeneous pure jump Markov process, that I will use in next section. I refer to [H03], [GS90] for more details and references on this subject.
Let $\mathcal{S}$ be a countable set, called space of the states.
Let $\mathbf{X} = \{X_t\}_{t \geq 0}$ be a $\mathcal{S}$-valued discrete-state continuous-time stochastic process on a probability space $(\Omega, \mathcal{F}, P)$.

Definition :

- **X** *is a* **time-homogeneous** *process if*

$$P(X_{t+r} \mid X_r) = P(X_t \mid X_0) \quad \forall r, t \geq 0$$

- **X** *is a* **Markov process** *if the following relation, called Markov property, is satisfied*

$$P(X_{t_n} = s_n \mid X_{t_1} = s_1, ..., X_{t_{n-1}} = s_{n-1}) = P(X_{t_n} = s_n \mid X_{t_{n-1}} = s_{n-1})$$

  for each $s_1, ..., s_n \in \mathcal{S}$ *and for each* $t_1 < t_2 < ... < t_n$ *in* $\Re$.

- *a* **pure jump Markov process X** *is a Markov process such that there exists a time sequence* $0 = t_0 < t_1 < t_2 < ...$ *in* $\Re$ *and a state sequence* $s_0, s_1, s_2, ... \in \mathcal{S}$, *where* $s_i \neq s_{i+1}$ *for each* $i \in \mathcal{N}$, *satisfying*

$$X_t = \begin{cases} s_i & \text{if } t_i \leq t < t_{i+1} \text{ for all } i \geq 0 \\ \\ \Delta & \text{if } t \geq t^* \quad \text{if } t^* = \lim_{i \to +\infty} t_i \end{cases}$$

where Δ is an element such that $\Delta \notin \mathcal{S}$.

In order to simplify the notation, I set

$$\mathcal{S} = \{1, 2, 3, ...\}$$

$$p_{ij}(r,t) = P(X_t = j \mid X_r = i)$$

$$p_{ij}(h) = P(X_{t+h} = j \mid X_t = i) \quad \text{if the process is homogeneous.}$$

Let **X** be a homogeneous pure jump Markov process.

Definition : *the* **generator** *of* **X** *is the matrix* $Q = (q_{ij})$ *such that*

$$\lim_{h \searrow 0} \frac{p_{ij}(h) - I_i(j)}{h} = q_{ij} \qquad i, j \in \mathcal{S}$$

where $I_i(j) = 1$ *if* $i = j$ *and* $I_i(j) = 0$ *if* $i \neq j$.

For $i \neq j$, the element q_{ij} of the generator represents the infinitesimal transition rate from the state i to the state j. The generator Q is such that:

$$\begin{aligned} q_{ij} &\geq 0 && \text{per } i, j \in \mathcal{S},\ i \neq j \\ q_{ii} &= -\sum_{j \in \mathcal{S}, j \neq i} q_{ij} && \text{per } i \in \mathcal{S} \end{aligned} \tag{1.16}$$

The following proposition is an operative characterization of the described process.

Proposition : *let* Q *be a matrix having the (1.16) properties and let* $\pi(0) = \{\pi_i(0), i \in \mathcal{S}\}$ *be a fixed probability distribution; then, there exists a unique homogeneous pure jump Markov process having generator* Q *and initial distribution* $\pi(0)$.

1.4 A state-space model

Poisson, stress release and ETAS models express different and contrasting behaviours of the seismic activity; but each of them is considered as reasonable for earthquake sequences occurring in specific time-size ranges.
Moreover, the same region can show a different seismic behaviour when time elapses, so that the choice of a unique model can appear unsuitable to provide always a good fitting.
A solution can be found in a state-space model: the system is in different states - corresponding to different models - as time goes by, and the move from a state to another is guided by a hidden process that establishes in which state the system is at any time t.

Definition : *a* **state-space model** *is composed by two stochastic processes* $\mathbf{X} = \{X_t\}_{t\geq 0}$ *and* $\mathbf{Y} = \{Y_t\}_{t\geq 0}$, *called* **state process** *and* **observation process** *respectively. They are linked through the relations*

$$X_{t_n} \mid X_{t_{1:n-1}} \sim g(X_{t_n} \mid X_{t_{1:n-1}}) \tag{1.17}$$

$$Y_{t_n} \mid X_{t_n}, Y_{t_{1:n-1}} \sim f_{X_{t_n}}(Y_{t_n} \mid Y_{t_{1:n-1}}) \tag{1.18}$$

where $\{f_i : i \in S\}$ *is a family of conditional densities of* $\mathbf{Y}$, g *is a conditional density function of* $\mathbf{X}$ *and* $Y_{t_{1:n}}$ *denotes the observations* $Y_{t_1}, ..., Y_{t_n}$ *of* $\mathbf{Y}$.
Each relationship specifying (1.17) and (1.18) is called respectively **state equation** *and* **observation equation.**

A simple choice for the state equation is

$$X_{t_n} \mid X_{t_{1:n-1}} \sim g(X_{t_n} \mid X_{t_{n-1}})$$

that is, $\mathbf{X}$ is a Markov process; in this case, since $\mathbf{X}$ is an unobservable (hidden) Markov process, the state-space model is also called a **hidden Markov model** (HMM).

In the state-space model for earthquake analysis proposed in this thesis, I assume that:

(i) the state process $\mathbf{X}$ is a homogeneous pure jump Markov process such that

- $\mathcal{S} = \{1, 2, 3\}$ is the space of the states where 1, 2, 3 correspond to Poisson, stress release and Etas models respectively
- the generator is a 3×3 real matrix $Q = (q_{ij})_{i,j \in \mathcal{S}}$.

(ii) the observation process $\mathbf{Y} = (\mathbf{T}, \mathbf{m})$ is a marked point process where at time t the conditional intensity function depends on time also through the process X_t:

$$\lambda_t = \lambda(X_t, t)$$

Precisely:

- if $X_t = 1$, then the Poisson model is active and

$$\lambda(1, t) = \mu$$

- if $X_t = 2$, then the stress release model is active and

$$\lambda(2, t) = e^{\alpha + \beta(\rho t - R(t))}$$

- if $X_t = 3$, then the ETAS model is active and

$$\lambda(3, t) = \sum_{i: t_i < t} \frac{k\, e^{\gamma(m_i - M_0)}}{(t - t_i + c)^p}$$

(iii) denoting $\mathbf{N}$ the counting process associated with $\mathbf{Y}$, a reasonable assumption is that an earthquake does not occur in correspondence of a state change, that is $\mathbf{X}$ and $\mathbf{N}$ have no common jumps.

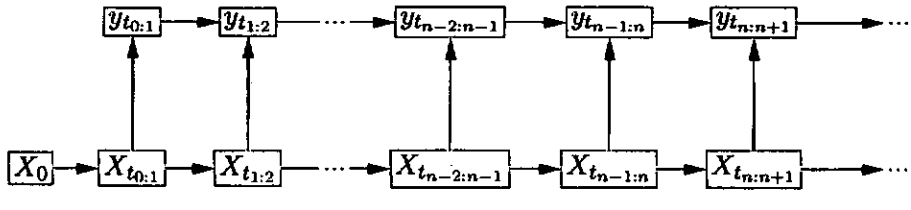

Figure 1.10: Graphical representation of the proposed state-space model structure; the variables of the state process are yellow and the observations are violet; arrows link dependent variables.

Let $\theta = (\theta_1, \theta_2, \theta_3, Q)$ denote all the parameters to estimate, where

$\theta_1 = \mu$	is the Poisson model parameter
$\theta_2 = (\alpha, \beta, \rho)$	are the SRM parameters
$\theta_3 = (k, \gamma, c, p)$	are the Etas model parameters
Q	is the infinitesimal generator of the homogeneous pure jump Markov process

and

$$\mu \in \mathcal{R}^+$$
$$\alpha \in \mathcal{R} \quad \beta \in \mathcal{R}^+ \quad \rho \in \mathcal{R}^+$$
$$k \in \mathcal{R}^+ \quad \gamma \in \mathcal{R}^+ \quad c \in \mathcal{R}^+ \quad p \in \mathcal{R}^+$$

and $Q = (q_{ij})_{i,j \in \mathcal{S}}$ is a 3×3 real matrix such that $q_{ij} \geq 0$ for $i \neq j$ and $\sum_{j \in \mathcal{S}} q_{ij} = 0$ for all $i \in \mathcal{S}$.

Figure 1.10 is a graphical representation of the proposed state-space model. It can be noted that the n^{th} observation Y_{t_n} depends on the current state X_{t_n}, as often assumed in the literature; since the observation process is a point process, a further dependence is introduced in the present model and must be considered during the analysis: for $t > t_n$, Y_t depends also on the past history $t_1, ..., t_n$ of the observed process.

In this context, two important problems arise : the estimation of the parameters and the filtering problem.

A filtering problem consists in estimating the hidden process by using information from the observable process; at each time t, a filtering solution provides an estimate of the probability that the hidden process visits a state, given the history of the observation process up to time t.

Chapter 2

Sequential estimation

2.1 The importance sampling principle

We are interested in

$$E_p(f) = \int f(x)p(dx)$$

where the expectation is with respect to a probability distribution having density $p(x)$ on a measurable space $(X, \mathcal{X})$ and $f(x)$ is an X-valued integrable function with respect to $p(x)$. There is often no analytic solution to that problem; an approximation $I(f)$ can be obtained by **perfect Monte Carlo integration** [DFG01]:

$$I(f) = \frac{1}{m}\sum_{i=1}^{m} f(x_i)$$

where $x_1, ..., x_m$ are independent samples drawn from $p(x)$. This estimator is unbiased and, if the variance of f is finite, that is $\sigma_f^2 = E_p(f^2) - E_p^2(f) < +\infty$, then the variance of $I(f)$ is σ_f^2/m. It follows that the strong law of large numbers and the central limit theorem hold :

$$I(f) \xrightarrow{a.s.} E_p(f) \quad \text{for } m \longrightarrow +\infty$$

$$\sqrt{m}[E_p(f) - I(f)] \xrightarrow{d} N(0, \sigma_f^2) \quad \text{for } m \longrightarrow +\infty$$

where $N(0, \sigma_f^2)$ is the normal distribution having zero mean and variance σ_f^2.

The advantage of the perfect Monte Carlo estimator is that its rate of convergence is independent of the dimension of the integrand, but sampling from $p(x)$ can be complex.

The **importance sampling principle** keeps this advantage and overcomes the sampling difficulties. In this method, an "easy-to-sample "distribution $\pi(x)$ on the measurable space $(X, \mathcal{X})$ is considered; $\pi(x)$ is called importance distribution. A different formulation of $E_p(f)$ is

$$E_p(f) = \int f(x)\omega(x)\pi(x)dx \tag{2.1}$$

where $\omega(x) = \frac{p(x)}{\pi(x)}$.
Let $x_1, ..., x_m$ be independent samples from $\pi(x)$ and let each x_i be associated with the following weight

$$\omega_i = \frac{p(x_i)}{\pi(x_i)}$$

then an approximation of (2.1) is given by

$$E_p(f) \approx \frac{1}{m}\sum_{i=1}^{m} f(x_i)\omega_i \tag{2.2}$$

Since the values of the weights are often known up to a normalizing constant, the above approximation is not practicable. To overcome this problem, one considers that

$$E_p(f) = \frac{\int f(x)\omega(x)\pi(x)dx}{\int \omega(x)\pi(x)dx} \tag{2.3}$$

Given an independent sample $x_1, ..., x_m$ from $\pi(x)$, unbiased estimates for numerator and denominator of (2.3) are respectively $\sum_{i=1}^{m} f(x_i)\omega_i$ and $\sum_{i=1}^{m} \omega_i$; setting

$$\tilde{\omega}_i = \frac{\omega_i}{\sum_{j=1}^{m} \omega_j}$$

it holds that $\sum_{i=1}^{m} \tilde{\omega}_i = 1$ and the ratio of the previous approximations provides the following estimation of (2.3)

$$J(f) = \sum_{i=1}^{m} f(x_i)\tilde{\omega}_i$$

In the literature, $x_1, ..., x_m$ are called **particles** and the associated weights $\tilde{\omega}_1, ..., \tilde{\omega}_m$ are called **importance weights**.
The set $\{x_i, \tilde{\omega}_i\}_{i=1,...,m}$, interpreted as a discrete probability distribution, can be

considered an approximation of the target distribution $p(x)$.
Precisely, since $J(f)$ is a ratio of two estimates, it is a biased estimate of $E_p(f)$; but under the weak assumptions [G89]

- the support of $\pi(x)$ includes the support of $p(x)$
- $E_p(f)$ exists and is finite

the strong law of large numbers holds

$$J(f) \xrightarrow{a.s.} E_p(f) \quad \text{for} \quad m \longrightarrow +\infty$$

Moreover, under the following further assumptions [G89]

- $E_p(\omega) < +\infty$
- $E_p(f^2\omega) < +\infty$

the central limit theorem holds

$$\sqrt{m}\,[E_p(f) - J(f)] \xrightarrow{d} N(0, \sigma^2)$$

$$m\,\hat{\sigma}_m^2 \xrightarrow{a.s.} \sigma^2$$

where

$$\sigma^2 = E_p[(f - E_p(f))^2\,\omega]$$

$$\hat{\sigma}_m^2 = \textstyle\sum_{i=1}^{m} [f(x_i) - J(f)]^2\,\tilde{\omega}_i^2$$

The previous conditions are generally satisfied when the importance distribution $\pi(x)$ is chosen so that its support contains the support of $p(x)$ and its tails don't decay more quickly than the tails of $p(x)$. If $\pi(x)$ is also close to $p(x)$, then faster convergence is observed.

2.2 Sequential importance sampling

Following the literature, I resume the theory about the sequential importance sampling assuming discrete-time processes [DFG01].

Let $(\mathbf{X}, \mathbf{Y}) = \{X_t, Y_t\}_{t \in \mathcal{N}}$ be a discrete-time state-space model, where $\mathbf{X}$ is a hidden Markov process and $\mathbf{Y}$ is an observed process.
Let $y_{1:t} = (y_1, y_2, ..., y_t)$ denote a realization of $\mathbf{Y}$ up to time $t \in \mathcal{N}$.
The state equation is represented by the initial distribution $p(X_0)$ of the hidden process $\mathbf{X}$ and its transition probability $p(X_t \mid X_{t-1})$; the observation equation is represented by $p(y_t \mid X_t)$.
For all $t \in \mathcal{N}$, we are interested in $E_{p(X_{0:t}|y_{1:t})}[f(X_{0:t})]$, where $p(X_{0:t} \mid y_{1:t})$ is the posterior distribution.
When a new observation becomes available, usually the posterior distribution needs to be newly and completely calculated. In a sequential structure, through the Bayes's theorem

$$p(x_{0:t} \mid y_{1:t}) = \frac{p(y_{1:t} \mid x_{0:t})p(x_{0:t})}{\int p(y_{1:t} \mid x_{0:t})p(x_{0:t})dx_{0:t}}$$

the posterior distribution at time t can be updated in order to include the new information at time $t+1$ and to produce the posterior distribution at time $t+1$

$$p(x_{0:t+1} \mid y_{1:t+1}) = p(x_{0:t} \mid y_{1:t}) \frac{p(y_{t+1} \mid x_{t+1})p(x_{t+1} \mid x_t)}{p(y_{t+1} \mid y_{1:t})}$$

where $p(x_{0:t})$ and $p(y_{1:t} \mid x_{0:t})$ denote respectively the prior distribution and the likelihood function.
The filtering distribution, as a marginal of the posterior, can also be calculated by a two-step iteration:

$$\textit{Prediction} \quad p(x_t \mid y_{1:t-1}) = \int p(x_t \mid x_{t-1})p(x_{t-1} \mid y_{1:t-1})dx_{t-1}$$

$$\textit{Filter} \quad p(x_t \mid y_{1:t}) = \frac{p(y_t \mid x_t)p(x_t \mid y_{1:t-1})}{\int p(y_t \mid x_t)p(x_t \mid y_{1:t-1})dx_t}$$

The calculation of the normalizing constant and of the integral parts is generally a very difficult problem and some approximation method is preferable, like a sequential importance sampling method.
Consequently, given an approximation of $E_{p(X_{0:t-1}|y_{1:t-1})}[f(X_{0:t-1})]$ and observed y_t at time t, a sequential analysis provides an approximation of $E_{p(X_{0:t}|y_{1:t})}[f(X_{0:t})]$ based on the approximation of $E_{p(X_{0:t-1}|y_{1:t-1})}[f(X_{0:t-1})]$, for all $t \in \mathcal{N}$.
A sequential importance sampling method carries out this task by exploiting the importance sampling principle; precisely at time $t-1$, by using an importance

distribution $\pi(X_{0:t-1} \mid y_{1:t-1})$, the following relation is available

$$E_{p(X_{0:t-1}|y_{1:t-1})}[f(X_{0:t-1})] \approx \sum_{i=1}^{m} \tilde{\omega}_{t-1}^{(i)} f(x_{0:t-1}^{(i)})$$

At time t, each particle $x_{0:t}^{(i)}$ is obtained by joining the old particle $x_{0:t-1}^{(i)}$ and a new sampled part $x_t^{(i)}$, so that the importance distribution at time $t-1$ is a marginal distribution of the importance distribution at time t; it means the following assumption

$$\pi(X_{0:t} \mid y_{1:t}) = \pi(X_t \mid X_{0:t-1}, y_{1:t})\pi(X_{0:t-1} \mid y_{1:t-1}) \tag{2.4}$$

that is equivalent to

$$\pi(X_{0:t} \mid y_{1:t}) = \pi(X_0) \prod_{s=1}^{t} \pi(X_s \mid X_{0:s-1}, y_{1:s})$$

Each particle $X_{0:t}^{(i)}$ is associated with a weight $\tilde{\omega}_t^{(i)}$, iteratively obtained by updating the weights at time $t-1$ in the following way:

$$\begin{aligned}
\tilde{\omega}_t &= \frac{p(X_{0:t} \mid y_{1:t})}{\pi(X_{0:t} \mid y_{1:t})} = \\
&\overset{(2.4)}{=} \frac{p(X_{0:t}, y_t \mid y_{1:t-1})}{\pi(X_t \mid X_{0:t-1}, y_{1:t})\pi(X_{0:t-1} \mid y_{1:t-1})p(y_t \mid y_{1:t-1})} = \\
&= \frac{p(X_t, y_t \mid X_{0:t-1}, y_{1:t-1})p(X_{0:t-1} \mid y_{1:t-1})}{\pi(X_{0:t-1} \mid y_{1:t-1})\pi(X_t \mid X_{0:t-1}, y_{1:t})p(y_t \mid y_{1:t-1})} = \\
&= \tilde{\omega}_{t-1} \frac{p(y_t \mid X_t)p(X_t \mid X_{t-1})}{\pi(X_t \mid X_{0:t-1}, y_{1:t})p(y_t \mid y_{1:t-1})}
\end{aligned}$$

2.2.1 Importance function choice

In some applications, the expression of the weights can be very complex to calculate. Then, the choice of the prior and importance distribution must aim at the tractability of the weights.

Moreover, the importance distribution must at least to satisfy some properties in order to guarantee a useful and suitable application of the importance sampling

principle; for example, as described in §2.1, its support must contain that of the target distribution, its tails must be heavier than those of the target distribution, etc...
Since fast convergence is observed when the importance function is close to the target distribution and since the prior distribution represents our belief and initial knowledge about the target distribution, a reasonable possibility is that prior distribution and importance distribution are equal. This choice is computationally desirable, because it leads to the following simple expression of the importance weights:

$$\tilde{\omega}_t = \tilde{\omega}_{t-1} \frac{p(y_t \mid X_t)}{p(y_t \mid y_{1:t-1})} \propto \tilde{\omega}_{t-1}\, p(y_t \mid X_t) \tag{2.5}$$

In particular, I use it to balance the high computational cost of the dynamic model applied to seismic sequences that I have proposed in this work.
Another reasonable choice is an importance distribution equal to the present posterior distribution, but the expression of the weights is generally so complicated as to require Markov chain Monte Carlo approximations; this is obviously more efficient, but computationally expensive.

2.2.2 Degeneracy of the particles

For the strong law of large numbers, an accurate approximation is obtained for a big "enough" number m of particles; in the applications it is noted that, as the number of iterations in the sequential method increases, the probability mass tends to concentrate only on few particles and the other particles have negligible weights: this phenomenon is called **degeneracy** of the weights. The reason of the degeneracy of the weights, proved in [KLW94], is that the variance of the weights is increasing as new data become progressively available. The following theorems show this property.

Theorem : $E_{\pi(X_{0:t}|y_{1:t})}(\tilde{\omega}_t) = 1$.
proof :
$E_{\pi(X_{0:t}|y_{1:t})}(\tilde{\omega}_t) = \int \tilde{\omega}_t \pi(x_{0:t} \mid y_{1:t}) dx_{0:t} = \int p(x_{0:t} \mid y_{1:t}) dx_{0:t} = 1.$ $\triangle$

Theorem : $\{\tilde{\omega}_t\}$ *is a martingale with respect to* $p(y_{t+1} \mid y_{1:t})$*; unconditional on* $y_{1:t}$*, their variances are increasing with respect to time* t.

proof : the sequence of the weights is a martingale because of

$$\begin{aligned} E_{p(y_{t+1}|y_{1:t})}(\tilde{\omega}_{t+1}) &= \int \tilde{\omega}_{t+1} p(y_{t+1} \mid y_{1:t}) dy_{t+1} = \\ &= \tilde{\omega}_t \int \frac{p(y_{t+1} \mid x_{t+1})}{p(y_{t+1} \mid y_{1:t})} p(y_{t+1} \mid y_{1:t}) dy_{t+1} = \\ &= \tilde{\omega}_t \int p(y_{t+1} \mid x_{t+1}) dy_{t+1} = \\ &= \tilde{\omega}_t \end{aligned}$$

Since $Var(\tilde{\omega}_t) = Var[E_{p(y_{t+1}|y_{1:t})}(\tilde{\omega}_{t+1})] \leq Var(\tilde{\omega}_{t+1})$, it is proved that $Var(\tilde{\omega}_t) \leq Var(\tilde{\omega}_{t+1})$ for all $t \in \mathcal{N}$. Δ

It is interesting to observe that, from the variance decomposition formula

$$\begin{aligned} Var(\tilde{\omega}_t) &= E[Var_{\pi(X_{0:t}|y_{1:t})}(\tilde{\omega}_t)] + Var[E_{\pi(X_{0:t}|y_{1:t})}(\tilde{\omega}_t)] = \\ &= E[Var_{\pi(X_{0:t}|y_{1:t})}(\tilde{\omega}_t)] + Var(1) = \\ &= E[Var_{\pi(X_{0:t}|y_{1:t})}(\tilde{\omega}_t)] \end{aligned}$$

and since $Var(\tilde{\omega}_t)$ is increasing, it follows that $Var_{\pi(X_{0:t}|y_{1:t})}(\tilde{\omega}_t)$ has an increasing, but not strictly increasing trend.

A rule of thumb allows to quantify the degeneracy phenomenon (see [L96]).
Suppose that, at a fixed time t, there are q particles having weight equal to $\tilde{\omega}^{(i)} = 1/q$ and $m - q$ particles having null weight. In order to find an approximation of q, we estimate the coefficient of variation of $\tilde{\omega}$:

$$c^2(\tilde{\omega}) = \frac{Var_\pi(\tilde{\omega})}{E_\pi^2(\tilde{\omega})}$$

where $\pi(X_{0:t} \mid y_{1:t})$. Observing that $c^2(\tilde{\omega}) = c^2(\omega)$, an empirical estimation of $E_\pi(\omega)$ is $\frac{1}{m}\sum_{j=1}^{m} \omega^{(j)}$. Analogously, an empirical estimate of $Var_\pi(\omega)$ is given by $\frac{1}{m}\sum_{j=1}^{m} \omega^{(j)2} - \frac{1}{m^2}\left(\sum_{j=1}^{m} \omega^{(j)}\right)^2$. By the ratio of the two previous approximations,

$$c^2(\tilde{\omega}) \approx m \sum_{j=1}^{m} \tilde{\omega}^{(j)2} - 1 = m\, q \frac{1}{q^2} - 1 = \frac{m}{q} - 1$$

and, inverting the relation, it results

$$q \approx \frac{m}{1 + c^2(\tilde{\omega})}$$

The value q represents an estimate of the number of particles having not close to zero weight; the estimation of q is denoted by ESS_t , **effective sample size** at time t.

Since $E_\pi(\tilde{\omega}) = 1$, then $c^2(\tilde{\omega}) = Var_\pi(\tilde{\omega})$ and

$$ESS_t = \frac{m}{1 + Var_\pi(\tilde{\omega})} \tag{2.6}$$

In order to obtain an estimation of ESS_t, further variables $\bar{\omega}^{(j)} = \omega^{(j)} \frac{m}{\sum_{i=1}^{m} \omega^{(i)}}$ are defined; their empirical means and variances are

$$\begin{aligned}
\textbf{empirical mean}(\bar{\omega}^{(j)}) &= \frac{1}{m} \sum_k \omega^{(k)} \frac{m}{\sum_i \omega^{(i)}} = 1 \\
\textbf{empirical variance}(\bar{\omega}^{(j)}) &= \frac{1}{m} \sum_k \left(\omega^{(k)} \frac{m}{\sum_i \omega^{(i)}} \right)^2 - \frac{1}{m} \left(\sum_k \omega^{(k)} \frac{m}{\sum_i \omega^{(i)}} \right)^2 = \\
&= \left(\frac{m}{\sum_i \omega^{(i)}} \right)^2 \cdot \textbf{empirical variance}(\omega^{(j)})
\end{aligned}$$

Since $c^2(\bar{\omega}) = c^2(\tilde{\omega})$, it is also true that

$$Var_\pi(\tilde{\omega}) \approx \left(\frac{m}{\sum_i \omega^{(i)}} \right)^2 \left[\frac{\sum_i \omega^{(i)2}}{m} - \left(\frac{\sum_i \omega^{(i)}}{m} \right)^2 \right] = m \cdot \sum_i \tilde{\omega}^{(i)2} - 1$$

Finally, by substituting $Var_\pi(\tilde{\omega})$ in (2.6), the following approximation of ESS_t is obtained:

$$ESS_t \approx \frac{1}{\sum_{j=1}^{m} \tilde{\omega}^{(j)2}}$$

In the light of the latest theorems, the weight variances are necessarily increasing in time; so, it is clear that ESS_t unavoidably decreases, revealing a loss of efficiency. It is worth observing that the above approximation of ESS_t has a globally, but not always regularly decreasing trend. This suggests that the choice of the number m of particles is guided by the observation of ESS_t's values in some explorative run.

2.2.3 Resampling step and Künsch-Hürzeler method

A particle having zero weight has null influence on the estimation and, consequently, a useless computational effort is employed for it.
Because of degeneracy of the particles, after some iterations of the sequential importance sampling method, the number of particles having null or almost null weight is considerable.
Then a resampling step of the particle set is introduced in order to improve the computational efficiency.
Let us assume that, at time t, the set of particles $\{x_i, \tilde{\omega}_t^{(i)}\}_{i=1,\dots,m}$ is an approximation of the posterior distribution $p(X_{0:t} \mid y_{1:t})$; the resampling step provides a new approximating distribution $\{\bar{x}_i, \frac{1}{m}\}_{i=1,\dots,m}$ obtained by sampling with replacement from the distribution $\{x_i, \tilde{\omega}_t^{(i)}\}_{i=1,\dots,m}$.
The resampling step does not change the essence of the method, even if it adds some randomness: indeed, the estimators based on the resampled particle set are less efficient than that based on the no resampled particle set.
Moreover, the resampling step causes an **impoverishment effect**: the set of distinct particles becomes more and more exiguous after each sequential step. Despite these disadvantages, the use of some resampling steps can be suitable: it optimizes the computational efforts reducing the number of null weight particles and, consequently, it focuses more attention on the particles with "big" weights, following the idea that these particles carry more information on the posterior distribution [LC95].
In order to improve the efficiency of the method, a balance between impoverishment effect and degeneracy of the particles is needed; these reasons suggest that resampling steps must be applied sometimes and not at every sequential iteration t.
A resampling step can be applied at some deterministic times t; otherwise, a dynamic criterion to establish the suitable times t for a resampling step can be based on the ESS_t values: for example, a resampling step is executed at time t if ESS_t is less than a fixed threshold $k \cdot m$, for $0 \leq k \leq 1$, that is only when there is a noteworthy number of particles with irrelevant weight.
Many resampling methods are proposed in the literature and I refer to [DFG01], [LC98] for a complete overview on this topic.
The simplest resampling step is based on a sampling from a multinomial distribution of parameters $\tilde{\omega}_1, \tilde{\omega}_2, \dots, \tilde{\omega}_m$: it takes computational time $O(m \log m)$. Other cheaper resampling methods, having cost $O(m)$, are generally applied and indifferently used. For example, the *residual sampling*, proposed in [LC98], has the following scheme: let $\{(x_t^{(i)}, \tilde{\omega}_t^{(i)})\}_{i=1,\dots,m}$ be the set of particle,

- set $k_i = m\omega_t^{(i)}$ for all $i = 1, ..., m$
- the resampled set $\{\bar{x}_t^{(i)}\}$ is obtained by retaining $[k_i]$ (the integer part of k_i) copies of $x_t^{(i)}$ for each $i = 1, ..., m$
- set $k = m - [k_1] - .... - [k_m]$
- draw the remaining k particles from the distribution $\{(x_t^{(i)}, k_i - [k_i]\}_{i=1,...,m}$
- reset all the weights to m^{-1}.

In particular, among the resampling methods having cost $O(m)$, I have chosen and applied the following, proposed in [CCF99] and based on a method of simulating order statistics:

Algorithm 1

1. $T_1 = t_1 \sim$ Exponential with mean 1
2. for $i = 2, ..., m$

   $t_i \sim$ Exponential with mean 1

   $T_i = T_{i-1} + t_i$
3. $i = 0$, $j = 1$
4. do while $i < m$

   if $\tilde{\omega}_j T_m > T_i$ then

   $i = i + 1$

   set $\bar{x}_t^{(i)} = x_t^{(j)}$

   else

   $j = j + 1$

   end if

   end do

In order to combat the particle impoverishment, some techniques aimed to rejuvenate the particle set are proposed in the literature. Some of them are based on a resampling step from a kernel density estimate of the posterior distribution as in [GSS93].

Künsch and Hürzeler [DFG01] propose to introduce a small local noise on each particle after the resampling step:

$$\bar{x}_j = x_s + v_j \tag{2.7}$$

Since the noise is negligible, the weights are approximately the same and are essentially unchanged. The noise v_j is chosen to have a decreasing variance, so as to keep, each time, a little more information of the previous observations; generally, it is set $Var(v_j) = a/t^2$, where a is a suitable fixed constant.

2.3 Self-organizing state-space model

In the previous section, the parameters of the state-space model are considered as known; in general to fit the model to a real data set, they have to be estimated.
A sequential solution for the estimation of the parameters is a discussed problem in the literature.
The Kalman filter and its extensions provide a solution for Gaussian linear models; but the state-space model proposed in this work is nonlinear and non-Gaussian.
There are in the literature two important classes of techniques for parameter estimation: the former is based on the Rao-Blackwellization procedure [KLW94], [CR96], [S02], while the latter was proposed by Kitagawa [K98].
Denoting by θ the vector of the parameters, we are interested in

$$I(f) = E_{p(\theta, x_{1:t}|y_{1:t})}(f(\theta, x_{1:t}))$$

Let us consider the following two reformulations of $I(f)$

$$\begin{aligned} I(f) &= E_{p(x_{1:t}|y_{1:t})}[E_{p(\theta|x_{1:t}, y_{1:t})}(f(\theta, x_{1:t}))] \\ I(f) &= E_{p(\theta|y_{1:t})}[E_{p(x_{1:t}|\theta, y_{1:t})}(f(\theta, x_{1:t}))] \end{aligned}$$

The Rao-Blackwellization procedure exploits the formulation that allows to solve explicitely the inner expectation on the basis of the specific expression of f. This is not the case of my application.
In the technique proposed by Kitagawa, the state-space model is reorganized, so that the unknown model parameters are included in the hidden state vector; then a Bayesian sequential Monte Carlo method is applied to solve simultaneously the fil-

tering and the parameter estimation problems: this is the self-organizing state-space model and I can apply it to my problem; its structure is illustrated in figure (2.2).

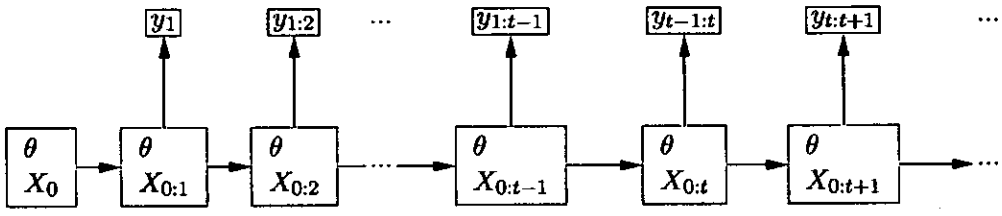

Figure 2.1: Self-organizing hidden Markov model structure

At each time t, a vector $(\theta, x_{1:t})$ represents a particle; the set of the weighted particles constitutes a discrete approximation of the distribution $p(\theta, x_{1:t} \mid y_{1:t})$.

2.4 Sequential solution of the problem

The state-space model I have proposed is rewritten as a self-organizing state-space model, taking into account that each observation depends not only on the current state, but also on the past observed history.
Moreover, since the previously described sequential theory is applicable to discrete-time models, a time-discretization of the proposed time-continuous model is necessary in order to carry out a Bayesian sequential analysis.
If I choose a thin partition of the observed time interval $[0, T]$, the impoverishment effect due to the sequential method is already evident during the analysis of the early part of $[0, T]$; consequently, the estimates obtained by studying the last part of $[0, T]$ are based on an insufficient exploration of the parameter and state space.
It seems reasonable to assign $t_0, t_1, ..., t_n, t_{n+1}$ as partition of $[0, T]$, where $t_1, ..., t_n$ correspond to the point process observations and $t_0 = 0$, $t_{n+1} = T$.

The corresponding self-organizing state-space model has augmented state vector $(X_{t_0:t_i}, \theta)$, where θ is time-invariant and $X_{t_0:t_i}$ is a finite random sequence

$$(\tau_1, s_1),\ (\tau_2, s_2),\ ...,\ (\tau_v, s_v)$$

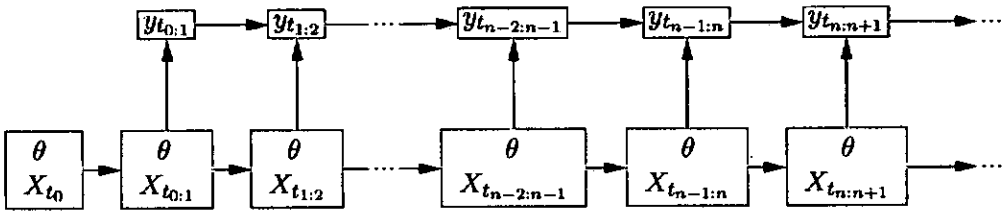

Figure 2.2: Self-organizing state-space structure for the proposed model

such that τ_i denotes the i^{th} jump time of the pure jump Markov process in $[t_0, t_i]$ and s_i the corresponding visited state (for simplicity, I denote $\tau_0 = t_0$ and $\tau_{v+1} = t_i$).
After a review of the expression (2.5) in the light of the proposed state-space model, the iterative expression of the weight $\tilde{\omega}_{t_i}$ associated to the time interval $(t_{i-1}, t_i]$ is given by

$$\tilde{\omega}_{t_i} \propto \tilde{\omega}_{t_{i-1}} p(y_{t_{i-1:i}} \mid X_{t_{i-1:i}}, y_{t_{0:i-1}})$$

where the likelihood function in $(t_{i-1}, t_i]$ is given by

$$p(y_{t_{i-1:i}} \mid X_{t_{i-1:i}}, y_{t_{0:i-1}}) = \lambda(s_{v+1}, t_i) \cdot exp\left\{ -\sum_{j=0}^{v} \int_{\tau_j}^{\tau_{j+1}} \lambda(s_j, t)\, dt \right\}$$

It can be noted that the dependence of the likelihood in $(t_{i-1}, t_i]$ on the past observed history $y_{t_{0:i-1}}$ is expressed through the conditional intensity functions $\lambda_t(\cdot)$ of the marked point processes.
The following *Algorithm 2* is a scheme of the sequential importance sampling method that I apply in the present work. In particular, I recall that I choose prior and importance distribution for parameters and states to be equal.
This algorithm can be usefully employed in many ways. Fixing a very large number m of particles, the goodness of the algorithm is guaranteed, but the computational time can be remarkable.
An alternative is to fix a reasonable number m of particle and to obtain the parameter estimates as means of the estimated parameters on many runs of the algorithm; then, fixed the parameters equal to their estimated values, the filtering distribution of the states can be obtained through a further run of the algorithm.
Otherwise, it is possible to run the algorithm a first time using the noising method of Künsch and Hürzeler and then to run it again without the Künsch and Hürzeler method using the estimated posterior distributions as prior distributions.

Algorithm 2

1. fix the prior distributions: $p(\theta)$ and $p(x)$
2. sample an initial particle set from the prior distributions:

$$(x_{t_0}^{(j)}, \theta^{(j)}) \qquad \text{for } j = 1, ..., m$$

and set the corresponding weights $\omega_{t_0}^{(j)}$ equal to $\frac{1}{m}$;

3. for i = 1, ..., n
   (a) sample from the importance distributions the updating of the particles and obtain the new particle set at time t_i

$$(x_{t_{0:i}}^{(j)}, \theta^{(j)}) \qquad \text{for } j = 1, ..., m$$

   (b) update the weights

$$\tilde{\omega}_{t_i} \propto \tilde{\omega}_{t_{i-1}} p(y_{t_{i-1:i}} \mid x_{t_{i-1:i}}, y_{t_{0:i-1}})$$

   (c) apply a resampling step following *Algorithm 1* if the effective sample size ESS_t is "small" and, in case, add a noise to the parameter values following the Künsch-Hürzeler method.

Chapter 3

Filtering problem

The present chapter will be concerned with the filtering of a pure jump Markov process with point process observations, the considered state-space model, under the assumption that all parameters are known.

For each $j \in \mathcal{S}$, let us assume

$$Z_t(j) = I(X_t = j) = \begin{cases} 0 & \text{if } X_t \neq j \\ 1 & \text{if } X_t = j \end{cases}$$

The distribution of interest is

$$\hat{Z}_t(j) = E[Z_t(j) \mid \mathcal{H}_t] = P(X_t = j \mid \mathcal{H}_t) \qquad \forall\, j \in \mathcal{S}$$

where

$\mathcal{H}_t$ is the internal history of the observed process $\mathbf{Y}$

$\mathcal{G}_t$ is the internal history of the hidden process $\mathbf{X}$

$\mathcal{F}_t$ is the internal history of the global process $(\mathbf{X}, \mathbf{Y})$.

3.1 Filtering equations

The innovation method provides the following filtering equations ([B81] pp. 85-94, [K91] pp. 194-200), the roots of which are the target distribution.

For each $j \in S$, $Z_t(j)$ admits the semimartingale representation

$$Z_t(j) = Z_0(j) + \int_0^t f_s(j)\, ds + M_t(j) \qquad \forall\, j \in \mathcal{S} \tag{3.1}$$

where $M_t(j)$ is a zero mean $\mathcal{F}_t$-martingale and $f_t(j)$ is an $\mathcal{F}_t$-progressive process such that, for all $t \geq 0$, $E(\int_0^t \mid f_s(j) \mid ds) < +\infty$ P-almost surely.
It can be shown that, for a homogeneous pure jump Markov process, it holds

$$f_s(j) = \sum_{i \in \mathcal{S}} q_{ij} Z_s(i) \qquad \forall\, j \in \mathcal{S}$$

where $Q = (q_{ij})_{i,j \in \mathcal{S}}$ is the infinitesimal generator of $\mathbf{X}$ ([B81], p.98).

Then (3.1) represents the state equation of the considered state-space model. Projecting the state equation on the observed history $\mathcal{H}_t$, one obtains the filtering equation

$$\hat{Z}_t(j) = \hat{Z}_0(j) + \int_0^t \hat{f}_s(j) ds + \hat{M}_t(j) \tag{3.2}$$

where

$$\hat{f}_s(j) = \sum_{i \in \mathcal{S}} q_{ij} \hat{Z}_s(i)$$

$$\hat{M}_t(j) = E[M_t(j) \mid \mathcal{H}_t]$$

and $\hat{M}_t(j)$ is a zero mean $\mathcal{H}_t$-martingale and $\hat{f}_t$(j) is a $\mathcal{H}_t$-progressive process such that, for all nonnegative bounded $\mathcal{H}_t$-progressive processes $\{C_t\}_{t \geq 0}$, it holds

$$E\left[\int_0^t C_s f_s(j) ds\right] = E\left[\int_0^t C_s \hat{f}_s(j) ds\right]$$

For the integral representation theorem of point process martingales (1.1), there exists a unique process $\{\Phi_s(j)\}_{s \in \mathcal{R}+}$ such that

$$\hat{M}_t(j) = \int_0^t \Phi_s(j)(dN_s - \hat{\lambda}_s ds) \tag{3.3}$$

where $\hat{\lambda}_t = E(\lambda_t \mid \mathcal{H}_t)$ and $\mathbf{N}$ is the counting process associated with the marked point process $\mathbf{Y}$. Note that $(dN_s - \hat{\lambda}_s ds)$ corresponds to the zero mean martingale in the Doob-Meyer decomposition for a marked point process.
Generally, $\Phi_s(j)$ is called the **$\mathcal{H}_t$-innovation gain** and $(dN_s - \hat{\lambda}_s ds)$ the **$\mathcal{H}_t$-innovation**

process. An involved proof leads to the following general expression for the innovation gain

$$\Phi_s(j) = \Phi_{1,s}(j) - \Phi_{2,s}(j) + \Phi_{3,s}(j)$$

where $\Phi_{1,s}(j)$, $\Phi_{2,s}(j)$, $\Phi_{3,s}(j)$ are $\mathcal{H}_t$-predictable processes such that

$$E\left[\int_0^t C_s \Phi_{1,s}(j)\hat{\lambda}_s ds\right] = E\left[\int_0^t C_s Z_s(j)\lambda(X_s, s)ds\right]$$

$$E\left[\int_0^t C_s \Phi_{2,s}(j)\hat{\lambda}_s ds\right] = E\left[\int_0^t C_s Z_s(j)\hat{\lambda}_s ds\right]$$

$$E\left[\int_0^t C_s \Phi_{3,s}(j)\hat{\lambda}_s ds\right] = E\left[\sum_{0<s\leq t} C_s \Delta M_s dN_s\right]$$

I will give the version of those equations for the considered state-space process.
It is useful to remark that, for all $j \in \mathcal{S}$, the conditional intensity functions of the models are by definition left continuous, that is $\lambda(j, s) = \lambda(j, s^-)$; and also the state probabilities are such that $\hat{Z}_s(j) = \hat{Z}_{s-}(j)$.

Since $E\left[\int_0^t C_s Z_s(j)\hat{\lambda}_s ds\right] = E\left[\int_0^t C_s \hat{Z}_s(j)\hat{\lambda}_s ds\right] = E\left[\int_0^t C_s \hat{Z}_{s-}(j)\hat{\lambda}_s ds\right]$, it is always true that

$$\Phi_{2,s}(j) = \hat{Z}_{s-}(j)$$

Observing that

$$\hat{\lambda}_s = E(\lambda_t \mid \mathcal{H}_t) = \sum_{i\in\mathcal{S}} \lambda(i, s^-)\hat{Z}_{s-}(i) = \sum_{i\in\mathcal{S}} \lambda(i, s)\hat{Z}_{s-}(i)$$

$$\lambda(X_s, s)Z_s(j) = \lambda(j, s)Z_s(j)$$

it holds

$$\Phi_{1,s}(j) = \frac{\lambda(j, s)\hat{Z}_{s-}(j)}{\sum_{i\in\mathcal{S}} \lambda(i, s)\hat{Z}_{s-}(i)}$$

where the denominator is strictly positive, since $\hat{Z}_{s-}(i)$ for $i \in \mathcal{S}$ is a discrete probability distribution and $\lambda(i, s)$ is by definition strictly positive for each $i \in \mathcal{S}$.

Remembering that the considered processes have no common jumps, $ds = 0$ when

$dN_s \neq 0$ and vice versa; then, obviously, it results

$$\Phi_{3,s}(j) = 0$$

Summarizing, for all $j \in \mathcal{S}$, (3.3) becomes

$$\hat{M}_t = \int_0^t \left(-\hat{Z}_{s^-}(j) + \frac{\lambda(j,s)\hat{Z}_{s^-}(j)}{\sum_{i\in\mathcal{S}} \lambda(i,s)\hat{Z}_{s^-}(i)} \right) \left(dN_s - \sum_{i\in\mathcal{S}} \lambda(i,s)\hat{Z}_{s^-}(i)\, ds \right)$$

and, by substitution in (3.2) of the previous relations, the filtering equation becomes

$$\begin{aligned} \hat{Z}_t(j) &= \hat{Z}_0(j) + \sum_{i\in\mathcal{S}} q_{ij} \int_0^t \hat{Z}_s(i) ds + \\ &+ \int_0^t \left(-\hat{Z}_{s^-}(j) + \frac{\lambda(j,s)\hat{Z}_{s^-}(j)}{\sum_{i\in\mathcal{S}} \lambda(i,s)\hat{Z}_{s^-}(i)} \right) \left(dN_s - \sum_{i\in\mathcal{S}} \lambda(i,s)\hat{Z}_{s^-}(i) ds \right) \end{aligned}$$

and, in differential form,

$$\begin{aligned} d\hat{Z}_t(j) &= \sum_{i\in\mathcal{S}} q_{ij}\hat{Z}_t(i)dt - \hat{Z}_{t^-}(j)dN_t + \\ &+ \frac{\lambda(j,t)\hat{Z}_{t^-}(j)}{\sum_{i\in\mathcal{S}} \lambda(i,t)\hat{Z}_{t^-}(i)} dN_t + \hat{Z}_{t^-}(j) \sum_{i\in\mathcal{S}} \lambda(i,t)\hat{Z}_{t^-}(i)dt - \\ &- \lambda(j,t)\hat{Z}_{t^-}(j)dt \end{aligned} \tag{3.4}$$

The resulting filtering equation belongs to the family of the Kushner-Stratonovich equations.

3.2 Solution of filtering equations

For each $j \in \mathcal{S}$, an iterative solution of (3.4) is proposed following [KKM90].

Remembering that $\mathbf{X}$ and $\mathbf{N}$ have no common jumps, it is $dN_t = 0$ when $t \in (t_n, t_{n+1})$, for all $n \in \mathcal{N}$. Then, by integration of (3.4) on (t_n, t) for $t < t_{n+1}$, it follows that

$$\hat{Z}_t(j) = \hat{Z}_{t_n}(j) + \sum_{i\in\mathcal{S}} q_{ij} \int_{t_n}^{t} \hat{Z}_s(i)ds + \int_{t_n}^{t} \hat{Z}_{s^-}(j) \sum_{i\in\mathcal{S}} \lambda(i,s)\hat{Z}_{s^-}(i)ds +$$
$$- \int_{t_n}^{t} \lambda(j,s)\hat{Z}_{s^-}(j)ds$$

For the same reason, $dt = 0$ when $t \in (t_{n+1}^-, t_{n+1}]$ and, by integration of (3.4) on $(t_{n+1}^-, t_{n+1}]$, it follows that

$$\hat{Z}_{t_{n+1}}(j) = \hat{Z}_{t_{n+1}^-}(j) - \hat{Z}_{t_{n+1}^-}(j) + \frac{\lambda(j,t_{n+1})\hat{Z}_{t_{n+1}^-}(j)}{\sum_{i\in\mathcal{S}} \lambda(i,t_{n+1})\hat{Z}_{t_{n+1}^-}(i)} =$$
$$= \frac{\lambda(j,t_{n+1})\hat{Z}_{t_{n+1}^-}(j)}{\sum_{i\in\mathcal{S}} \lambda(i,t_{n+1})\hat{Z}_{t_{n+1}^-}(i)}$$

Hence, the filtering equations are equivalent to the following iterative system

$$\begin{cases} \hat{Z}_0(j) = P(X_0 = j) \\ \hat{Z}_t(j) = \hat{Z}_{t_n}(j) + \sum_{i\in\mathcal{S}} q_{ij} \int_{t_n}^{t} \hat{Z}_s(i)ds + \int_{t_n}^{t} \hat{Z}_{s^-}(j) \sum_{i\in\mathcal{S}} \lambda(i,s)\hat{Z}_{s^-}(i)ds + \\ \quad - \int_{t_n}^{t} \lambda(j,s)\hat{Z}_{s^-}(j)ds \qquad \text{for} \quad t \in (t_n, t_{n+1}) \\ \hat{Z}_{t_{n+1}}(j) = \dfrac{\lambda(j,t_{n+1})\hat{Z}_{t_{n+1}^-}(j)}{\sum_{i\in\mathcal{S}} \lambda(i,t_{n+1})\hat{Z}_{t_{n+1}^-}(i)} \end{cases} \tag{3.5}$$

Following [KKM90], the conditional probability $\hat{Z}_t(j)$ can be written as the ratio

$$\hat{Z}_t(j) = \frac{\varphi_t(j)}{\sum_{i\in\mathcal{S}} \varphi_t(i)}$$

where $\varphi_t(j)$ is the solution of the linear system

$$\begin{cases} \varphi_0(j) = P(X_0 = j) \\ \varphi_t(j) = \varphi_{t_n}(j) + \sum_{i \in \mathcal{S}} q_{ij} \int_{t_n}^{t} \varphi_s(i) ds + \\ \qquad - \int_{t_n}^{t} \lambda(j,s) \varphi_{s^-}(j) ds \quad \text{for} \quad t \in (t_n, t_{n+1}) \\ \varphi_{t_{n+1}}(j) = \lambda(j, t_{n+1}) \varphi_{t_{n+1}^-}(j) \end{cases} \tag{3.6}$$

The system (3.6) is obtained by deleting nonlinear terms in (3.5).
Note that the conditional probabilities $\hat{Z}_t(j)$ are obtained by normalization of $\varphi_t(j)$. Then, the problem reduces to solve (3.6).
Through the Feynman-Kac formula ([L98] [B99]), it is possible to show that a solution of (3.6) is the process $\varphi(j) = \{\varphi_t(j)\}$ such that, for all $t \in (t_n, t_{n+1})$,

$$\varphi_t(j) = E^{(n)} \left[I(X_t^{(n)} = j) \; exp \left\{ - \int_{t_n}^{t} \lambda(X_s^{(n)}, s) \; ds \right\} \right] \tag{3.7}$$

where $\{X_t^{(n)}\}_{t \in (t_n, \infty)}$ is a homogeneous pure jump Markov process with respect to a suitable probability space $(\Omega^{(n)}, \mathcal{F}^{(n)}, P^{(n)})$; it has the same infinitesimal generator Q of $\{X_t\}$ on the interval (t_n, t_{n+1}) and initial distribution $\left\{ \varphi_{t_n}(j) / \sum_{i \in \mathcal{S}} \varphi_{t_n}(i) \right\}_{j=1,\dots,K}$.
Finally, $E^{(n)}$ denotes the expectation with respect to $(\Omega^{(n)}, \mathcal{F}^{(n)}, P^{(n)})$.

Then, (3.6) becomes

$$\begin{cases} \varphi_0(j) = P(X_0 = j) \\ \varphi_t(j) = E^{(n)} \left[I(X_t^{(n)} = j) \; exp \left\{ - \int_{t_n}^{t} \sum_{i \in \mathcal{S}} \lambda(i,s) I(X_s^{(n)} = i) ds \right\} \right] \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{for } t \in (t_n, t_{n+1}) \\ \varphi_{t_{n+1}}(j) = \lambda(j, t_{n+1}) \varphi_{t_{n+1}^-}(j) \end{cases} \tag{3.8}$$

This suggests that the Monte Carlo method can be applied to obtain an approximation of the expectation in (3.8). Precisely, let $\mathcal{D}_{t_n,t}$ be a countable set of simulations from the process $\{X_t^{(n)}\}$ in the interval (t_n, t), for $t_n < t < t_{n+1}$:

$$\mathcal{D}_{t_n,t} = \left\{ \quad d_u = \{(\tau_0,\sigma_0),(\tau_1,\sigma_1),...,(\tau_{i_u},\sigma_{i_u}),(\tau_{i_u+1},\sigma_{i_u+1})\} : \quad u = 1,...,D \quad \right\}$$

where $D = card(\mathcal{D}_{t_n,t})$ denotes the fixed cardinality of the set $\mathcal{D}_{t_n,t}$; an element d_u of $\mathcal{D}_{t_n,t}$ is a sequence of couples (τ_k, σ_k), for $k = 0,...,i_u+1$, where τ_k is a jump time of the homogeneous pure jump Markov process $\{X_t^{(n)}\}$ in (t_n, t) such that

$$t_n = \tau_0 < \tau_1 < ... < \tau_{i_u} < \tau_{i_u+1} = t$$

and σ_k is the corresponding visited state, that is

$$X_{\tau_k}^{(n)} = \sigma_k \quad \text{for} \quad \sigma_k \in \mathcal{S} \ , \ \ k = 0,...,i_u+1$$

Each element of $\mathcal{D}_{t_n,t}$ is obtained by following the *Algorithm 4* for the simulation of a homogeneous pure jump Markov process (section 4.1) with infinitesimal generator Q and initial distribution $\left\{\varphi_{t_n}(j)/\sum_{i\in\mathcal{S}}\varphi_{t_n}(i)\right\}_{j=1,...,K}$.
Given $d_u \in \mathcal{D}_{t_n,t}$, the integral in (3.8) holds

$$\begin{aligned}\int_{t_n}^{t}\sum_{i\in\mathcal{S}}\lambda(i,s)I(X_s^{(n)}=i)\,ds &= \sum_{k=0}^{i_u}\int_{\tau_k}^{\tau_{k+1}}\sum_{i\in\mathcal{S}}\lambda(i,s)I(X_s^{(n)}=i)\,ds = \\ &= \sum_{k=0}^{i_u}\int_{\tau_k}^{\tau_{k+1}}\lambda(\sigma_k,s)\,ds\end{aligned}$$

Then (3.8) is approximated by

$$\begin{cases} \varphi_0(j) = P(X_0 = j) \\ \\ \varphi_t(j) = card(\mathcal{D}_{t_n,t})^{-1}\displaystyle\sum_{d_u\in\mathcal{D}_{t_n,t}} I_j(\sigma_{i_u+1})\,exp\left\{-\sum_{k=0}^{i_u}\int_{\tau_k}^{\tau_{k+1}}\lambda(\sigma_k,s)\,ds\right\} \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{for } t\in(t_n,t_{n+1}) \\ \\ \varphi_{t_{n+1}}(j) = \lambda(j,t_{n+1})\varphi_{t_{n+1}^-}(j) \end{cases} \tag{3.9}$$

When the parameters are unknown, they can be estimated by using the sequential importance sampling method of the *Algorithm 2* in section 2.4 and then it is possible to obtain a better approximation of the filtering distribution by exploiting (3.8). *Algorithm 3* is a scheme procedure for (3.8).

Algorithm 3

1. load the data set $\{y_i = (t_i, m_i)\}_{i=0,...,n+1}$
2. read the model parameters θ and the infinitesimal generator Q
3. fix a "sufficiently large" integer value D (representing $card(\mathcal{D}_{t_n,t})$)
4. set $\hat{Z}_{t_0}(1) = \hat{Z}_{t_0}(2) = \hat{Z}_{t_0}(3) = 1/3$
5. for $i = 0, ..., n$
   (a) set $\varphi_{t_{i+1}^-}(1) = \varphi_{t_{i+1}^-}(2) = \varphi_{t_{i+1}^-}(3) = 0$
   (b) produce a number D of simulations on (t_i, t_{i+1}) from the homogeneous pure jump Markov process having infinitesimal generator Q and initial distribution $\{\hat{Z}_{t_i}(1), \hat{Z}_{t_i}(2), \hat{Z}_{t_i}(3)\}$ in order to form the set

$$\mathcal{D}_{t_i,t_{i+1}} = \left\{ \quad d_u = \{(\tau_0, \sigma_0), (\tau_1, \sigma_1), ..., (\tau_{i_u}, \sigma_{i_u}), (\tau_{i_u+1}, \sigma_{i_u+1})\} \quad \right\}_{u=1,...,D}$$

   where $\tau_0 = t_i$, $\tau_{i_u+1} = t_{i+1}$ and $\sigma_{i_u+1} = \sigma_{i_u}$.

   (c) for $j = 1, 2, 3,$

$$\varphi_{t_{i+1}^-}(j) = D^{-1} \cdot \sum_{d_u \in \mathcal{D}_{t_i,t_{i+1}}} I_j(\sigma_{i_u+1}) \cdot exp\left\{ -\sum_{k=0}^{i_u} \int_{\tau_k}^{\tau_{k+1}} \lambda(\sigma_k, s)\, ds \right\}$$

   (d) for $j = 1, 2, 3,$ $\varphi_{t_{i+1}}(j) = \dfrac{\varphi_{t_{i+1}^-}(j) \cdot \lambda(j, t_{i+1})}{\sum_{k=1}^{3} \varphi_{t_{i+1}}(k) \cdot \lambda(k, t_{i+1})}$

Chapter 4

Simulation, analysis and results

In this chapter, I generate a data set according to the proposed state-space model and I apply to this set some of the sequential methods for estimation presented in the previous chapters.
The first part of this chapter concerns the theory and the algorithms for the simulation of the considered state-space model. The first step consists in getting a realization

$$\{x_{\tau_v} : v = 1, ..., V\}$$

of the hidden process $\mathbf{X}$ on time interval $[0, T]$.
For convenience, let $x_{\tau_0} = 0$ and $x_{\tau_{V+1}} = T$.
Then, for each sub-interval $(\tau_v, \tau_{v+1}]$, one has to produce a realization

$$\{(t_i, m_i) : i = 1, ..., n_{\tau_v}\}$$

of the marked point process with conditional intensity function $\lambda(x_{\tau_v}, \cdot)$.

4.1 Simulation of a pure jump Markov process

Let $\mathbf{X}$ be a homogeneous pure jump Markov process on the state space $\mathcal{S} = \{1, 2, 3\}$, having generator Q and initial distribution $(\hat{Z}_0(1), \hat{Z}_0(2), \hat{Z}_0(3))$.

Let the **holding time process** of $\mathbf{X}$ be the process $\mathbf{H} = \{H_v\}_{v \in \mathcal{N}}$ of the time between subsequent jumps :

$$H_0 = min\{t \geq 0 : X_t \neq X_0\}$$

$$H_v = min\{t \geq 0 : X_{\tau_v - 1 + t} \neq X_{\tau_v - 1}\} \quad \forall v \in \mathcal{N}$$

and $\tau_v = H_0 + ... + H_{v-1}$.

Let the **jump chain** of the process $\mathbf{X}$ be the sequence $\mathbf{J} = \{J_v\}_{v \in \mathcal{N}}$ of the subsequently visited states of $\mathbf{X}$:

$$J_0 = X_0$$

$$J_v = X_{\tau_v - 1} \quad \forall\, v \in \mathcal{N}$$

Proposition [GS90]:

1. *the jump chain* $\mathbf{J}$ *is a homogeneous discrete time Markov chain such that its transition probabilities are*

$$p_{ij}^{(J)} = \begin{cases} -\frac{q_{ij}}{q_{ii}} & per\ i \neq j \\ 0 & per\ i = j \end{cases}$$

2. *given* X_0, J_1 *is conditionally independent from* H_0

3. *given* $J_0 = j_0, ..., J_n = j_n$, *the random variables* $H_0, ..., H_n$ *are conditionally independent; moreover, for all* $i = 0, ..., n$, *the conditional distribution of* H_i *is exponential with parameter* $-q_{j_i j_i}$.

The previous proposition defines a simulation procedure of a homogeneous pure jump Markov process; precisely, if a change of state occurred at time t, the time up to next change is drawn from the holding time distribution and the new state is generated according to the transition probabilities.

Algorithm 4 : homogeneous pure jump Markov process simulation

1. Assign the initial distribution $(\hat{Z}_0(1), \hat{Z}_0(2), \hat{Z}_0(3))$ and the generator matrix $Q(\cdot, \cdot)$

2. set $v = 0$, $\tau_v = 0$

3. let J_v be a sample from $(\hat{Z}_0(1), \hat{Z}_0(2), \hat{Z}_0(3))$

4. while $\tau_v < T$:

   (a) sample H_v from an exponential distribution with mean $-\dfrac{1}{Q(J_v, J_v)}$

   (b) $\tau_{v+1} = \tau_v + H_v$

   (c) let J_{v+1} be a sample from the following probability distribution

$$P(J_{v+1} = j \mid J_v) = \begin{cases} -\dfrac{Q(J_v, j)}{Q(J_v, J_v)} & \text{per } j \neq J_v \\ 0 & \text{for } j = J_v \end{cases} \qquad j = 1, 2, 3$$

   (d) if $\tau_{v+1} < T$, then $v = v + 1$

5. set $V = v + 1$, $\tau_{V+1} = T$ and $J_{V+1} = J_V$.

4.2 Simulation of a marked point process

The simulation of a marked point process can be subdivided in two steps: first, the simulation of the occurrence time for the point process t and then the simulation of the corresponding magnitude mark m.

4.2.1 Simulation of a point process

Let **N** be a marked point process with conditional intensity function λ_t.
Two simulation methods for point processes are available in the literature: the inverse transform method [WVjZ91] and the thinning method [LS76], [O98], [DVj88].
The **inverse transform method** is a classical simulation technique that can also be employed in *Algorithm 4* to sample from the exponential distribution. It is based on the following popular result.

Theorem [R76]: *if X is a real-valued random variable on a probability space $(\Omega, \mathcal{F}, P)$ with a continuous distribution function $F(t) = P(X \leq t)$, then $F(t)$ and $S(t) = 1 - F(t)$ have both uniform distribution on $[0, 1]$.*

From the previous theorem, the survivor function $S_n(t \mid t_1, ..., t_{n-1})$ of a point process $\mathbf{N}$, defined in (1.2), is uniformly distributed on $[0, 1]$. It follows that, if u is a sample from a uniform distribution with support $[0, 1]$ and if S_n^{-1} denotes the inverse function of the survivor function, the value $t = S_n^{-1}(u)$ is a simulated point of the point process $\mathbf{N}$.

The inverse transform method can be applied to the Poisson model and to the stress release model, because their survivor function can be easily inverted. Now, I give full details of the corresponding algorithm.

Assume that, in a time interval (τ_v, τ_{v+1}), a Poisson model with conditional intensity function μ is active. For $r < t$, the expression of the survivor function becomes

$$S(t \mid \mathcal{H}_r) = exp\left\{-\int_r^t \mu \, ds\right\} = exp\{-\mu \, (t - r)\}$$

Solving the equation $u = S(t \mid \mathcal{H}_r)$, where u is a uniform sample on $[0, 1]$, it results

$$t = r - \frac{\log u}{\mu}$$

Algorithm 5 : Simulation of a Poisson process

1. *set* $i = 0, \;\; t_i = \tau_v$
2. $u \sim Uniform(0, 1)$
3. $t_{i+1} = t_i - \dfrac{\log u}{\mu}$
4. *if* $t_{i+1} < \tau_{v+1}$, *then* $i \leftarrow i + 1$ *and go to the step (2). otherwise print the realization* $t_1, ..., t_i$.

Now, using the same notation, I assume that a stress release model is active in (τ_v, τ_{v+1}), having conditional intensity function $\lambda_t = e^{\alpha + \beta(\rho \, t - R(t))}$. For $r < t$, the expression of the survivor function becomes

$$S(t \mid \mathcal{H}_r) = exp\left[-\int_r^t e^{\alpha + \beta(\rho \, s - R(s))} ds\right] = exp\left[-A(e^{\beta\rho(t-r)} - 1)\right]$$

setting $A = \dfrac{\alpha + \beta + (\rho r - R(r))}{\beta\rho}$. Solving the equation $u = S(t \mid \mathcal{H}_r)$, where u is a

uniform random sample on $[0,1]$, it results

$$t = r + \frac{1}{\beta\rho} \log\left(1 - \frac{\log u}{A}\right)$$

Algorithm 6 : Simulation of a stress release model

1. *set* $i = 0$, $t_i = \tau_v$
2. $A = \dfrac{e^{\alpha+\beta(\rho\ t_i - R(t_i))}}{\beta\rho}$
3. $u \sim Uniform(0,1)$
4. $t_{i+1} = t_i + \dfrac{1}{\beta\rho} \log\left(1 - \dfrac{\log u}{A}\right)$
5. *if* $t_{i+1} < \tau_{v+1}$, *then* $i \leftarrow i+1$ *and go to the step (2) otherwise print the realization* $t_1, ..., t_i$.

The **thinning method**, devised by Lewis and Shedler in 1976, is used when some difficulty arises in the inversion of the survivor function, as in the Etas model.
Let $\mathbf{N}$ be a point process with conditional intensity function λ_t.
The thinning method can be applied when there exists a constant Λ such that $0 \leq \lambda_t \leq \Lambda$ for each $t \geq 0$. In that case, there exists a stationary Poisson process $\mathbf{N}^*$ with conditional intensity function Λ such that $\mathbf{N}^*$ dominates $\mathbf{N}$.
So, the process $\mathbf{N}^*$ can be simulated using the inverse transform method: let $r_1, ..., r_m$ be one of its realizations.
The sequence $t_1, ..., t_n$, obtained from $r_1, ..., r_m$ by deleting each r_i that does not satisfy the random relationship

$$u_i \leq \frac{\lambda_{r_i}}{\Lambda} \quad \text{where } u_i \sim U(0,1)$$

is a realization of the point process $\mathbf{N}$.
It can be noted that this procedure is based on the same principle of the acceptance/rejection method [R01].
The thinning method presents some practical problems. Sometimes the determination of a constant Λ, such that $\lambda_t \leq \Lambda$ for each $t \geq 0$, is not obvious; moreover, if Λ is much greater than λ_t, a large number of the generated realizations of $\mathbf{N}^*$ will be rejected, with a consequent wasted high computational cost.
The Etas model typically presents the above-mentioned difficulties.

Let $\lambda(3,t) = \sum_{i:t_i<t} \frac{ke^{\gamma(m_i-M_0)}}{(t-t_i+c)^p}$ denote the conditional intensity function of an Etas model. If the parameter c is very small and two events occurred in a very small time interval, the conditional intensity function λ_t could attain a very large value; then it is practically impossible to fix a global efficient upper bound Λ on $[0,T]$. But for a sub-interval (t_i, t_{i+1}) between two subsequent events, a partial upper bound can be given by any value Λ greater than λ_{t_i}, since λ_t is a decreasing function on (t_i, t_{i+1}).

Algorithm 7: Simulation of an ETAS model

1. *set* $i = 0, \ t_i = \tau_v$
2. *for some* s *(for example* $s = 1.5$*),* $\Lambda = s \cdot \lambda(3, t_i)$
3. $j = 0, \ r_{old} = 0$
4. $u \sim Uniform(0,1)$
5. $r_{new} = r_{old} - \dfrac{\log u}{\Lambda}$,
6. *if* $r_{new} \geq \tau_{v+1}$*, print the realization* $t_1, ..., t_i$.
7. $u \sim Uniform(0,1)$
8. *if* $\dfrac{\lambda(3, t_i + r_{new})}{\Lambda} < u$,
   then $t_{i+1} = t_i + r_{new}$, $i \leftarrow i+1$ *and go to (2)*
   otherwise $r_{old} = r_{new}$ *and go to (4).*

Remark : in the first jump interval $[0, \tau_1)$, since no event is known before time 0, it is necessarily $\lambda(3,t) = 0$ for $t \in [0, \tau_1)$. It follows that the Etas simulation does not produce events in $[0, \tau_1)$, but this is not seismologically reasonable, even if it is mathematically correct. A possible trick to overcome this problem could be a choice of the initial time $t = 0$ in correspondence of an occurrence; practically, an event is simulated at time 0.

4.2.2 Magnitude simulation

The magnitude distribution is assumed to be time independent and to follow the Gutenberg-Richter law, in which the number $F(m)$ of earthquakes with magnitude

m or larger is given by

$$F(m) = 10^{a-bm} \tag{4.1}$$

Let $\Phi(m)$ represent the magnitude distribution conditionally to $m > M_0$, for a fixed magnitude threshold; using the Gutenberg-Richter law, it holds

$$1 - \Phi(m) = P(M > m \mid M > M_0) = \frac{P(M > m)}{P(M > M_0)} = \frac{10^{a-bm}}{10^{a-bM_0}} =$$

$$= 10^{-b(m-M_0)} = e^{-b \log_e 10 (m-M_0)}$$

Then, the magnitude distribution is an exponential with mean the reciprocal of $b \log_e 10$ and support $[M_0, +\infty)$.
The inverse transform method provides simulations from the magnitude distribution. The following procedure assigns the respective mark m_i to each time occurrence t_i.

Algorithm 8: magnitude simulation

1. $u \sim Uniform(0, 1)$
2. $m_i = M_0 - \frac{\log u}{b \log_e 10}$

4.3 Simulated data set

I considered a time interval $[0, T] = [0, 2000]$, in years.
In the literature, numerous studies on each of the point processes we are considering are available for different seismic regions. On the basis of these studies, I fixed the parameter values of the considered point processes as follows

Poisson model	$\mu = 0.50$			
SRM	$\alpha = -4.50$	$\beta = 0.02$	$\rho = 2.80$	
ETAS model	$k = 0.10$	$\gamma = 0.60$	$c = 0.50$	$p = 1.00$

About the hidden state process, less information is available.
So, I chose a uniform initial distribution $(\hat{Z}_0(1), \hat{Z}_0(2), \hat{Z}_0(3))$. The generator Q of the hidden pure jump Markov process is built by exploiting the relationship between the infinitesimal probability, e.g. $1+q_{11}$ of stay in the present state 1 and the mean holding time, given by $-1/q_{11}$. So, assuming that the mean holding times are about 50 years for the Poisson and the stress release models and 20 years for the ETAS model, we have $q_{11} \approx q_{22} \approx -0.02$ and $q_{33} \approx -0.06$.
The remaining jump probabilities q_{12} and q_{13} are subdivided in the proportions of 90% and 10%; precisely, assuming that a state jump occurs when we are in state 1, I supposed that there is more probability of jumping from the state 1 to the state 2, than from the state 1 to the state 3; analogously, there is more probability of jump from the state 2 to the state 3, than from the state 2 to the state 1; and so on.

$$\hat{Z}_0(1) = \frac{1}{3} \qquad \hat{Z}_0(2) = \frac{1}{3} \qquad \hat{Z}_0(3) = \frac{1}{3}$$

$$Q = \begin{pmatrix} -0.02 & 0.018 & 0.002 \\ 0.0015 & -0.015 & 0.0135 \\ 0.054 & 0.006 & -0.06 \end{pmatrix}$$

The simulation of the state-space model produces a sequence of 72 state changes and 908 earthquakes.
In the lower part of figure (4.1), each earthquake is represented by a vertical coloured line proportional to its magnitude; the colour corresponds to the active point process: yellow for the Poisson process, red for the SRM, green for the ETAS model. In the upper part of figure (4.1), the black line represents the conditional intensity function of the simulated process and the horizontal coloured line the states active at each istant. Since the Etas model has a different scale with respect to the other models, two units of measure are used in the representation of the conditional intensity function.

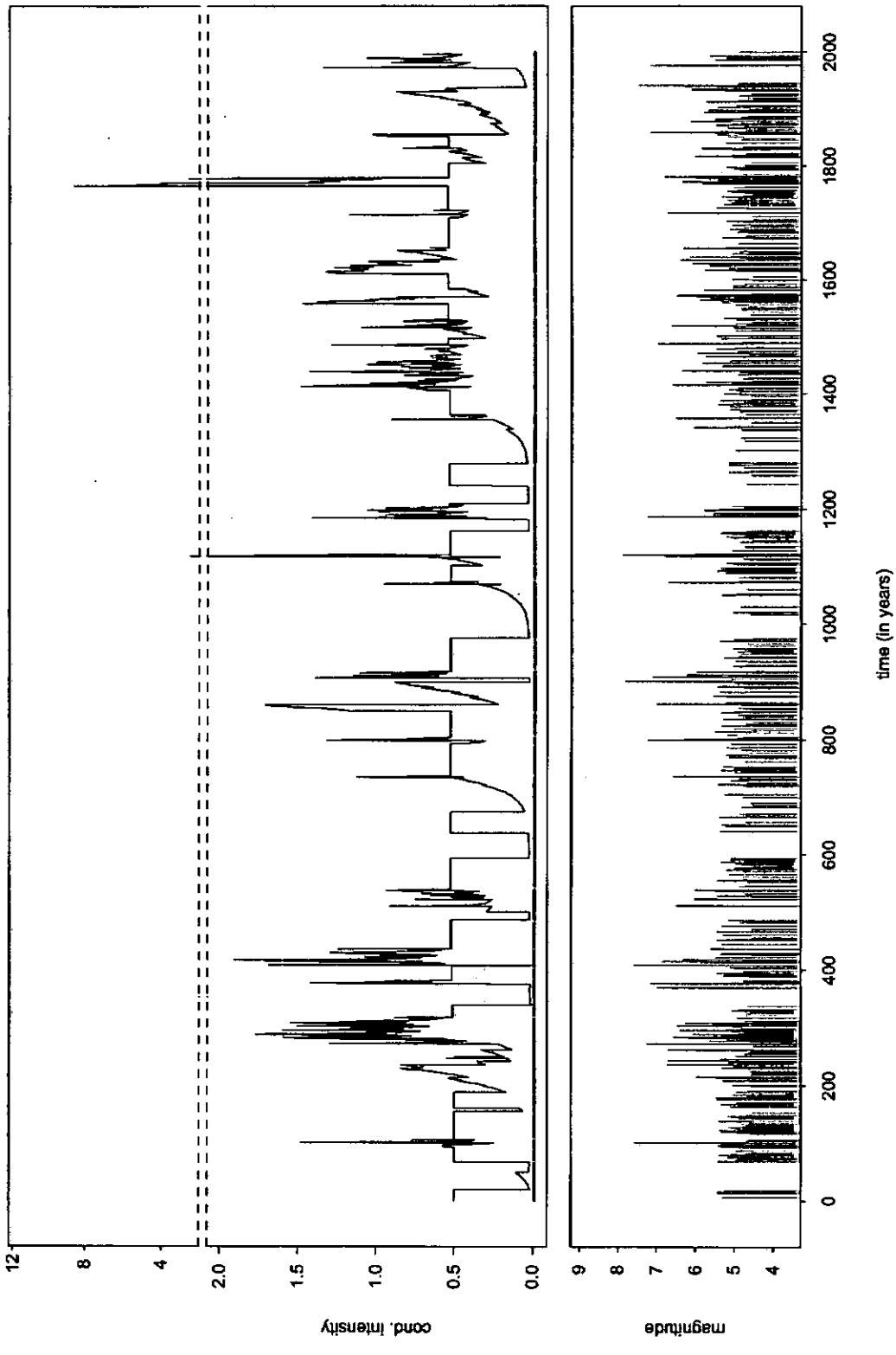

Figure 4.1: Conditional intensity function of the point process associated with the state simulated at each time t (upper picture) - simulated data set (lower picture)

4.4 Filtering

In this section, I suppose that the parameter values are known and I provide an approximation of the filtering problem by exploiting *Algorithm 3* in chapter 3.
In *Algorithm 3*, I set $D = 100000$, that is each Monte Carlo approximation of the involved integrals is obtained by using 100000 samples from the corresponding distributions.
All the algorithms in the present thesis are implemented in Fortran.
The results are summarized in some graphical representations.

Each graph of Figure (4.2) represents the filtering probability $\hat{Z}_t(j) = P(X_t = j)$, for each state $j = 1, 2, 3$; the colours indicate the instants when the considered state matches the simulated state. I recall that yellow, red and green colours are associated respectively with to Poisson, stress release and ETAS model.

Figure (4.3) represents the maximum filtering probability, $\max_{j=1,2,3} \hat{Z}_t(j)$ for each time $t \in [0, 2000]$ and the colours correspond to $argmax_{j=1,2,3}\hat{Z}_t(j)$; the simulated data set is represented in the lower part of the graph and the colour of the bars denote the state in which the system is in that instant.

Figure (4.4) represents the maximum filtering probability at the instants t when $argmax_{j=1,2,3}\hat{Z}_t(j)$ does not correspond to the simulated state.

In Figure (4.5), the expected conditional intensity function, $\hat{\lambda}_t = E(\lambda_t \mid \mathcal{H}_t)$ (violet line), is compared with the true conditional intensity function (black line).

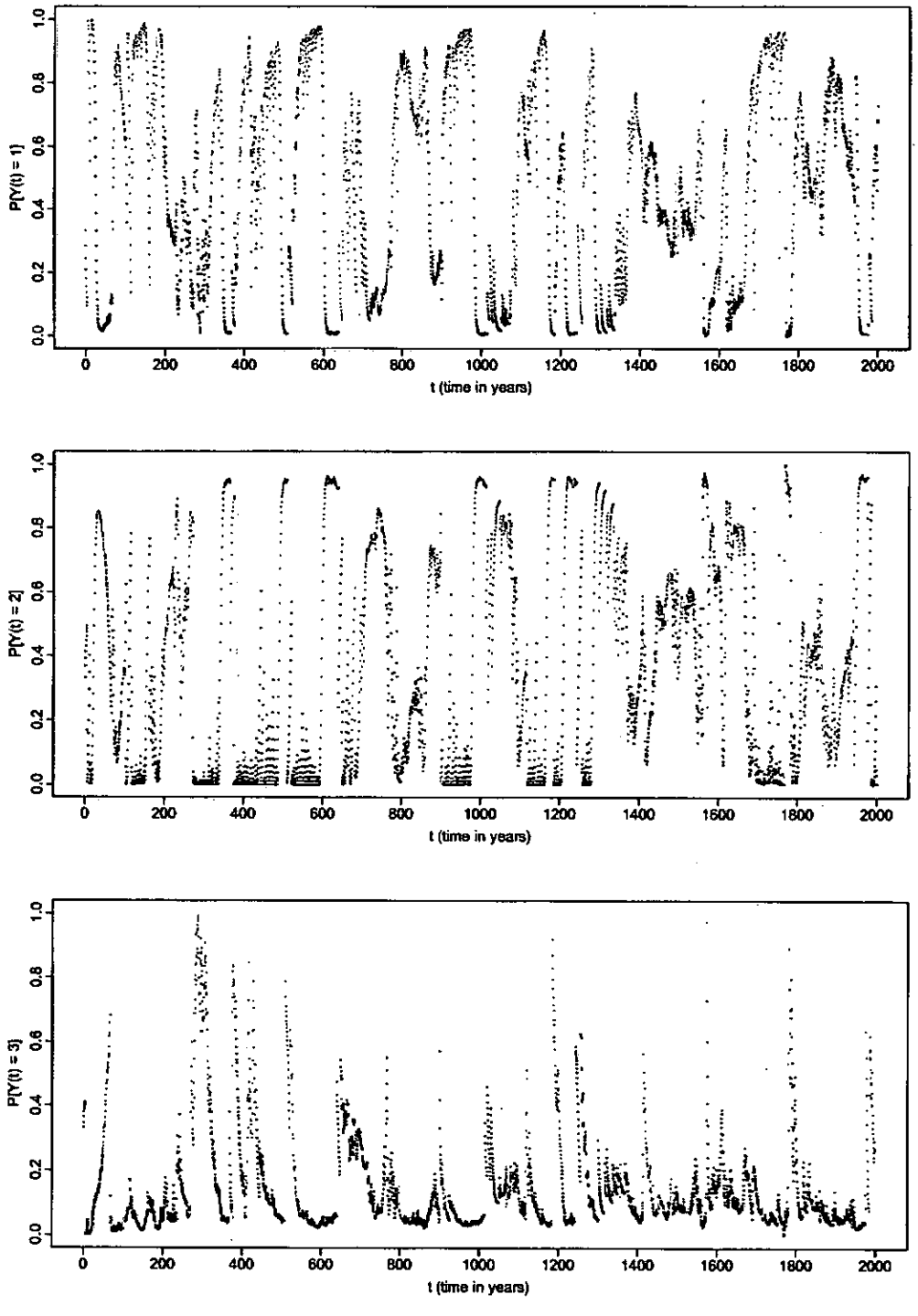

Figure 4.2: Filtering probability of each state.

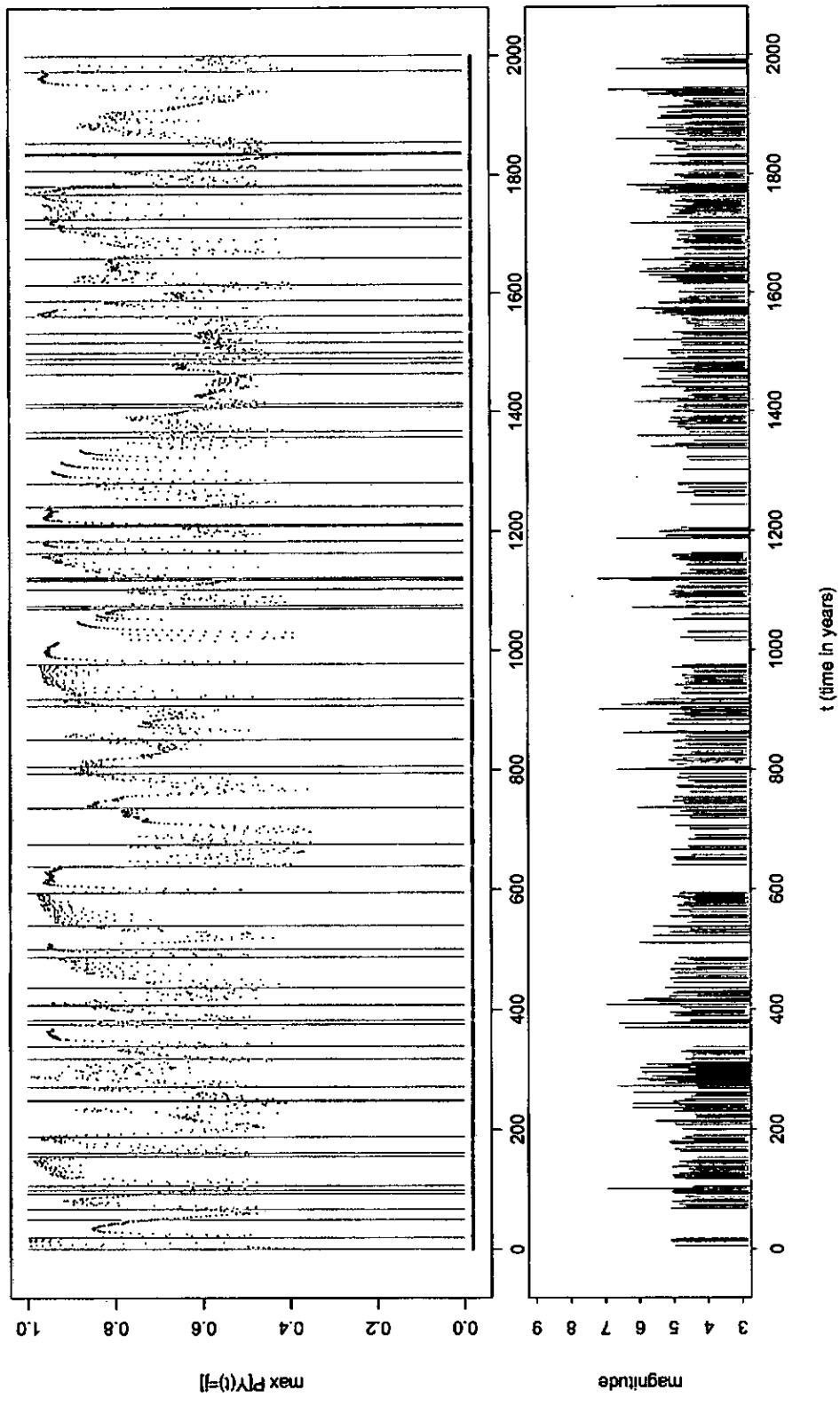

Figure 4.3: Maximum of the filtering probability of the states at each time t.

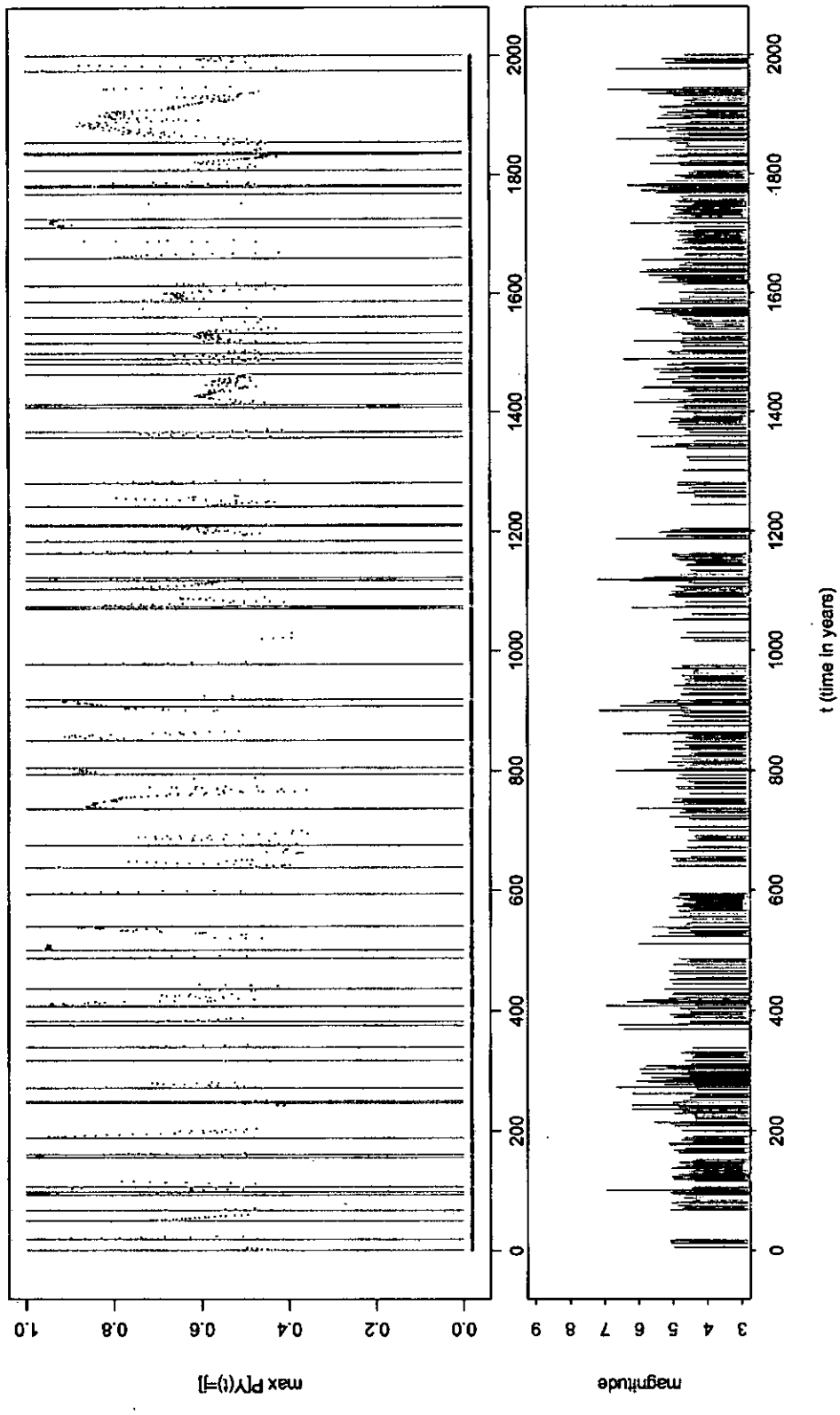

Figure 4.4: Maximum filtering probability at the times t when there is not agreement between the state with maximum probability and the simulated state.

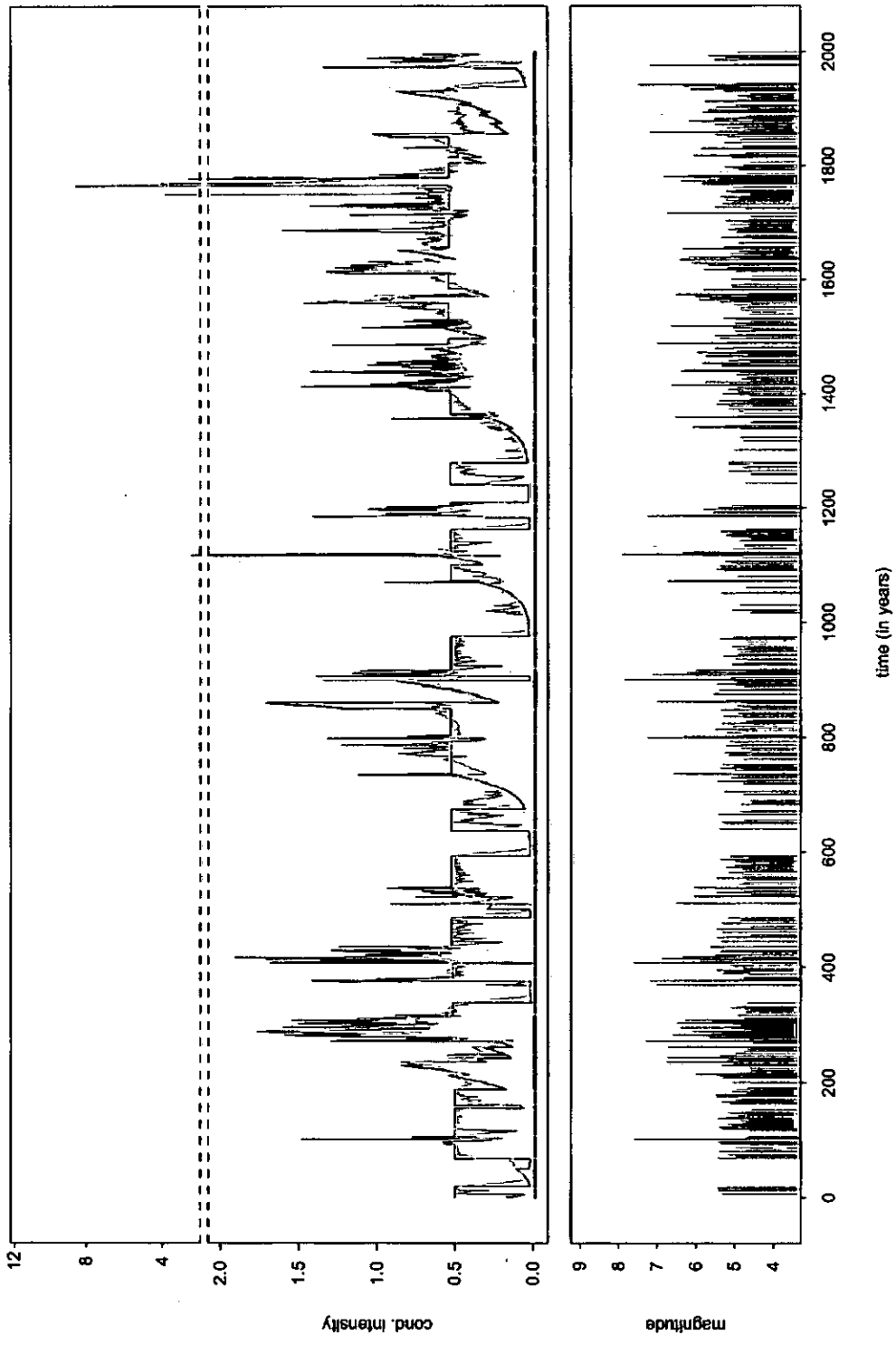

Figure 4.5: Conditional intensity function related to the simulation of the state-space model (black line) and mixture of the conditional intensity functions with weights given by the estimated filtering probabilities (violet line).

4.5 Results on parameter estimation

In this section, I assume that all the parameters are unknown and, in order to estimate them, I apply a sequential importance sampling method following *Algorithm 2* in chapter 2.
The analysis is carried out in two steps: first Algorithm 2 is applied including the Künsch-Hürzeler method in order to explore better the domain of each parameter; then Algorithm 2 is applied without the Künsch-Hürzeler method in order to remove the bias introduced by this technique.

4.5.1 First step: prior and importance distributions

First of all it is necessary to fix the input for the Bayesian sequential analysis of the proposed state-space model: prior and importance distributions.
I recall that importance and prior distributions are equal, so that the weights have a simple iterative expression.
We obtain information on mean and range of variability of their parameters from the literature regarding the considered point processes; on this basis, we assign the prior distributions reported in the following table:

parameter	distribution	mean	variance
μ	Gamma	0.4	0.03
α	Normal	-5.	3.
β	Gamma	0.01	0.00005
ρ	Gamma	1.5	0.7
k	Gamma	0.5	0.5
γ	Gamma	0.7	0.2
c	Gamma	0.3	0.02
p	Gamma	1.	0.005

Table 4.1: Prior and importance distributions for the parameter of the point process models at the first step.

The initial distribution for the hidden process is chosen uniform.
Denoting the identity matrix by I, since the rows of $I+Q$ have sum equal to one, I have chosen a Dirichlet prior distribution for the elements of Q belonging to the same row. The diagonal parameters of $I+Q$ are related to the holding probabilities

fixed around 93%; moreover, I assumed no information is available on the transition probabilities between different states. So, the Dirichlet prior distributions for each row are the following

parameters	distribution and its parameters	means and variances
$(1-q_{11}, q_{12}, q_{13})$	Dirichlet(35., 2., 0.5)	(0.9333, 0.0533, 0.0133) (0.0016, 0.0013, 0.0003)
$(q_{21}, 1-q_{22}, q_{23})$	Dirichlet(0.5, 35., 2.)	(0.0133, 0.9333, 0.0533) (0.0003, 0.0016, 0.0013)
$(q_{31}, 1-q_{32}, q_{33})$	Dirichlet(2., 0.5, 35.)	(0.0533, 0.0133, 0.9333) (0.0013, 0.0003, 0.0016)

Table 4.2: Prior and importance distributions for the parameters of the hidden process at the first step.

4.5.2 First step: considerations on d in the Künsch-Hürzeler method

I have executed runs of the sequential estimation procedure by using a number $m = 30000$ of particles.

When the effective sample size value ESS_t at time t is less than 24000 effective particles (the 80% of m), a resampling step and a Künsch-Hürzeler rejuvenation of the particle set is carried out.

The Künsch-Hürzeler technique is applied adding to each parameter in the particles a zero-mean Normal noise with standard deviation A/t^d.

The constant A depends on the size of the parameter; in particular, Table 4.3 contains the values I used:

A					
μ	0.01	γ	0.1	q_{21}	0.0001
α	0.05	c	0.02	q_{23}	0.001
β	0.0005	p	0.005	q_{31}	0.001
ρ	0.1	q_{12}	0.001	q_{32}	0.0001
k	0.05	q_{13}	0.0001		

Table 4.3: Choice of the A values in the application of the Künsch-Hürzeler method.

With regard to the value of d, Künsch and Hürzeler affirmed in [DFG01] that *"considerations from the recursive estimation suggest taking the variance (of the noise) of the order* t^{-2}*"* so that the difference between the original and the perturbed particle is a quantity of the order t^{-1} (that is $d = 1$).

In the case studies in [DFG01], they analysed sequences where the time interval length is of some tens, for example $t \in [0, 30]$; since I fixed $t \in [0, 2000]$, the assumption $d = 1$ imply that the noise effect becomes negligible too early during the sequential estimation. In order to show this phenomenon, I considered the first 200 simulated data covering almost 400 years and I run three versions of the *Algorithm 3*: version A does not apply Künsch and Hürzeler method, version B has $d = 1$ and version C has $d = 0.2$.

In Figure 4.6, I compared the histogram of the parameter posterior distributions for each run. It can be noted that versions A and B are almost indistinguishable and subject to a remarkable impoverishment of the particle set. Since this effect is observed for the sequential analysis of the fifth part of the data set, it is reasonable to deduce that versions A and B lead to the same inference and I checked it through further runs. On the contrary, version C is again able to perturb the particles and to combat the impoverishment effect.

My conclusion is that $d = 1$ is not a suitable value for my sequence of data; in particular, I choose $d = 0.2$ because it keeps the value t^{-d} of the same order for each $t \in [0, 2000]$ so that the added noise to the particles decreases, but is not negligible during the sequential analysis.

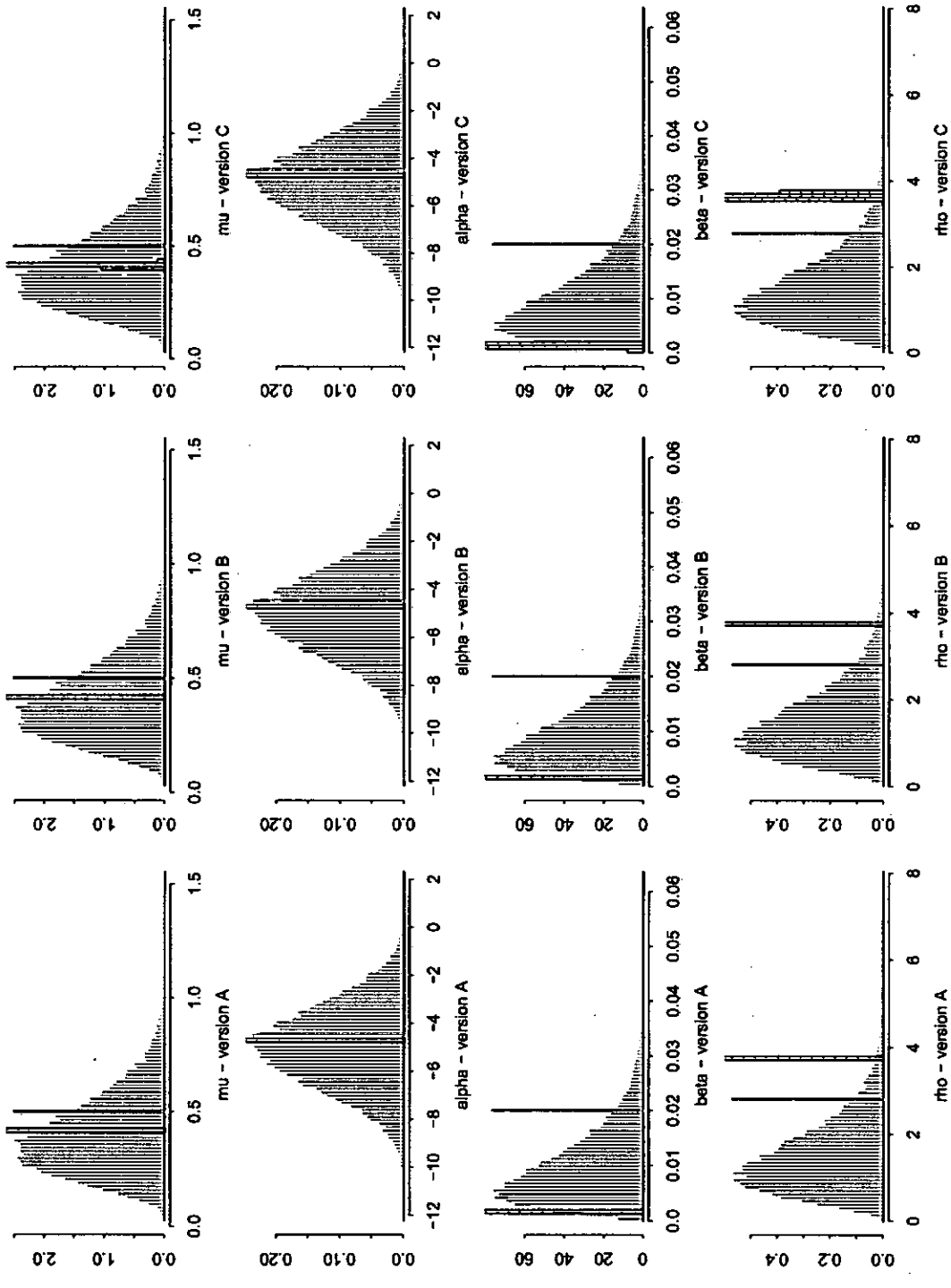

Figure 4.6: For each parameter prior (violet histogram) and scaled posterior (ruled histogram) particle set, value used in the simulation (green bar).

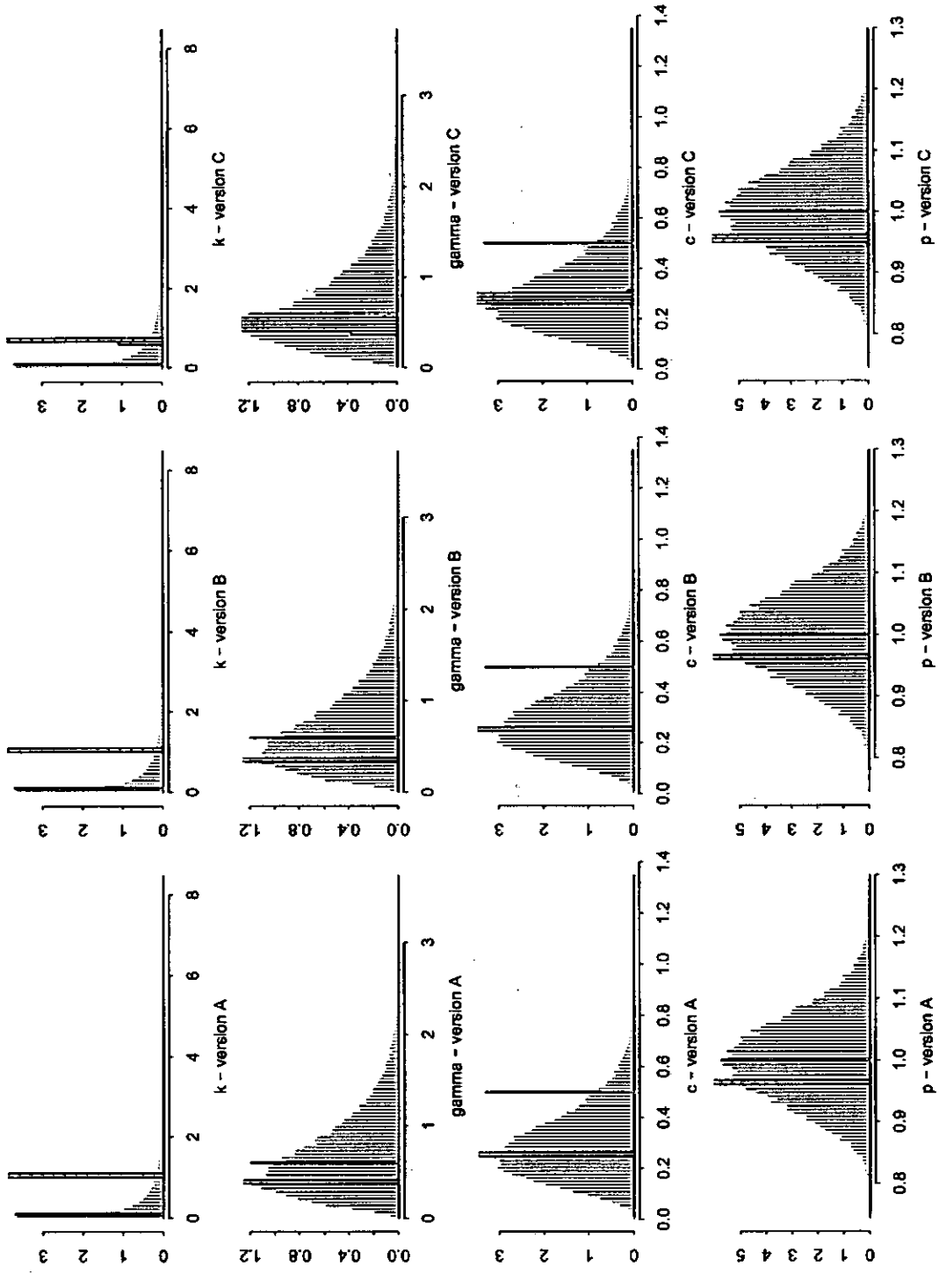

Figure 4.6: to be continued.

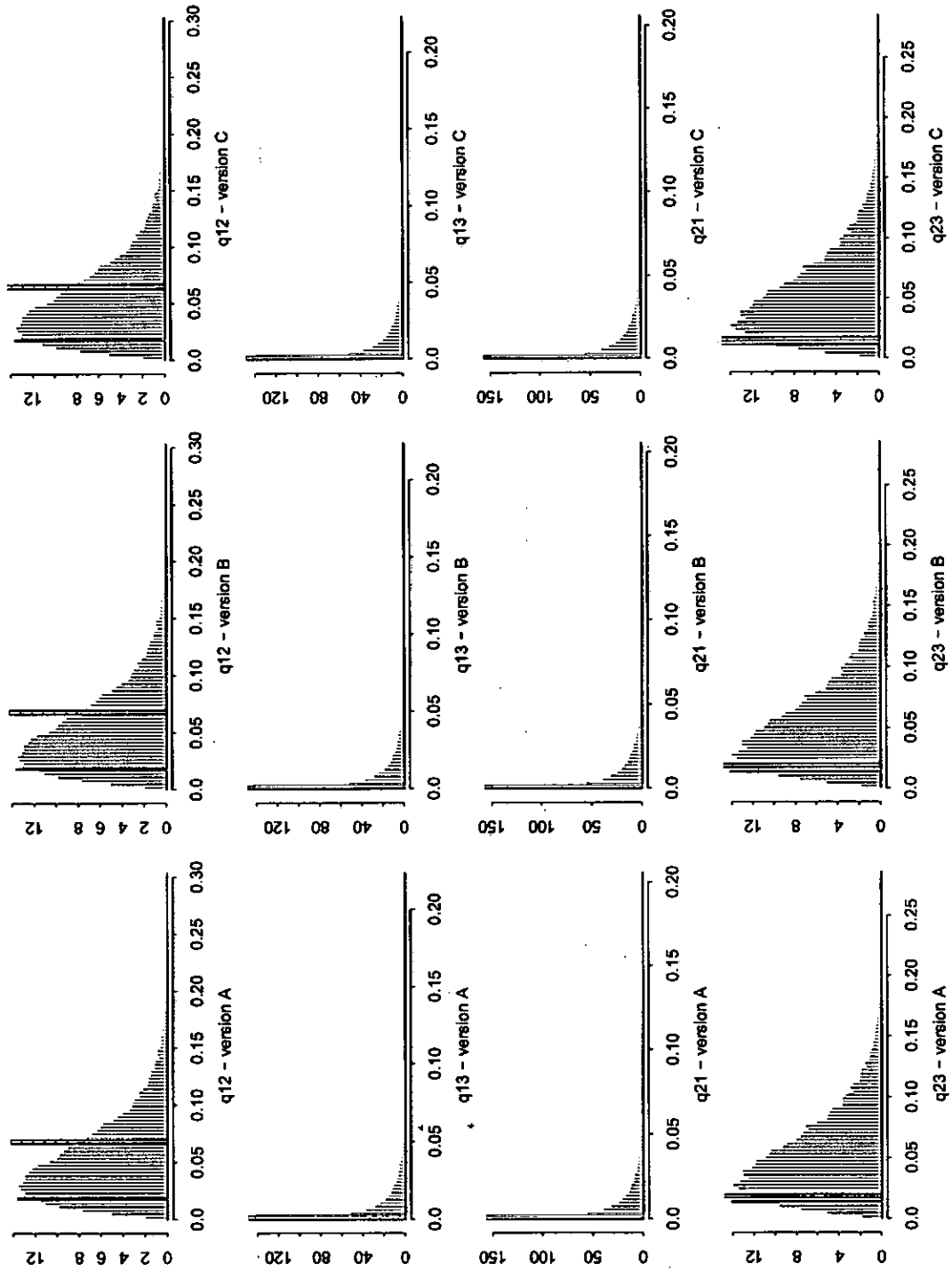

Figure 4.6: to be continued.

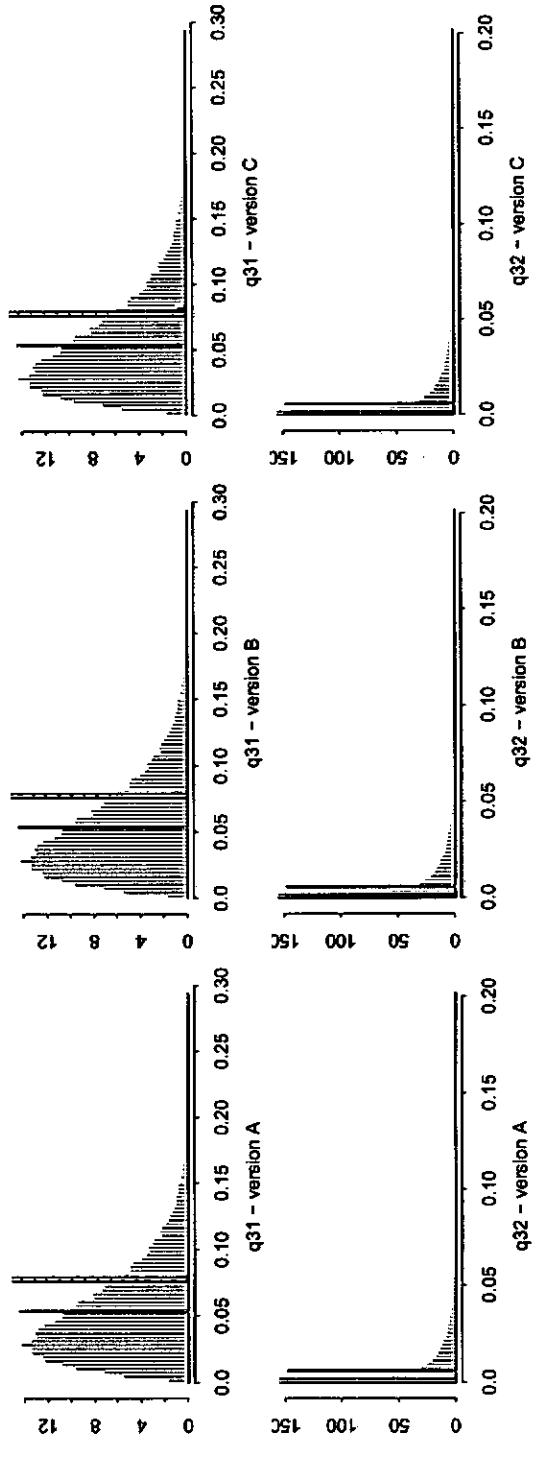

Figure 4.6: to be continued.

4.5.3 First step results

Summarizing the first step, I applied *Algorithm 2* using a particle set of size $m = 30000$ and the resampling method described in *Algorithm 1* when the effective sample size ESS_t is less than 24000.
Moreover, in correspondence of each resampling, the Künsch-Hürzeler method rejuvenates the set of the particles introducing a normal noise having zero mean and standard deviation $A \cdot t^{-d}$, where $d = 0.2$ and A is indicated in Table 4.3.
Prior and importance distributions of the parameters are equal and summarized in Table 4.1 and 4.2.
I executed three different runs of this procedure. Each run produces a set of weighted particles that approximates the posterior distribution of the parameters; in Table 4.4, I reported the estimated posterior means of the parameters and their posterior standard deviations; moreover, at the end of each table, I averaged out the estimated posterior means and I assigned a value to the standard deviation such that the corresponding confidence interval covers all the confidence intervals associated with the estimates obtained by the three runs. These means and standard deviations constituted the input of the sequential procedure at the second step.

	$\hat{\mu}$		$\hat{\alpha}$		$\hat{\beta}$	
run	mean	st. dev.	mean	st. dev.	mean	st. dev.
1	0.3497	$2.18 \cdot 10^{-3}$	-4.4536	$1.10 \cdot 10^{-2}$	$6.3571 \cdot 10^{-4}$	$1.09 \cdot 10^{-4}$
2	0.7632	$2.1836 \cdot 10^{-3}$	-7.1704	$1.1019 \cdot 10^{-2}$	$1.1085 \cdot 10^{-2}$	$1.0865 \cdot 10^{-4}$
3	0.2659	$2.19 \cdot 10^{-3}$	-2.4867	$1.09 \cdot 10^{-2}$	$3.0335 \cdot 10^{-3}$	$1.09 \cdot 10^{-4}$
	0.4596	0.15	-4.7036	1.63	$4.9182 \cdot 10^{-3}$	$8.76 \cdot 10^{-3}$

	$\hat{\rho}$		$\hat{k}$		$\hat{\gamma}$	
run	mean	st. dev.	mean	st. dev.	mean	st. dev.
1	2.9970	$2.17 \cdot 10^{-2}$	0.2990	$1.09 \cdot 10^{-2}$	0.6818	$2.19 \cdot 10^{-2}$
2	2.9914	$2.16 \cdot 10^{-2}$	0.3187	$1.09 \cdot 10^{-2}$	0.4732	$2.19 \cdot 10^{-2}$
3	2.5765	$2.17 \cdot 10^{-2}$	0.2468	$1.09 \cdot 10^{-2}$	0.1399	$2.18 \cdot 10^{-2}$
	2.8550	0.38	0.2882	0.78	0.4316	0.54

	$\hat{c}$		$\hat{p}$		$\hat{q}_{12}$	
run	mean	st. dev.	mean	st. dev.	mean	st. dev.
1	0.2113	$4.36 \cdot 10^{-3}$	0.9545	$1.10 \cdot 10^{-3}$	$6.1668 \cdot 10^{-2}$	$2.19 \cdot 10^{-4}$
2	0.3483	$4.36 \cdot 10^{-3}$	0.9459	$1.09 \cdot 10^{-3}$	$6.1674 \cdot 10^{-2}$	$2.18 \cdot 10^{-4}$
3	0.2620	$4.40 \cdot 10^{-3}$	1.0046	$1.09 \cdot 10^{-3}$	$1.9956 \cdot 10^{-2}$	$2.20 \cdot 10^{-4}$
	0.2739	0.15	0.9683	0.08	$4.7766 \cdot 10^{-2}$	0.04

	$\hat{q}_{13}$		$\hat{q}_{21}$		$\hat{q}_{23}$	
run	mean	st. dev.	mean	st. dev.	mean	st. dev.
1	$5.4604 \cdot 10^{-4}$	$2.16 \cdot 10^{-5}$	$2.9426 \cdot 10^{-3}$	$2.21 \cdot 10^{-5}$	$1.2537 \cdot 10^{-2}$	$2.19 \cdot 10^{-4}$
2	$2.0387 \cdot 10^{-2}$	$2.18 \cdot 10^{-5}$	$3.4082 \cdot 10^{-3}$	$2.19 \cdot 10^{-5}$	0.1195	$2.17 \cdot 10^{-4}$
3	$1.8226 \cdot 10^{-2}$	$2.19 \cdot 10^{-5}$	$1.4416 \cdot 10^{-2}$	$2.19 \cdot 10^{-5}$	$7.1115 \cdot 10^{-2}$	$2.20 \cdot 10^{-4}$
	$1.3053 \cdot 10^{-2}$	0.02	$6.9222 \cdot 10^{-3}$	0.02	$6.7712 \cdot 10^{-2}$	0.03

	$\hat{q}_{31}$		$\hat{q}_{32}$	
run	mean	st. dev.	mean	st. dev.
1	$7.7548 \cdot 10^{-2}$	$2.19 \cdot 10^{-4}$	$9.2085 \cdot 10^{-4}$	$2.18 \cdot 10^{-5}$
2	$6.6437 \cdot 10^{-2}$	$2.16 \cdot 10^{-4}$	$9.2881 \cdot 10^{-4}$	$2.20 \cdot 10^{-5}$
3	$5.2666 \cdot 10^{-2}$	$2.20 \cdot 10^{-4}$	$2.5168 \cdot 10^{-3}$	$2.19 \cdot 10^{-5}$
	$6.555 \cdot 10^{-2}$	0.03	$1.4555 \cdot 10^{-3}$	0.02

Table 4.4: For each run at the first step, the estimated posterior mean and standard deviation of each parameter are reported; at the end of each table, there are mean and standard deviation I used in the second step of the sequential analysis.

Each picture in Figure 4.7 compares the histogram of the prior particle set (violet histogram) with the histogram of the posterior particle set (ruled histogram).

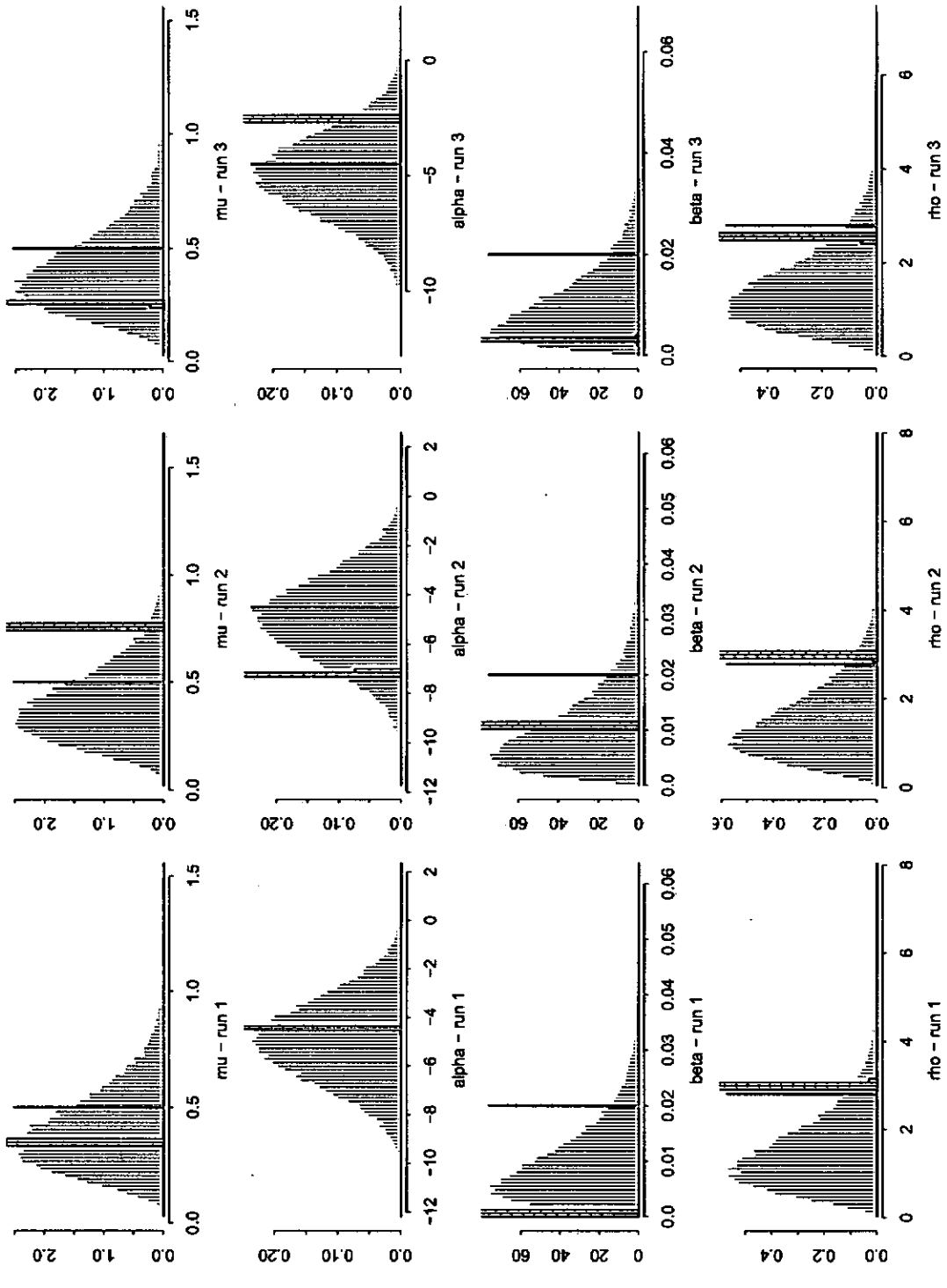

Figure 4.7: First step of the sequential procedure - for each parameter prior (violet histogram) and scaled posterior (ruled histogram) particle set, value used in the simulation (green bar).

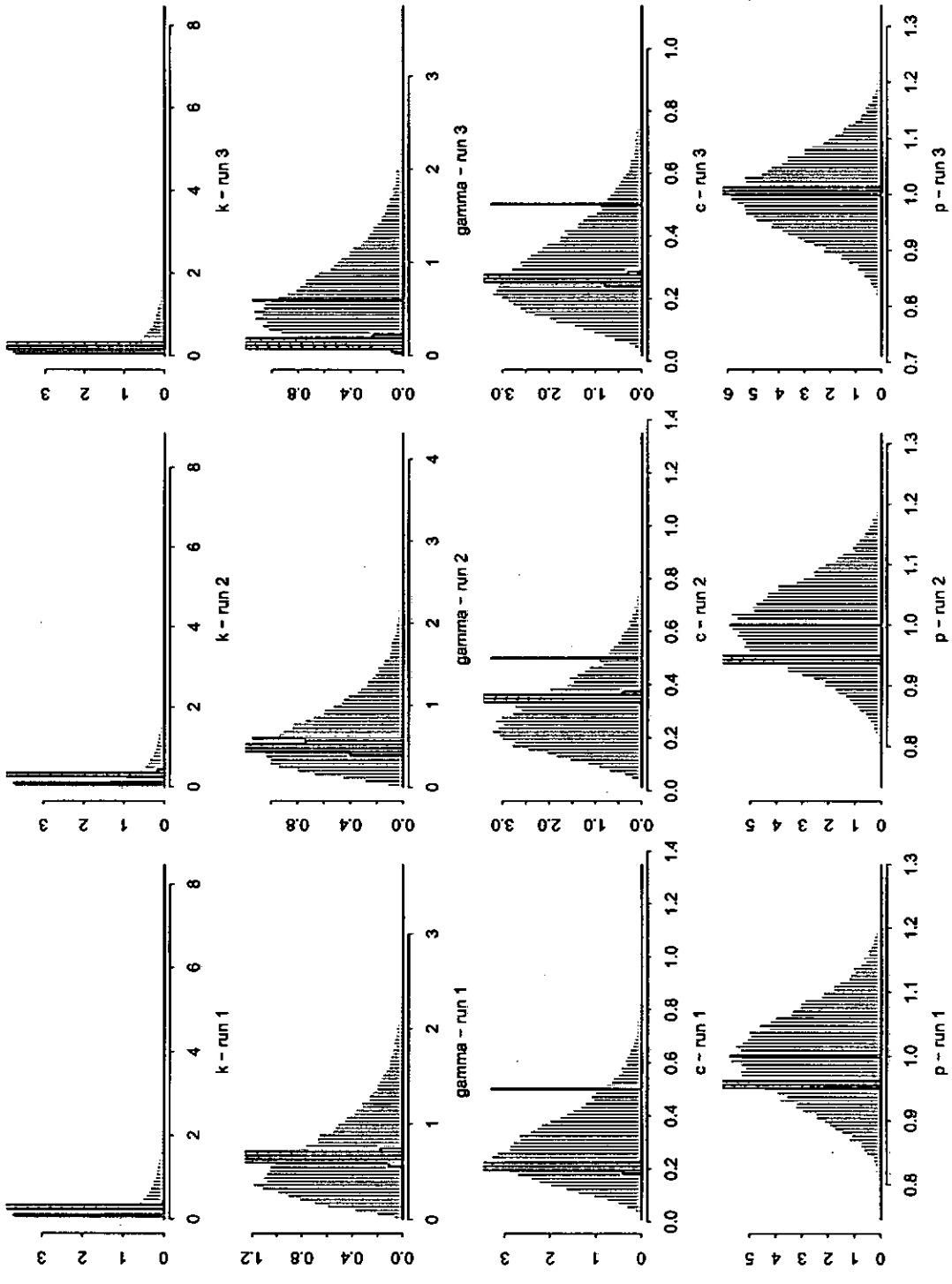

Figure 4.7: to be continued.

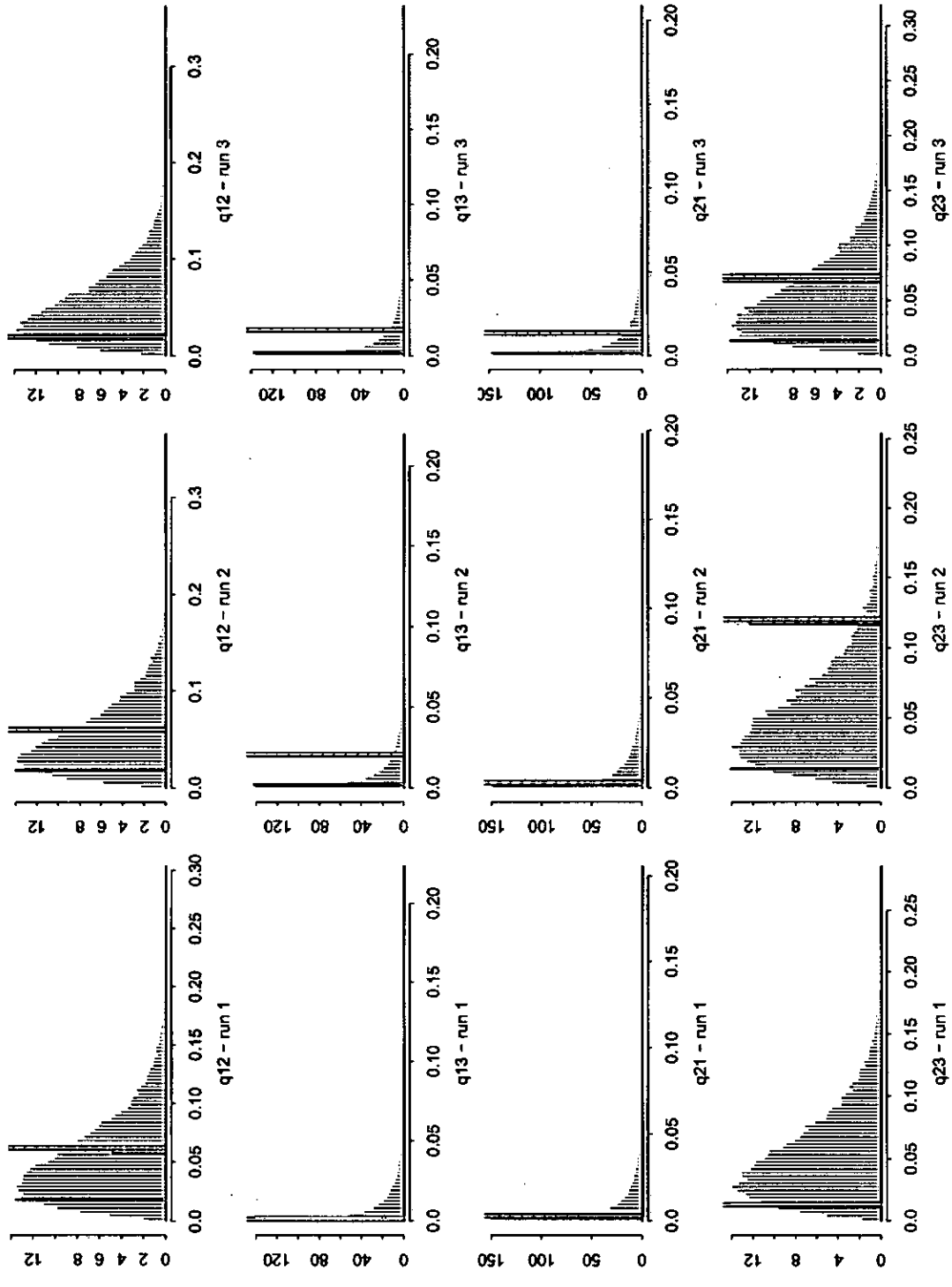

Figure 4.7: to be continued.

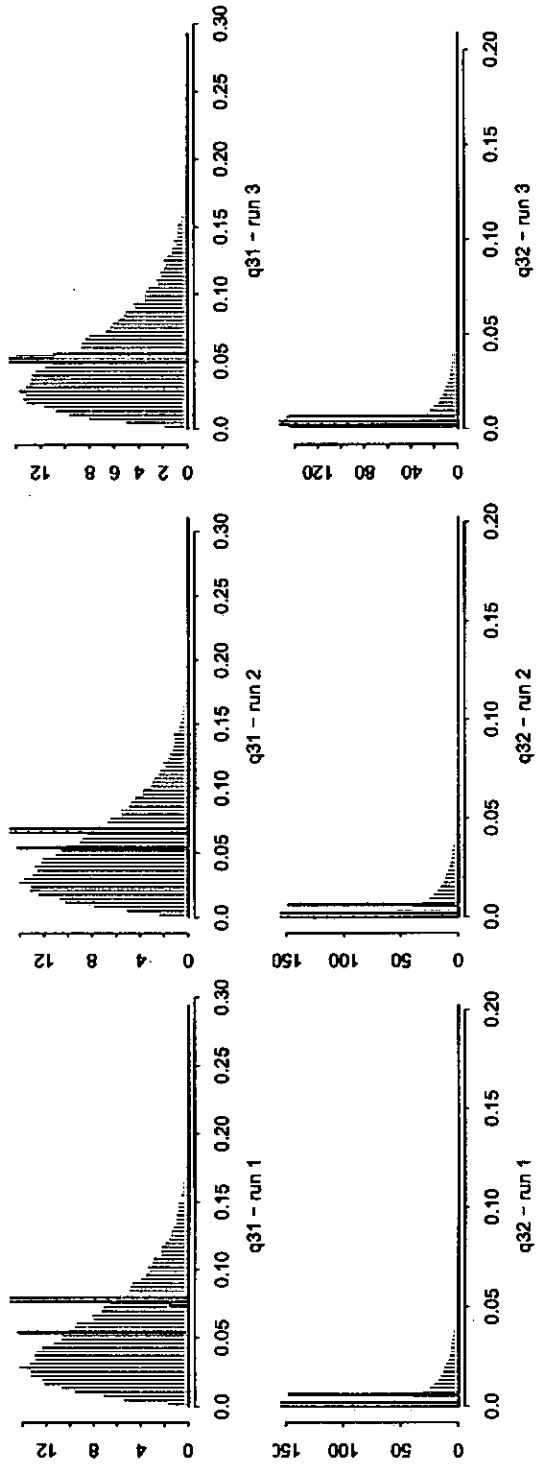

Figure 4.7: to be continued.

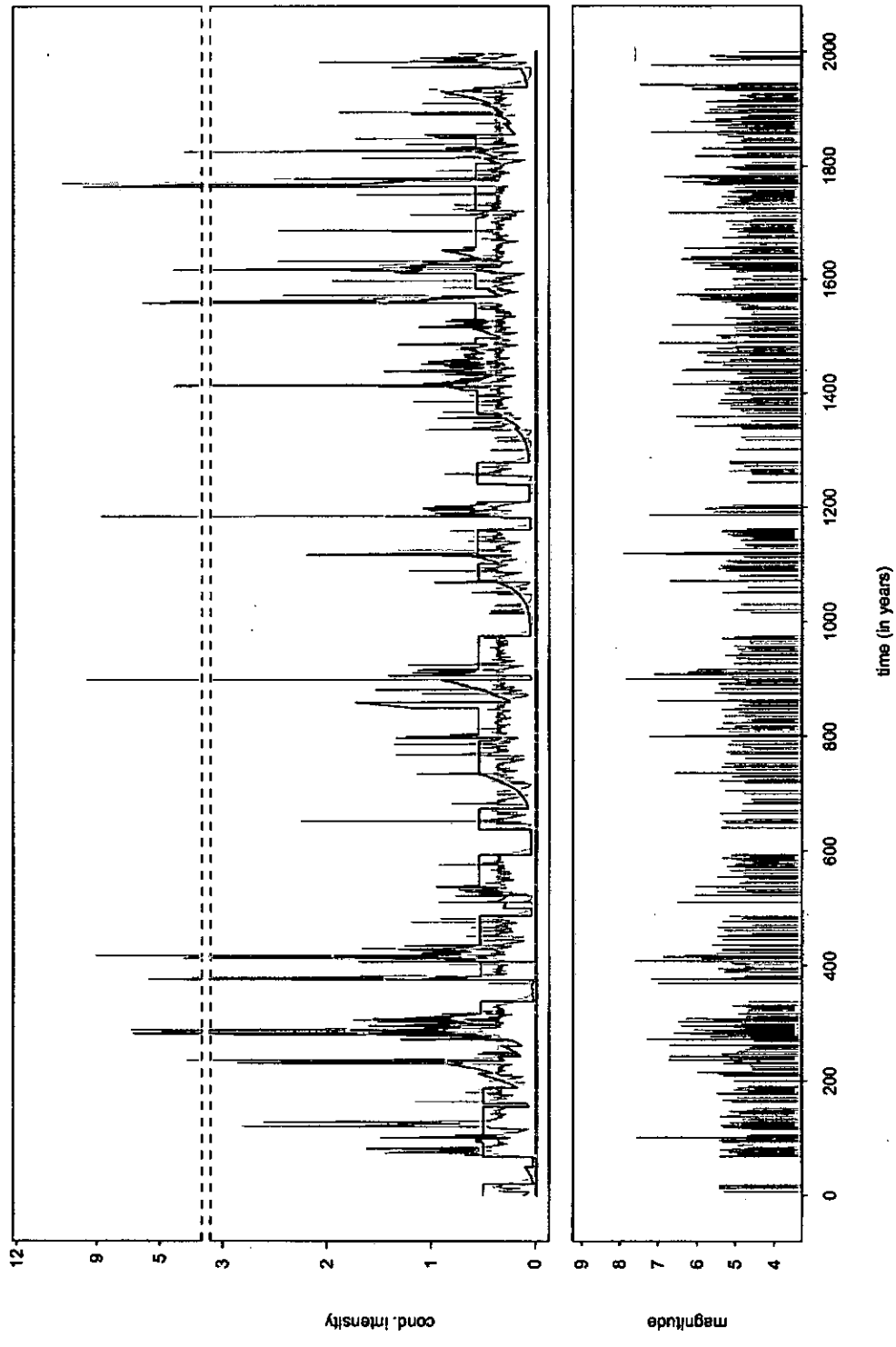

Figure 4.8: First step - Run 1 - Conditional intensity function related to the simulation of the state-space model (black line) and mixture of the plug-in estimates of the conditional intensity functions (violet line) with weights given by the estimated filtering probabilities (see *Algorithm 3*).

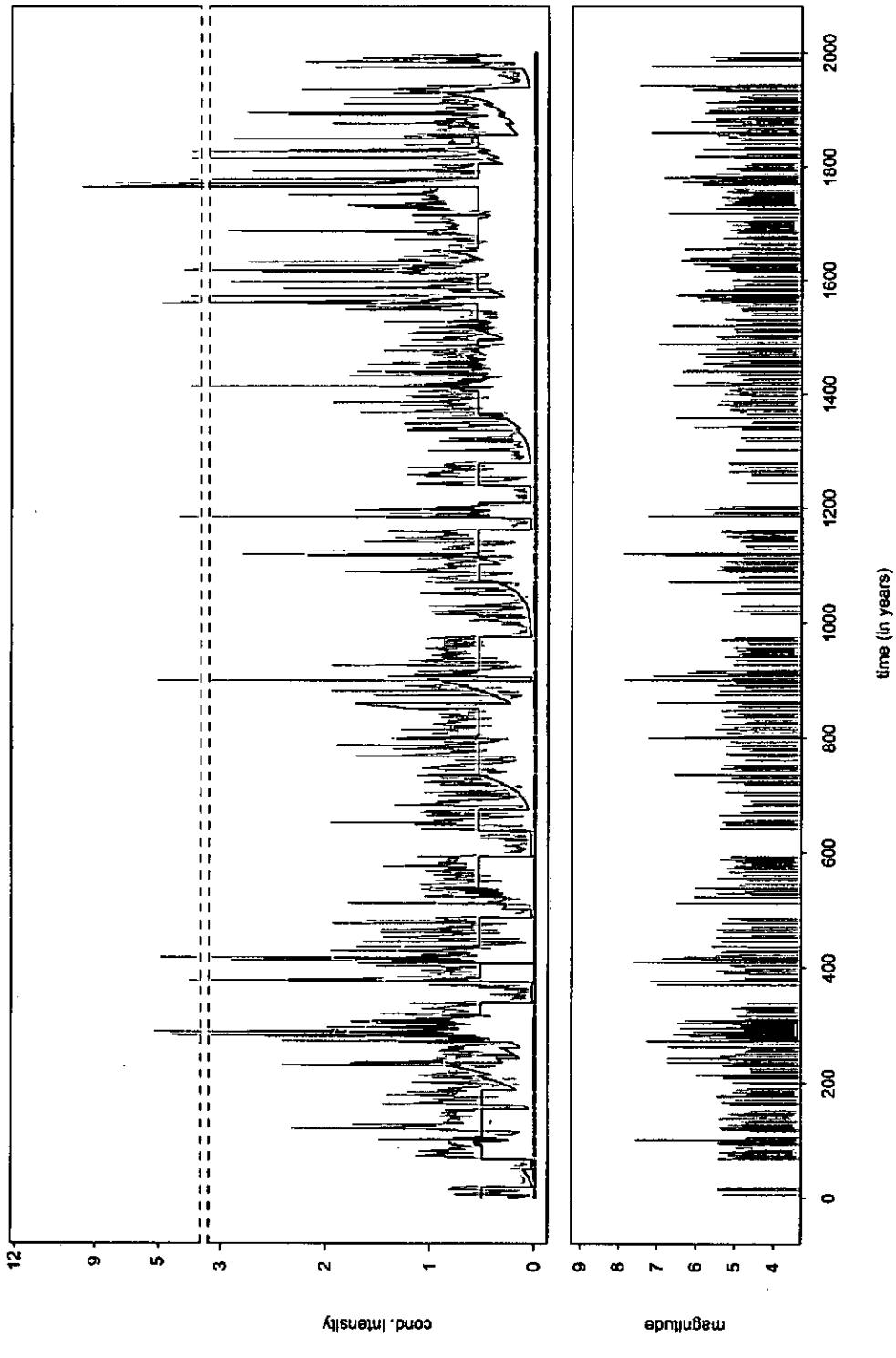

Figure 4.9: First step - Run 2 - Conditional intensity function related to the simulation of the state-space model (black line) and mixture of the plug-in estimates of the conditional intensity functions (violet line) with weights given by the estimated filtering probabilities.

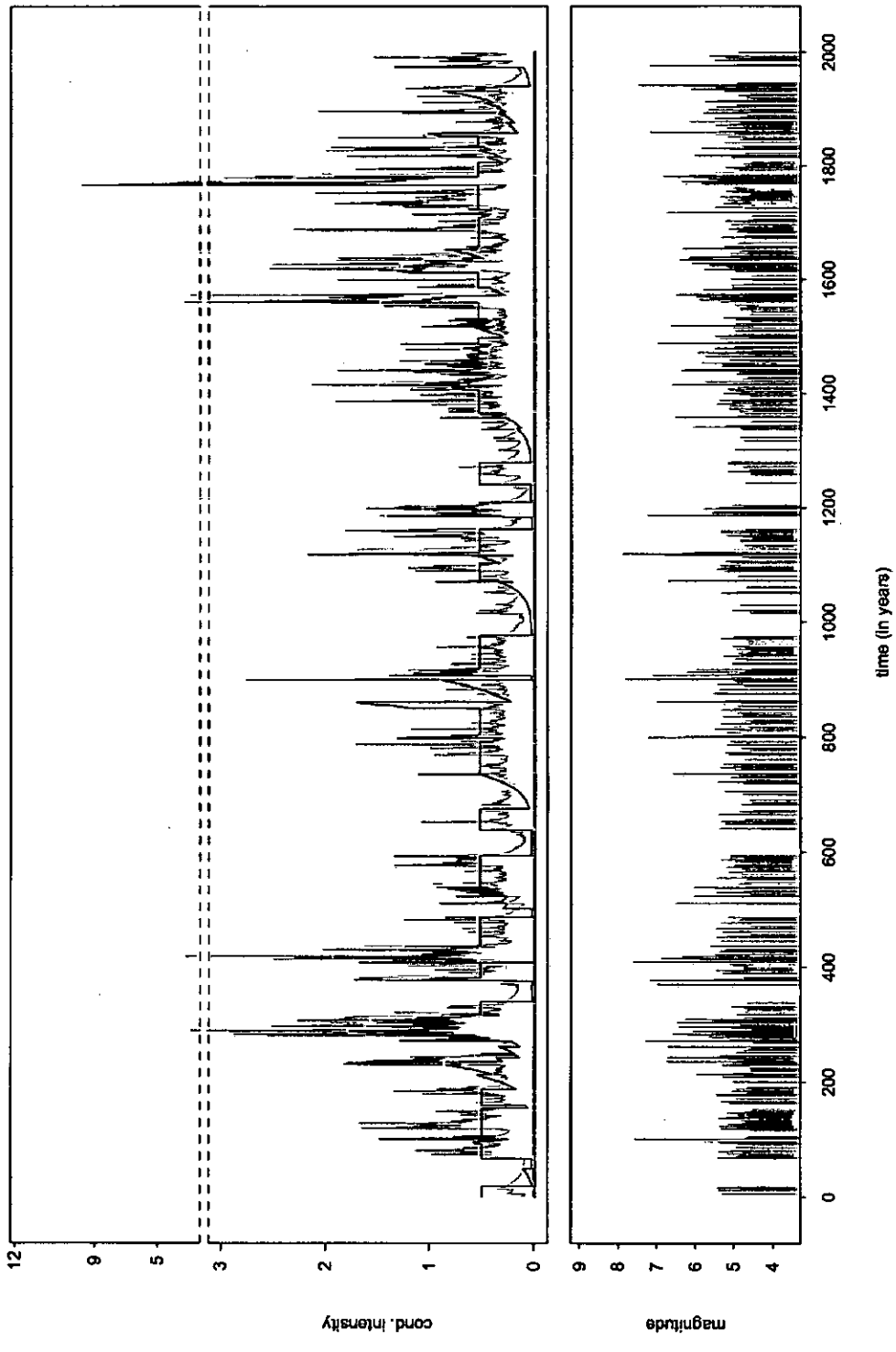

Figure 4.10: First step - Run 3 - Conditional intensity function related to the simulation of the state-space model (black line) and mixture of the plug-in estimates of the conditional intensity functions (violet line) with weights given by the estimated filtering probabilities.

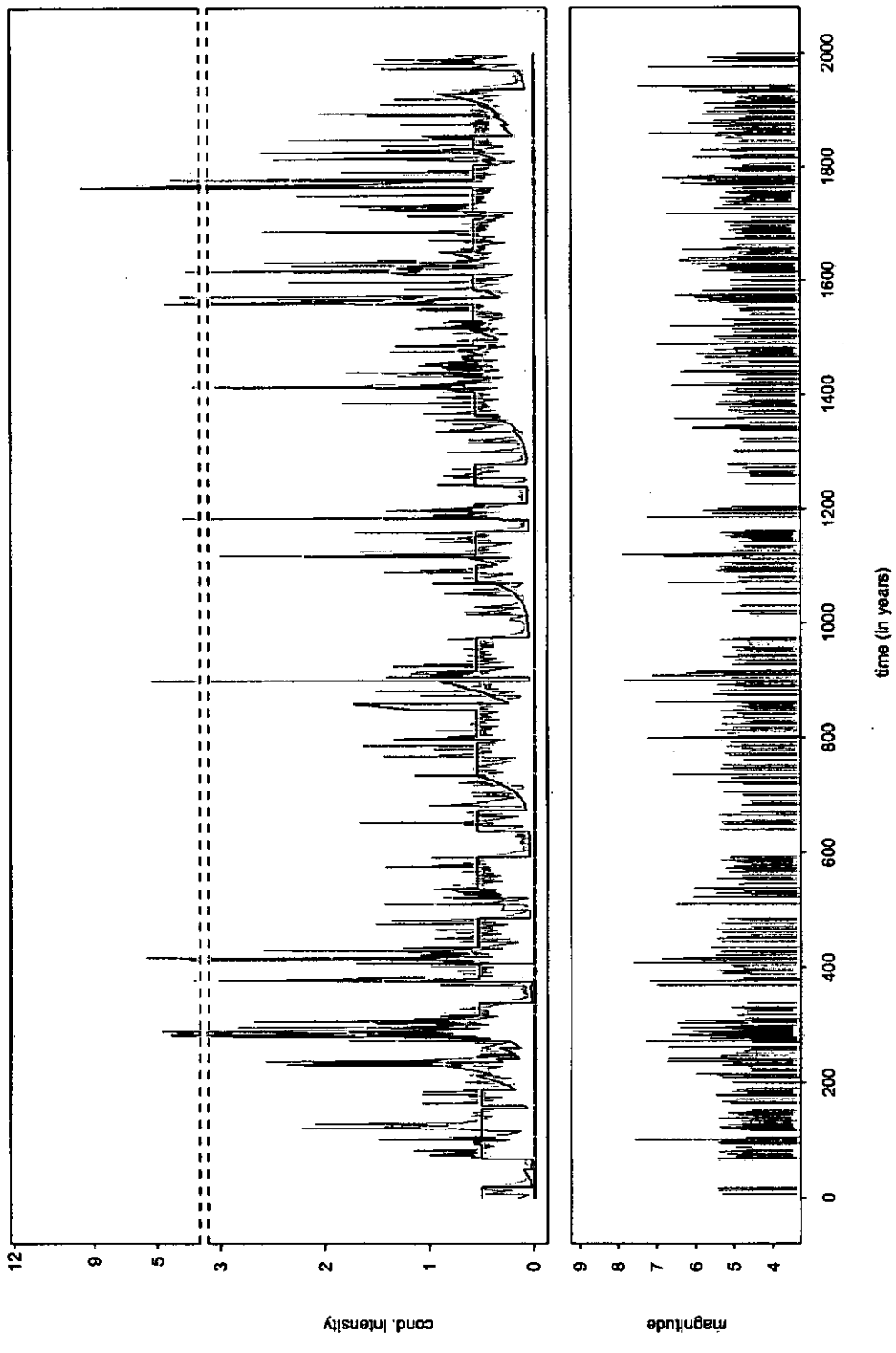

Figure 4.11: First step - Conditional intensity function related to the simulation of the state-space model (black line) and mixture of the plug-in estimates of the conditional intensity functions (violet line) obtained by using the average of the means as estimator of each parameters.

4.5.4 Second step

In the second step I performed the *Algorithm 2* which implements the resampling method, as the *Algorithm 1*, but without the application of the Kün $\cdot 10^{-2}$thod.
I executed two runs of such a procedure, where the number m of particles is 50000 and the resampling method is applied when the effective sample size is less that 40000.
The prior and importance distributions are again set equal. Mean and variance of these distributions for the point process parameters are calculated by exploiting their estimates got at the first step and are reported in Table 4.5; prior and importance distributions for the hidden process parameters are kept equal to those in Table 4.2.

parameter	distribution	mean	variance
μ	Gamma	0.4596	0.0231
α	Normal	-4.7036	2.6775
β	Gamma	0.0050	$7.682 \cdot 10^{-5}$
ρ	Gamma	2.8549	0.1482
k	Gamma	0.2882	0.6048
γ	Gamma	0.4316	0.2880
c	Gamma	0.2739	0.0225
p	Gamma	0.9683	0.0066

Table 4.5: Prior and importance distribution for the parameters of the point process models at the second step.

Table 4.6 contains the estimated posterior mean and standard deviation of each parameter and the corresponding average of the means; in Figure 4.12 the prior distribution (violet), the estimated posterior distribution (red and ruled) and the true value (green) of each parameter are compared. Finally, Figure 4.13 represents the root mean square error of each parameter, that is the square root of the mean square error between the true and the estimated value of each parameter up to t as varying the time during the sequential procedure.
A weak point of the present sequential analysis consists in the choice of the importance distributions equal to the prior distributions. I think that it is worth initially trying this choice, because of the complexity of the proposed model. In the future I plan to investigate the possibility of a choice of the importance distribution that is close to the target distribution, even though it has an iterative expression for the updating of the weights computationally easy to solve.

	$\hat{\mu}$		$\hat{\alpha}$		$\hat{\beta}$	
run	mean	st. dev.	mean	st. dev.	mean	st. dev.
1	0.2676	0.	-7.4669	0.	$1.0405 \cdot 10^{-2}$	$1.06 \cdot 10^{-8}$
2	0.6419	0.	-4.2607	0.	$1.9571 \cdot 10^{-3}$	$7.40 \cdot 10^{-10}$
	0.4548		-5.8632		$6.1810 \cdot 10^{-3}$	

	$\hat{\rho}$		$\hat{k}$		$\hat{\gamma}$	
run	mean	st. dev.	mean	st. dev.	mean	st. dev.
1	3.8294	0.	$6.2228 \cdot 10^{-6}$	0.	0.3076	$6.94 \cdot 10^{-8}$
2	2.8409	0.	0.1360	0.	$7.0164 \cdot 10^{-2}$	0.
	3.3351		$6.8014 \cdot 10^{-2}$		0.18888	

	$\hat{c}$		$\hat{p}$		$\hat{q_{12}}$	
run	mean	st. dev.	mean	st. dev.	mean	st. dev.
1	0.3457	$4.16 \cdot 10^{-7}$	0.9854	0.	$1.4477 \cdot 10^{-2}$	$1.56 \cdot 10^{-8}$
2	0.5318	$4.92 \cdot 10^{-7}$	0.9821	$1.08 \cdot 10^{-6}$	$7.5966 \cdot 10^{-2}$	$5.63 \cdot 10^{-8}$
	0.4387		0.9837		$4.5221 \cdot 10^{-2}$	

	$\hat{q_{13}}$		$\hat{q_{21}}$		$\hat{q_{23}}$	
run	mean	st. dev.	mean	st. dev.	mean	st. dev.
1	$3.6659 \cdot 10^{-3}$	$4.17 \cdot 10^{-9}$	$1.4815 \cdot 10^{-2}$	0.	$4.8643 \cdot 10^{-3}$	0.
2	$9.6831 \cdot 10^{-2}$	$1.02 \cdot 10^{-7}$	$2.3789 \cdot 10^{-3}$	0.	$2.8710 \cdot 10^{-2}$	$2.52 \cdot 10^{-8}$
	$5.0248 \cdot 10^{-2}$		$8.5970 \cdot 10^{-3}$		$1.6787 \cdot 10^{-2}$	

	$\hat{q_{31}}$		$\hat{q_{32}}$	
run	mean	st. dev.	mean	st. dev.
1	$8.8503 \cdot 10^{-2}$	$1.13 \cdot 10^{-7}$	$3.4222 \cdot 10^{-4}$	$3.01 \cdot 10^{-10}$
2	$6.3069 \cdot 10^{-2}$	$4.79 \cdot 10^{-8}$	$1.6415 \cdot 10^{-2}$	0.
	$7.5786 \cdot 10^{-2}$		$8.3785 \cdot 10^{-3}$	

Table 4.6: Estimated posterior means and standard deviations of the parameters at the second step; the average of the means are reported at the end of each table.

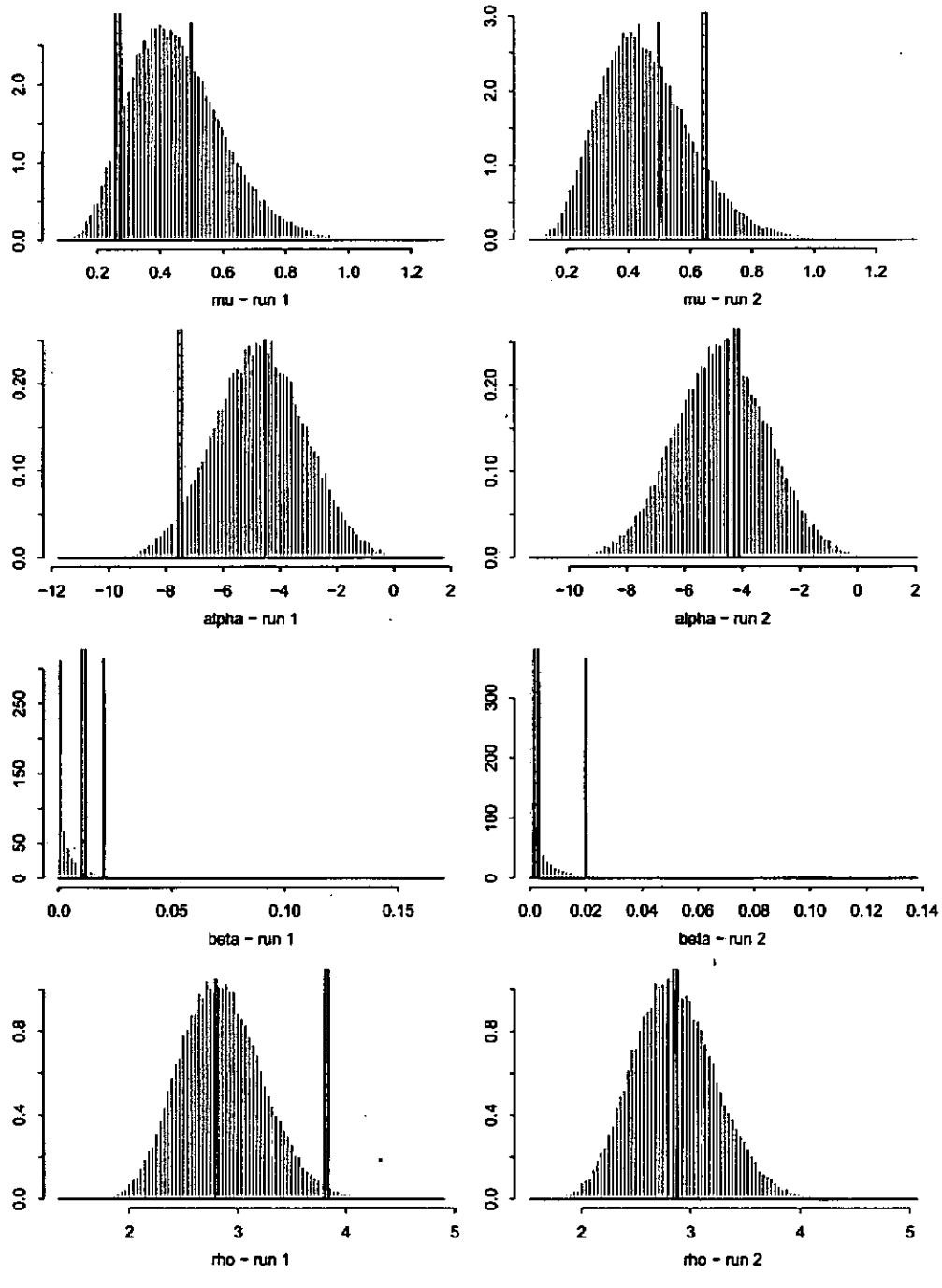

Figure 4.12: Second step of the sequential procedure - for each parameter prior (violet histogram) and scaled posterior (ruled histogram) particle set, value used in the simulation (green bar).

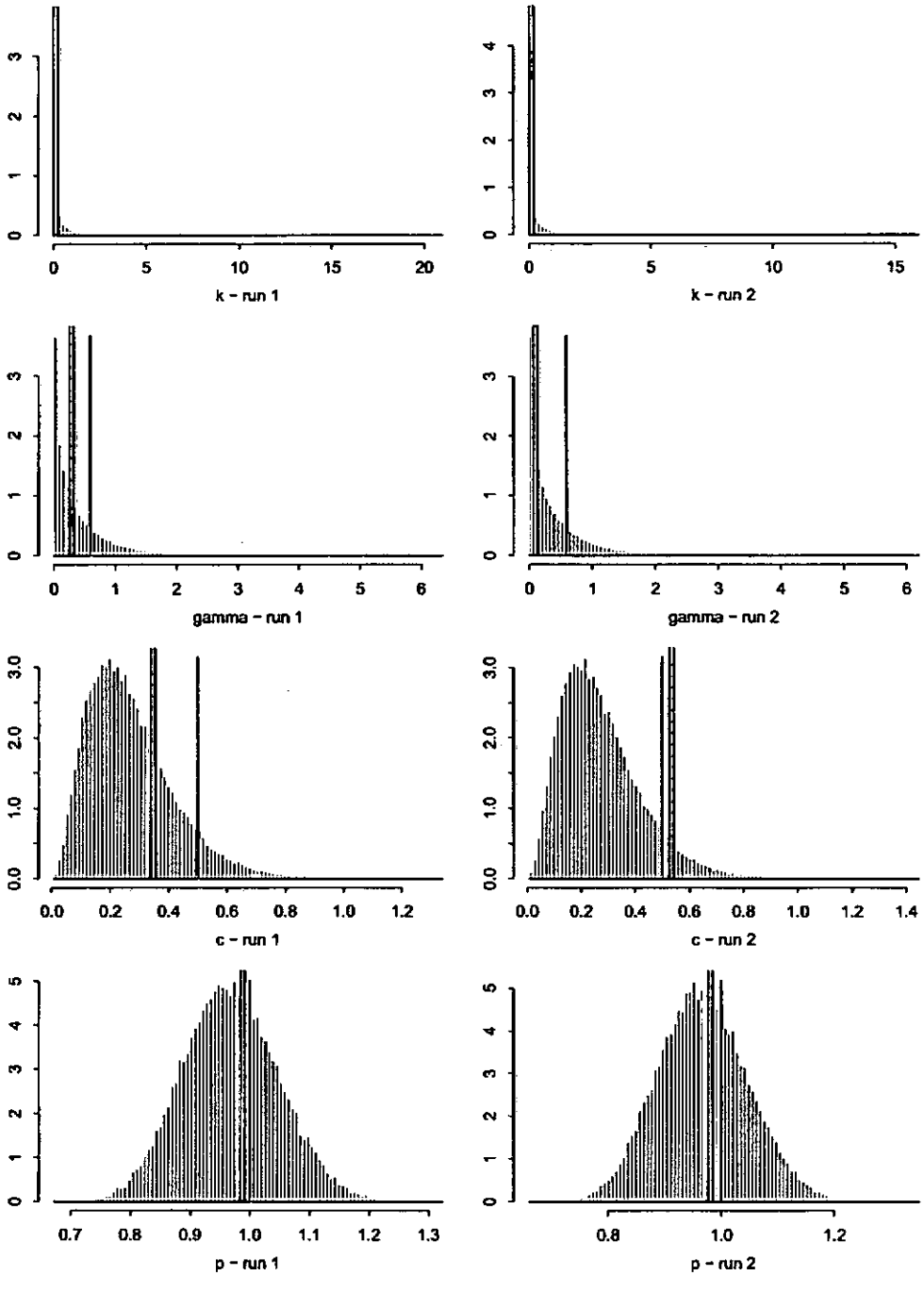

Figure 4.12: to be continued.

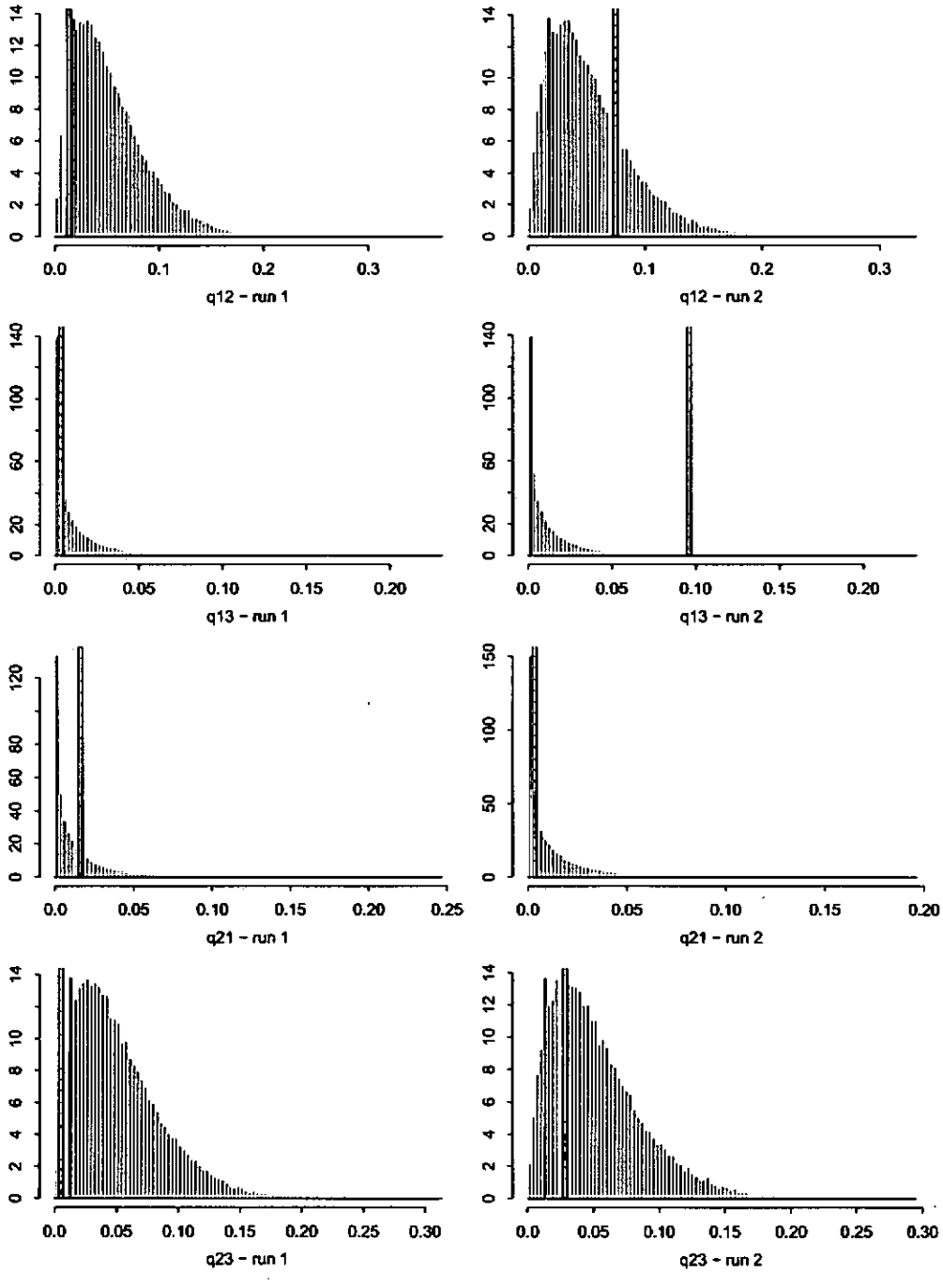

Figure 4.12: to be continued.

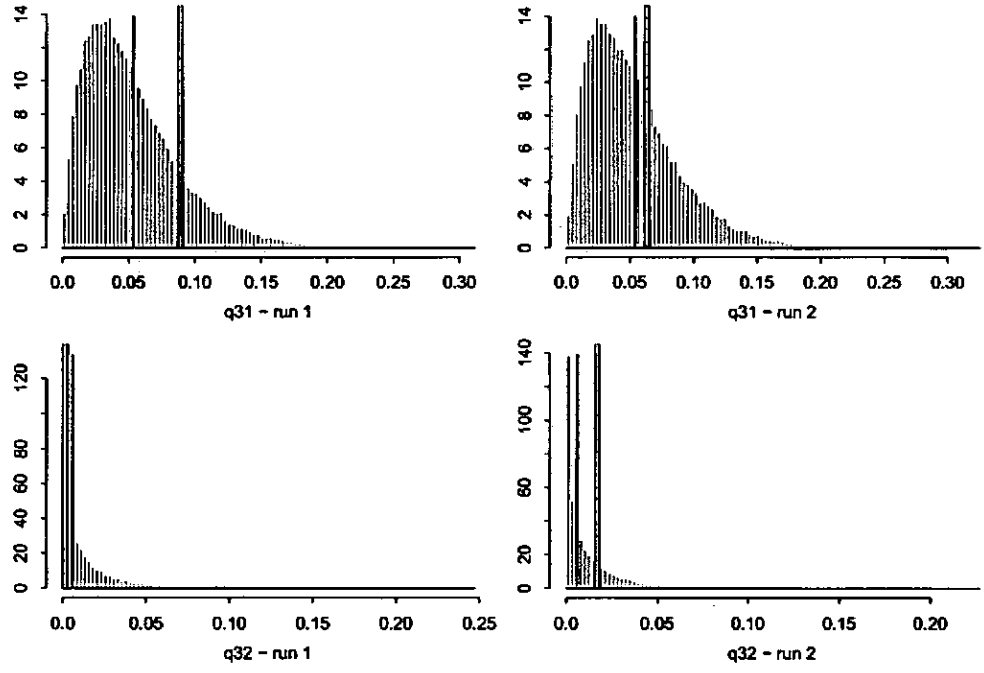

Figure 4.12: to be continued.

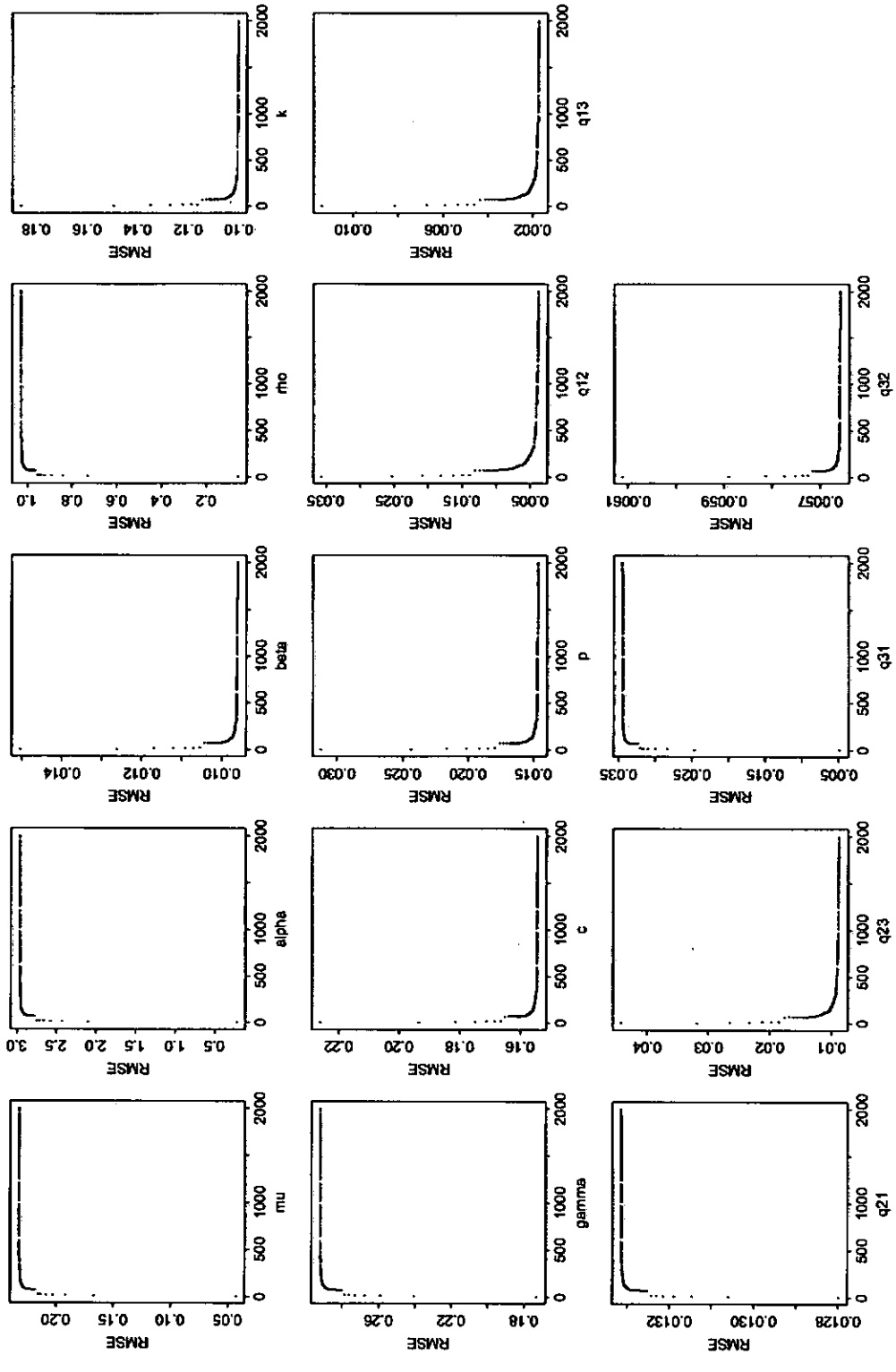

Figure 4.13: Second step - Run 1 - Root mean square error for each parameter.

4.5.5 Final results

The global averages in Table 4.6 constitute the final result of the multi-step procedure of estimation presented in the previous sections.
First I substituted the parameter estimates in *Algorithm 3* to refine the approximation of the filtering distribution through the sequential Monte Carlo method; the new estimate so obtained is represented in Figure 4.14, while Figure 4.15 shows the maximum of the filtering probability at each time t and Figure 4.16 points out the times when the simulated state and the state with the maximum filtering probability disagree. Finally, the mixture of the plug-in estimates of the conditional intensity functions is given in Figure 4.17.
Comparing 4.11 and 4.17, I observe that

Fig. 4.11 – the global behaviour of the estimated state-space model is very similar to that of the simulated model when the rejuvenation method by Künsch-Hürzeler is applied;

Fig. 4.17 – the fitting of the estimate produced by the standard particle filtering procedure appears not homogeneous.

I think that two critical aspects emerge from this comparison: the extremely large number of iterations required by the sequential procedure and the fast impoverishment of the particle set. In the literature some recipes have been proposed among which the Künsch-Hürzeler method seems to be the easiest to apply in general conditions, but its theoric properties must be studied in depth.

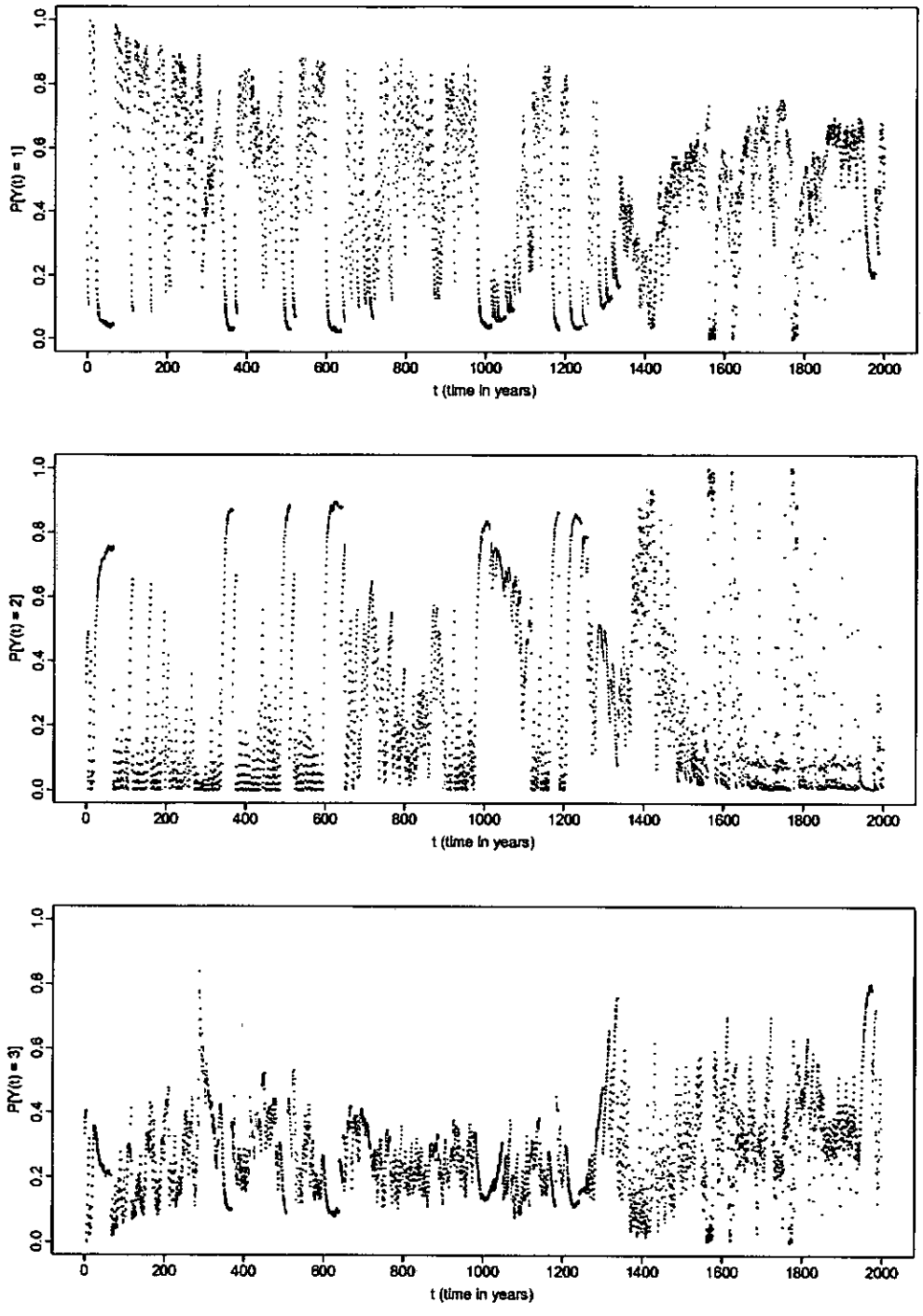

Figure 4.14: Filtering probability of each state.

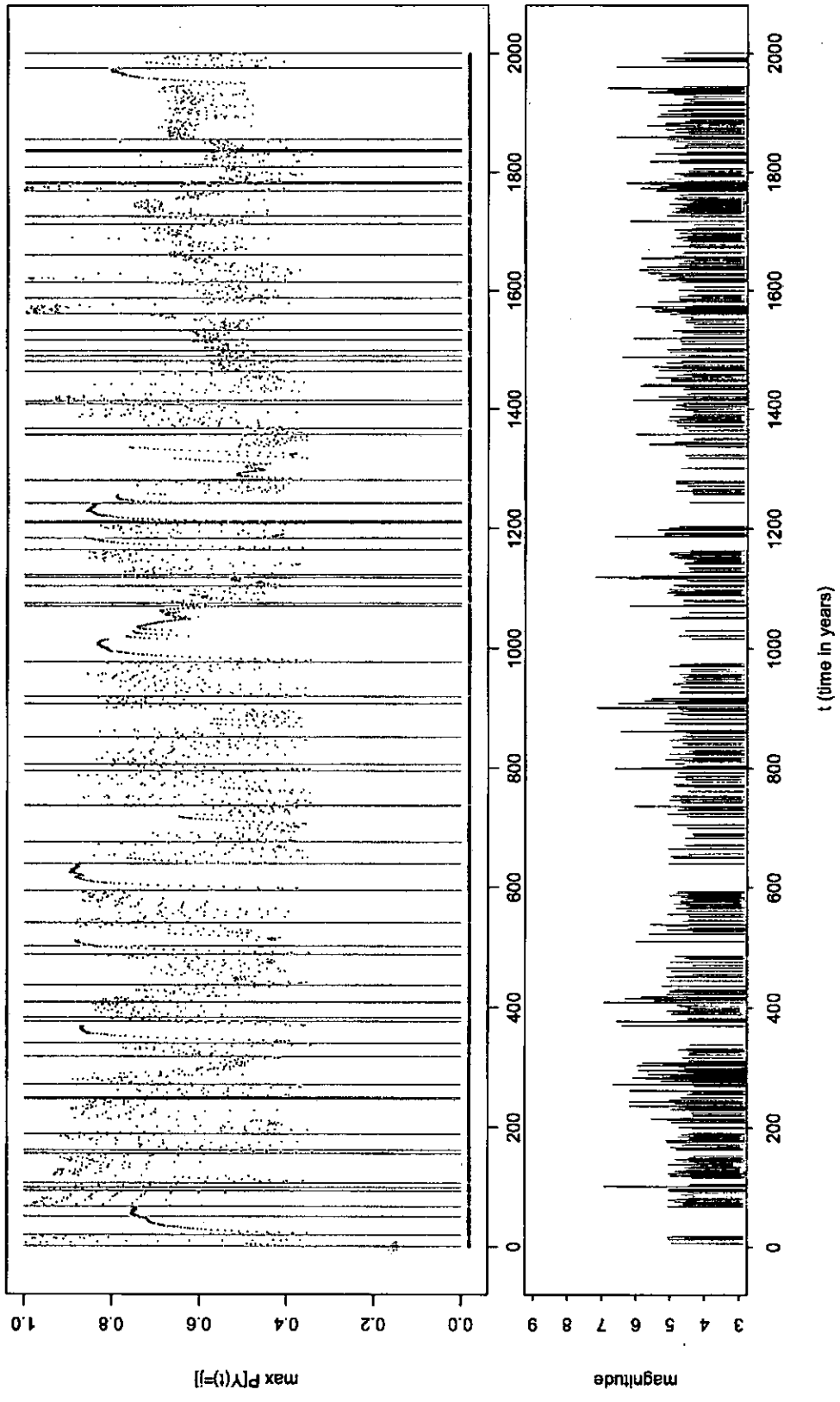

Figure 4.15: Maximum of the filtering probability of the states at each time t.

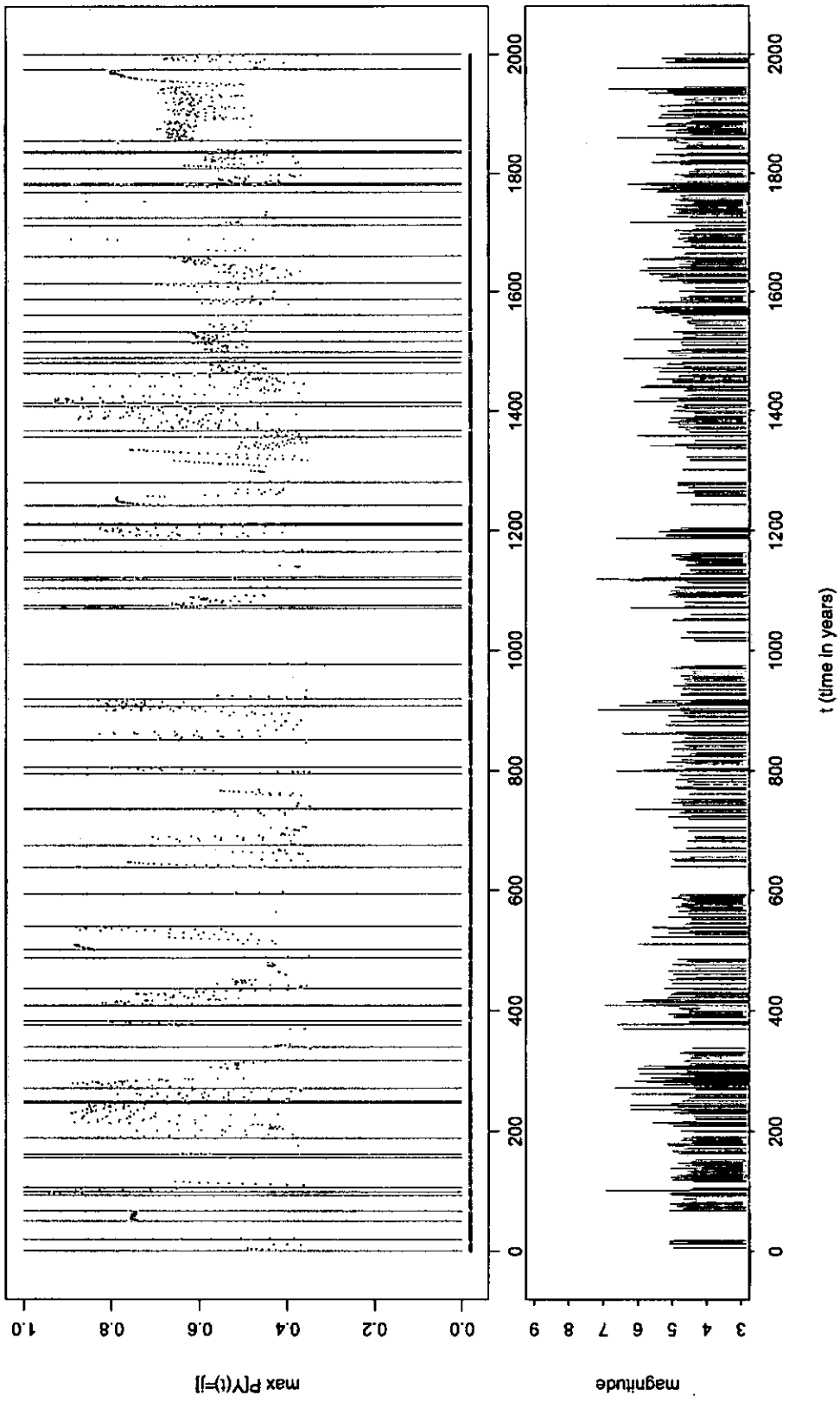

Figure 4.16: Maximum filtering probability at times t when there is not agreement between the state with maximum probability and the simulated state.

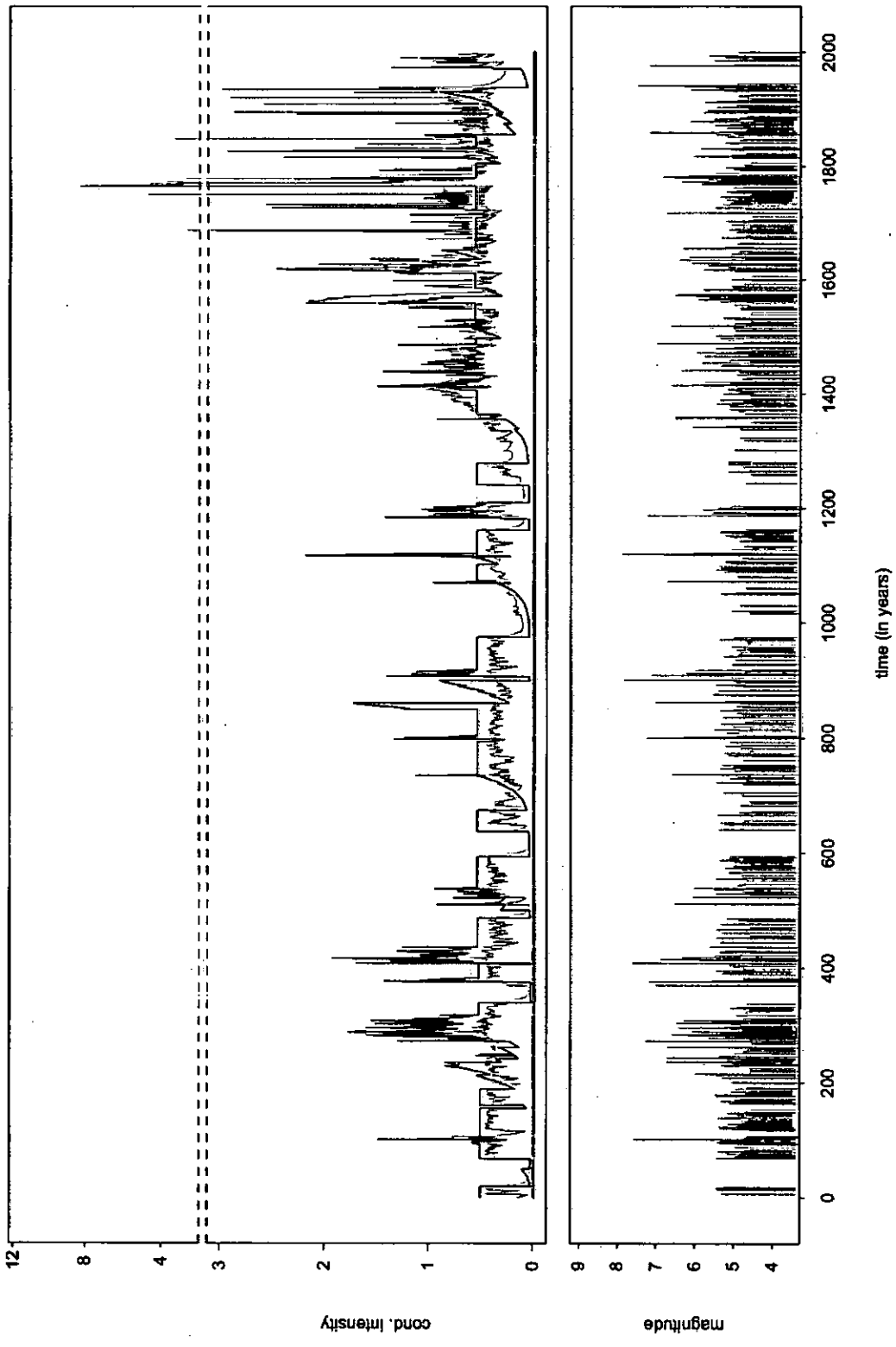

Figure 4.17: Conditional intensity function related to the simulation of the state-space model (black line) and mixture of the plug-in estimates of the conditional intensity functions (violet line) obtained by using the average of the means as estimator of each parameters.

Appendix A : NT4.1.1 and ZS4

The following table contains earthquakes of the NT4.1.1 catalogue (a parametric catalogue of damaging earthquakes in the Italian area) having magnitude equal to or greater than 5.0 and occurred between 1600 and 1992 in the zones from 56 to 80 of the Seismogenic zonation ZS4.

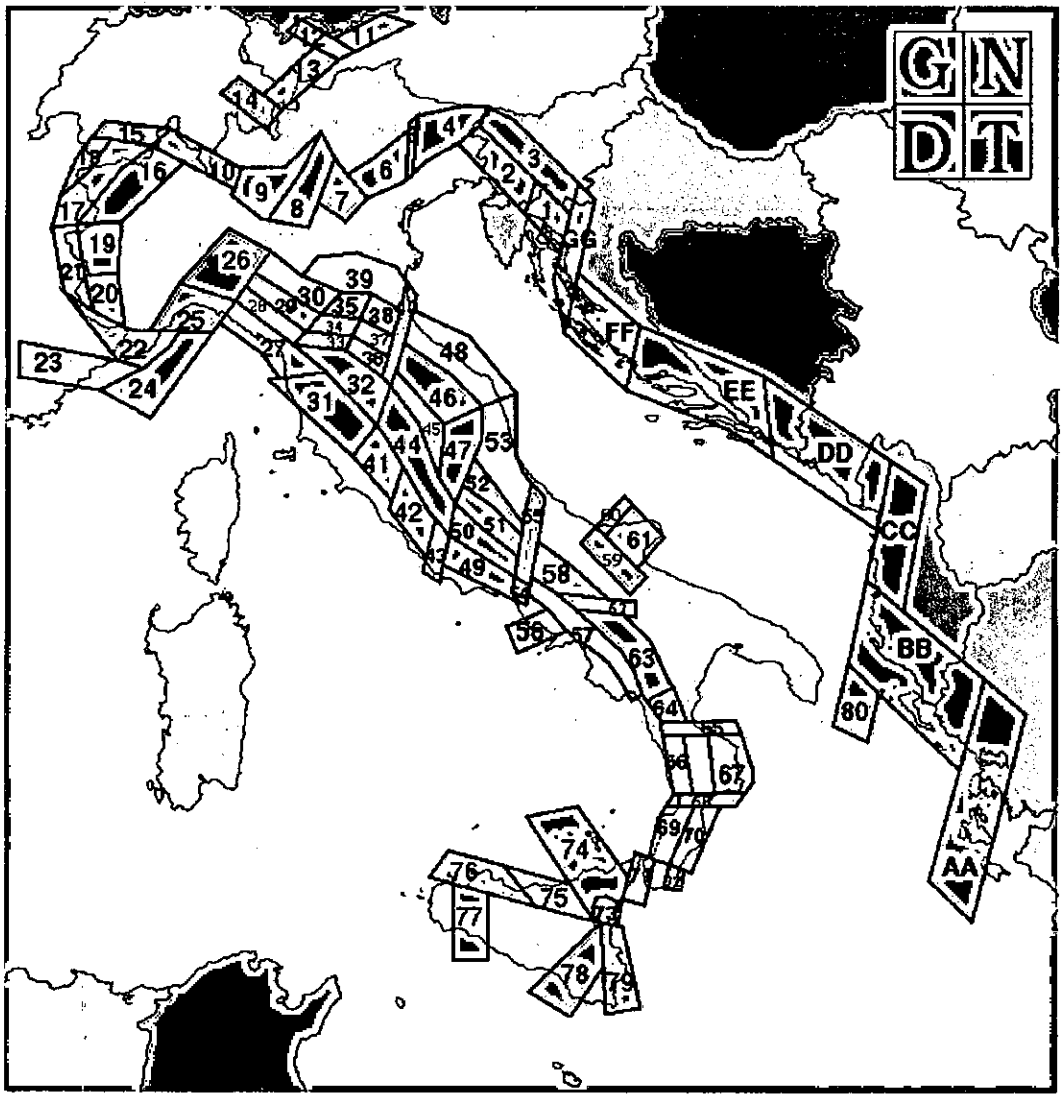

Figure 4.18: Seismogenic zonation ZS4. (The illustration is courtesy of the National Institute of Geophysics and Volcanology)

year	month	day	hour:min:sec	latitude	longitude	magnitude	zone
1609	7	20	00:00:00	38.967	16.350	5.5	66
1613	0	0	00:00:00	38.117	14.783	5.9	74
1619	1	5	00:00:00	39.000	16.500	5.0	66
1621	8	9	00:00:00	39.417	16.083	5.0	66
1624	10	3	00:00:00	37.267	14.733	5.5	78
1626	3	27	00:00:00	38.817	16.417	6.2	68
1627	7	30	00:00:00	41.733	15.267	7.0	59
1631	12	17	00:00:00	40.833	14.417	5.0	56
1638	3	27	15:10:00	39.083	16.283	7.3	66
1638	6	9	00:20:00	39.200	16.817	6.4	67
1646	5	31	00:00:00	41.833	16.000	6.4	61
1657	1	0	00:00:00	41.833	15.333	5.0	60
1659	11	5	00:00:00	38.700	16.333	6.4	69
1687	10	2	00:00:00	38.667	15.900	5.0	69
1688	6	5	16:15:00	41.317	14.567	7.3	58
1693	1	11	00:00:00	37.443	15.192	7.0	79
1693	1	11	00:00:00	37.203	14.762	5.9	78
1694	9	8	11:30:00	40.900	15.433	7.0	63
1698	4	12	00:00:00	37.217	14.767	5.0	78
1702	3	14	05:15:00	41.117	14.950	6.4	62
1707	3	3	00:00:00	38.500	16.250	5.0	70
1708	1	26	00:00:00	39.900	16.200	5.2	64
1714	8	0	00:00:00	40.750	14.750	5.0	57
1716	12	1	00:00:00	37.500	15.083	5.0	79
1717	4	22	00:00:00	38.100	15.267	5.2	74
1724	8	3	21:30:00	38.250	15.667	5.0	71
1726	9	1	00:00:00	38.117	13.350	5.9	76
1727	1	7	00:00:00	36.883	15.067	5.2	79
1727	7	4	00:00:00	37.500	13.000	5.0	77
1731	3	20	02:00:00	41.317	15.800	6.2	59
1732	11	29	12:30:00	41.083	15.117	6.4	62
1736	8	16	00:00:00	37.900	13.567	5.2	76
1739	2	13	00:00:00	41.500	15.500	5.0	59
1739	5	10	16:00:00	38.117	14.783	5.9	74
1740	6	13	00:00:00	37.500	13.000	5.2	77
1743	2	20	16:30:00	39.667	19.000	7.0	80
1743	12	7	00:00:00	38.417	15.833	5.2	69
1767	7	14	00:00:00	39.167	16.500	5.2	66
1777	6	6	16:15:00	38.167	16.167	5.2	70
1779	8	8	20:00:00	40.750	14.250	5.0	56

year	month	day	hour:min:sec	latitude	longitude	magnitude	zone
1783	2	5	00:00:00	38.267	15.917	7.3	69
1783	2	7	00:00:00	38.583	16.217	7.0	69
1783	3	1	00:00:00	38.700	16.250	5.9	69
1783	3	28	00:00:00	38.800	16.467	6.7	68
1783	11	15	09:15:00	41.667	15.333	5.0	59
1783	11	20	12:58:00	38.600	16.100	5.0	69
1784	3	20	13:30:00	38.600	16.100	5.0	69
1784	10	14	00:00:00	38.300	16.200	5.0	70
1786	3	9	00:00:00	38.100	15.000	5.9	74
1791	10	13	01:20:00	38.659	16.248	6.2	69
1794	6	12	20:30:00	41.000	15.000	5.0	62
1794	6	15	21:30:00	40.750	14.417	5.0	56
1796	3	18	00:00:00	40.750	13.917	5.0	56
1805	7	26	21:01:00	41.500	14.533	6.7	58
1805	10	13	22:00:00	41.167	14.333	5.0	57
1807	11	11	00:00:00	40.300	15.717	5.0	63
1818	2	20	18:15:00	37.616	15.099	6.2	79
1818	2	28	00:00:00	37.250	14.700	5.2	78
1818	9	8	00:00:00	37.833	14.083	5.2	75
1819	2	24	00:00:00	37.917	14.083	5.2	75
1821	8	2	00:00:00	38.933	16.400	5.2	66
1823	3	5	16:37:00	38.133	14.750	5.6	74
1823	3	5	16:37:00	38.051	13.692	5.9	76
1824	12	11	00:00:00	39.583	16.633	5.2	65
1826	2	1	16:00:00	40.583	15.683	5.2	63
1828	2	2	00:00:00	40.750	13.883	5.2	56
1828	3	12	00:00:00	38.350	15.850	5.0	69
1831	1	2	14:07:00	40.067	15.800	5.5	63
1832	3	8	00:00:00	39.050	16.917	6.4	67
1835	10	12	00:00:00	39.317	16.317	5.9	66
1836	4	25	00:00:00	39.567	16.700	6.4	65
1836	5	4	00:00:00	38.100	15.650	5.0	71
1836	11	20	07:27:00	40.150	15.800	5.5	63
1841	2	21	00:00:00	41.717	15.633	5.2	61
1841	3	6	12:00:00	40.750	13.917	5.0	56
1841	3	20	00:00:00	38.100	15.650	5.0	71
1848	1	11	00:00:00	37.217	15.233	5.5	79
1850	1	1	00:00:00	37.633	14.850	5.0	74
1851	8	14	13:20:00	40.950	15.650	6.4	62
1852	1	23	07:03:00	38.100	15.650	5.0	71

year	month	day	hour:min:sec	latitude	longitude	magnitude	zone
1852	5	13	21:40:00	38.100	15.650	5.0	71
1853	4	9	12:45:00	40.850	15.250	6.2	63
1854	2	12	17:50:00	39.270	16.273	6.4	66
1855	9	20	00:00:00	39.300	16.250	5.0	66
1857	12	16	00:00:00	40.350	15.833	7.0	63
1858	8	6	12:15:00	40.750	15.550	5.0	63
1864	12	28	12:00:00	41.833	15.583	5.0	61
1869	3	31	13:00:00	41.717	15.750	5.0	61
1870	10	4	00:00:00	39.250	16.333	6.4	66
1871	8	1	22:00:00	41.917	15.633	5.0	61
1872	10	8	13:25:00	39.300	16.250	5.0	66
1873	9	11	09:03:00	39.300	16.250	5.0	66
1874	8	30	22:30:00	37.883	14.950	5.0	74
1875	12	6	00:00:00	41.683	15.683	5.2	61
1876	5	25	03:35:00	37.817	13.300	5.0	77
1876	9	13	23:01:00	38.100	15.650	5.0	71
1878	10	4	00:46:00	37.300	14.700	5.0	78
1881	3	4	12:15:00	40.750	13.917	5.5	56
1882	6	6	05:40:00	41.500	14.183	5.2	58
1883	7	28	20:25:00	40.750	13.886	6.2	56
1884	1	10	20:35:00	37.833	15.167	5.0	74
1885	9	17	09:35:00	41.133	14.800	5.0	62
1885	12	26	00:00:00	41.550	14.650	5.0	58
1886	3	6	00:00:00	39.344	16.172	5.2	66
1887	12	3	03:45:00	39.540	16.215	5.9	66
1889	5	29	09:00:00	38.067	16.100	5.0	72
1889	6	30	21:15:00	38.583	14.583	5.0	74
1889	12	8	00:00:00	41.750	15.583	5.0	61
1891	6	27	12:11:48	38.550	14.833	5.0	74
1892	3	16	12:38:00	38.533	14.250	5.2	74
1893	1	25	11:21:00	40.583	15.417	5.0	63
1893	8	10	00:00:00	41.700	16.050	5.2	61
1894	3	25	00:00:00	41.867	15.350	5.0	60
1894	5	28	20:15:00	39.970	16.063	5.0	64
1894	8	3	06:29:26	38.550	14.833	5.0	74
1894	11	16	17:52:00	38.278	15.903	5.9	69
1894	12	27	00:00:00	38.467	14.500	5.0	74
1903	5	4	03:44:00	41.033	14.533	5.0	57
1905	9	8	01:43:00	38.754	16.026	7.5	69
1905	11	26	00:00:00	41.133	15.033	5.1	62

year	month	day	hour:min:sec	latitude	longitude	magnitude	zone
1906	9	11	19:03:25	38.000	13.600	5.0	76
1907	3	20	13:28:06	41.600	14.517	5.0	58
1907	5	8	00:00:00	37.500	15.000	5.0	79
1907	10	23	20:28:00	38.155	16.024	5.9	72
1908	6	30	05:44:00	38.600	14.500	5.0	74
1908	8	15	10:00:00	37.900	14.300	5.0	75
1908	12	10	06:20:00	38.017	15.083	5.0	74
1908	12	28	04:20:00	38.133	15.667	7.3	71
1909	7	1	06:24:00	38.167	15.617	5.5	71
1909	11	20	12:50:00	38.167	15.583	5.0	71
1909	12	3	11:50:00	37.900	13.100	5.0	77
1910	6	7	02:04:00	40.979	15.320	5.9	63
1910	11	18	02:42:48	38.167	15.583	5.0	71
1913	6	28	08:53:00	39.580	16.190	5.5	65
1913	10	4	18:26:00	41.483	14.667	5.2	58
1914	12	19	03:50:42	41.583	14.250	5.0	58
1926	8	17	01:42:00	38.567	14.832	5.0	74
1928	3	7	10:55:00	38.633	15.983	5.9	69
1930	7	23	00:08:00	41.050	15.300	6.7	62
1932	1	2	23:36:00	39.100	17.290	5.5	67
1940	1	15	13:19:00	38.033	13.433	5.1	76
1941	8	20	10:36:32	41.700	15.400	5.1	59
1941	9	7	11:13:00	41.200	15.000	5.0	58
1947	5	11	06:32:00	38.712	16.581	5.6	70
1948	8	18	21:12:00	41.700	15.783	5.4	61
1950	4	8	00:00:00	37.717	15.167	5.1	73
1950	4	10	03:55:00	38.183	15.550	5.0	71
1954	8	6	19:21:12	40.667	15.883	5.0	63
1962	8	21	18:19:00	41.167	14.967	6.2	58
1963	2	13	12:45:00	40.550	15.667	5.0	63
1967	10	31	21:08:00	37.850	14.367	5.1	75
1968	1	15	02:01:00	37.750	12.967	5.9	77
1974	10	20	11:25:53	39.600	18.900	5.4	80
1978	3	11	19:20:44	37.967	16.183	5.0	72
1978	4	15	23:33:00	38.150	14.983	6.1	74
1980	5	28	19:51:00	38.417	14.283	5.6	74
1980	11	23	18:34:00	40.800	15.267	6.9	63
1990	12	13	00:24:00	37.270	15.070	5.3	79

Appendix B : CPTI and ZS7

The following table contains earthquakes of the CPTI catalogue (July 2003 version of the Parametric Catalog of the Italian Earthquakes) having magnitude equal to or greater than 4.0 and occurred in the zone 729 of seismogenic zonation ZS7 between 1600 and 1992.

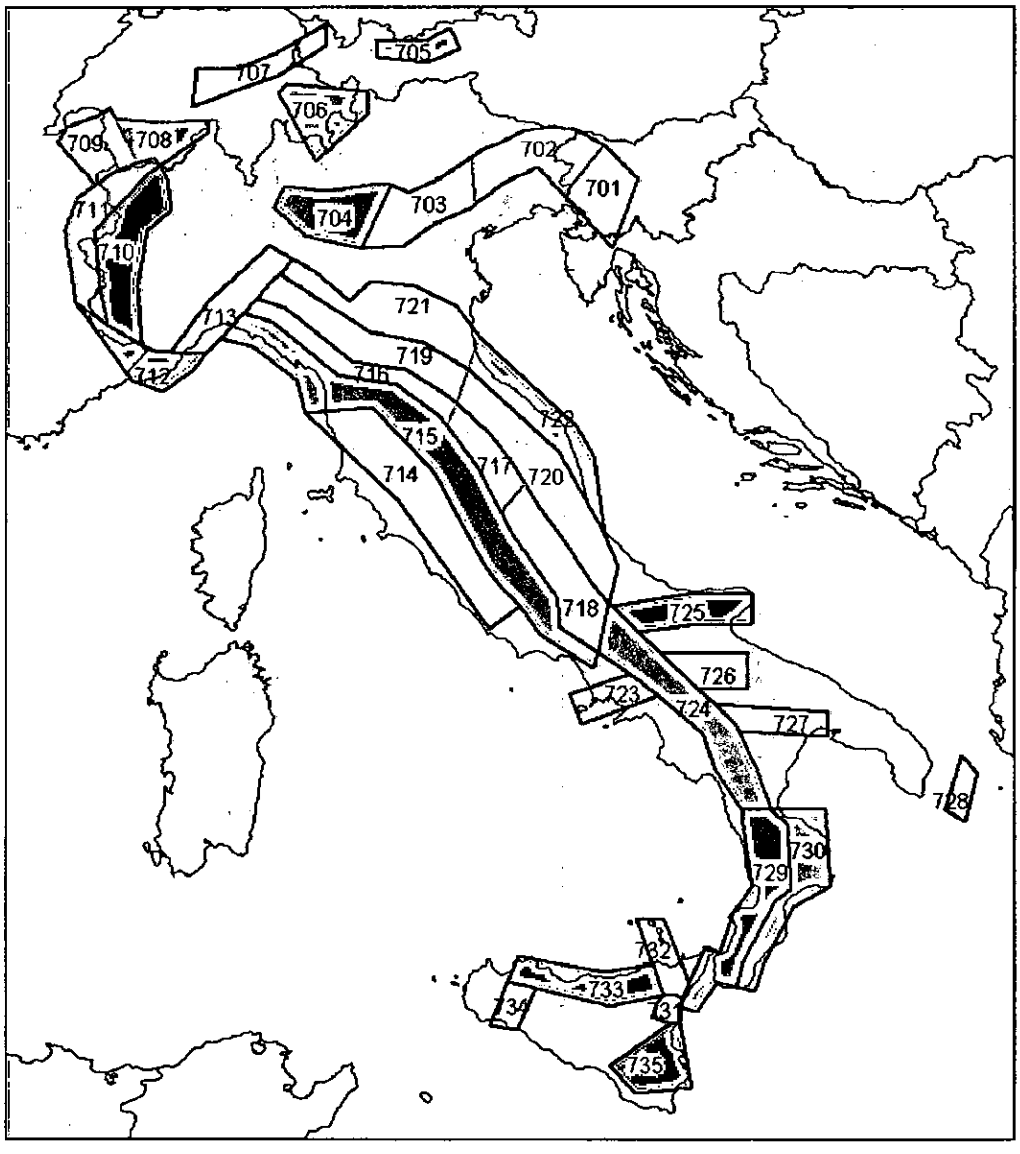

Figure 4.19: Seismogenic zonation ZS7. (The illustration is courtesy of the National Institute of Geophysics and Volcanology)

year	month	day	hour:min	latitude	longitude	magnitude
1601	8	10	07:00	38.25	15.67	4.30
1609	7	20	00:00	38.97	16.35	5.40
1619	1	5	00:00	39.00	16.50	4.80
1621	8	9	00:00	39.42	16.08	4.80
1626	4	4	12:45	38.82	16.42	6.00
1635	8	12	00:00	38.19	15.55	4.60
1638	3	27	15:05	39.03	16.28	7.10
1649	1	0	00:00	38.19	15.55	4.60
1659	11	5	22:15	38.70	16.25	6.60
1687	10	2	00:00	38.67	15.90	4.80
1706	3	19	00:00	38.20	15.64	4.00
1712	7	16	00:00	38.20	16.03	4.00
1715	2	21	00:00	38.23	15.81	4.00
1724	8	3	21:30	38.25	15.67	4.80
1728	5	0	00:00	39.00	16.25	4.30
1735	9	6	00:00	38.68	16.13	4.30
1743	12	7	00:00	38.58	16.14	5.10
1747	9	0	00:00	38.17	15.67	4.30
1767	7	14	01:05	39.38	16.28	5.80
1770	6	8	00:00	38.11	15.65	4.60
1777	6	6	16:15	38.98	15.62	5.10
1780	3	28	00:00	38.19	15.55	4.30
1783	2	5	12:00	38.30	15.97	7.10
1783	2	6	00:20	38.22	15.63	5.80
1783	2	7	13:10	38.58	16.20	6.80
1783	3	1	01:40	38.77	16.30	6.00
1783	3	28	18:55	38.78	16.47	6.60
1783	11	20	12:58	38.60	16.10	4.80
1784	3	20	13:30	38.60	16.10	4.80
1785	3	17	03:33	38.10	15.65	4.60
1785	4	13	06:00	38.60	16.10	4.60
1786	2	23	07:40	38.60	16.10	4.30
1787	9	19	22:00	38.18	15.55	4.60
1790	6	0	00:00	38.17	16.00	4.30
1791	10	13	01:20	38.63	16.27	6.00
1817	10	19	02:30	38.17	15.50	4.30
1821	8	2	00:00	38.94	16.45	5.10
1824	12	11	00:00	39.54	16.59	5.10
1828	3	12	00:00	38.53	16.00	4.80
1832	4	12	00:00	38.77	16.27	4.30

year	month	day	hour:min	latitude	longitude	magnitude
1835	10	12	22:35	39.33	16.30	6.00
1836	5	4	00:00	38.10	15.65	4.80
1839	8	18	01:00	39.30	16.25	4.30
1839	8	27	11:00	38.10	15.65	4.30
1840	4	24	01:30	38.10	15.65	4.30
1841	3	20	00:00	38.10	15.65	4.80
1841	8	15	00:30	38.10	15.65	4.30
1848	10	7	03:00	38.10	15.65	4.30
1851	2	15	00:00	38.90	16.60	4.00
1851	4	11	06:00	38.17	15.58	4.30
1852	1	23	07:03	38.10	15.65	4.80
1852	5	13	21:40	38.10	15.65	4.80
1854	2	12	17:50	39.25	16.30	6.30
1855	9	20	00:00	39.30	16.25	4.80
1869	11	28	00:00	38.69	16.22	4.60
1870	10	4	16:55	39.22	16.33	6.30
1871	6	29	20:45	39.15	16.23	4.60
1872	10	8	13:25	39.30	16.25	4.80
1873	9	11	09:03	39.30	16.25	4.80
1876	9	13	23:01	38.10	15.65	4.80
1881	4	23	04:12	39.37	16.05	4.30
1882	3	10	00:00	38.00	15.50	4.30
1883	7	25	09:50	39.30	16.27	4.30
1884	5	4	20:13	38.60	16.10	4.30
1886	3	6	00:00	39.34	16.19	5.10
1887	12	3	03:45	39.57	16.22	5.40
1889	1	10	19:10	38.95	16.52	4.30
1889	10	5	13:52	38.68	15.90	4.00
1894	11	16	17:52	38.28	15.87	5.80
1895	3	9	22:01	38.28	15.57	4.00
1895	7	26	17:44	38.30	15.98	4.30
1895	11	18	16:32	38.27	15.55	4.00
1896	4	1	07:18	38.27	15.88	4.30
1898	4	21	05:30	39.10	16.40	4.30
1898	8	12	00:00	38.18	15.39	4.60
1901	6	20	17 30	38.90	16.60	4.30
1901	12	13	00:09	38.60	16.10	4.00
1905	9	8	01:43	38.67	16.07	7.10
1908	3	1	05:23	39.13	16.31	4.00
1908	12	28	04:20	38.15	15.68	7.10

year	month	day	hour:min	latitude	longitude	magnitude
1909	7	1	06:24	38.15	15.60	5.40
1909	11	20	12:50	38.17	15.58	4.80
1909	12	8	23:45	38.37	16.08	4.00
1910	3	31	19:00	38.87	16.52	4.00
1910	6	13	23:30	38.17	15.58	4.30
1910	11	18	02:42	38.17	15.58	4.80
1910	12	12	06:07	38.60	16.03	4.00
1911	3	11	03:29	38.17	15.58	4.30
1911	6	18	16:59	38.17	15.58	4.30
1912	11	7	14:12	38.80	16.27	4.00
1912	12	22	08:05	38.17	15.58	4.80
1912	11	7	14:12	38.80	16.27	4.00
1912	12	22	08:05	38.17	15.58	4.80
1913	1	5	07:56	38.33	16.10	4.00
1913	6	27	17:00	38.98	16.27	4.30
1913	6	28	02:47	38.17	15.58	4.00
1913	6	28	08:53	39.53	16.23	5.40
1915	9	11	08:25	38.10	15.70	4.30
1918	11	24	13:38	39.28	16.32	4.30
1918	12	3	10:42	38.60	16.05	4.00
1919	3	18	14:17	38.17	15.58	4.00
1920	1	27	04:38	39.30	16.30	4.30
1921	6	19	12:52	38.30	16.00	4.30
1923	8	16	12:50	38.20	15.60	4.00
1928	3	7	10:55	38.54	16.04	5.10
1929	2	22	08:17	38.20	15.60	4.00
1932	5	22	17:01	37.91	15.46	3.75
1934	11	9	23:09	39.28	16.33	4.60
1936	4	7	05:41	38.72	16.20	4.00
1940	4	26	07:49	38.51	16.07	4.00
1941	5	28	00:00	38.27	15.87	4.30
1946	3	15	07:02	38.20	15.60	4.30
1947	6	29	20:20	39.30	16.25	4.30
1948	8	2	19:25	39.10	16.40	4.30
1949	12	9	13:45	38.12	15.58	4.30
1950	4	10	03:55	38.18	15.55	4.80
1950	12	9	14:45	38.00	15.48	4.30
1953	2	25	00:07	38.70	16.10	4.00
1955	7	23	21:05	39.58	16.42	4.30
1958	10	27	10:09	38.98	16.43	4.60

year	month	day	hour:min	latitude	longitude	magnitude
1961	3	24	10:36	38.17	15.65	4.00
1963	11	12	08:22	39.50	16.32	4.00
1965	10	1	18:38	39.25	16.25	4.00
1967	9	7	14:09	37.90	15.30	4.00
1968	7	17	19:11	38.70	16.20	4.00
1975	1	16	00:09	38.12	15.65	5.10
1980	12	9	05:50	38.73	16.17	0.00

List of Tables

List of Figures

Bibliography

[A77] AKAIKE H. *On entropy maximization principle*, in *Applications of Statistics*, pp. 27-41, ed. Krishnaiah P.R., North-Holland, Amsterdam (1977).

[B81] BREMAUD P. *Point processes and queues*, Springer-Verlag, New York (1981).

[B99] BREMAUD P. *Markov chains*, Springer-Verlag, New York (1999).

[BH03] BEBBINGTON M., HARTE D., *The linked stress release model for spatio-temporal seismicity: formulations, procedures and applications*, Geophysical Journal International, 154 (2003) 925-946.

[C01] CHOPIN N., *Sequential inference and state number determination for discrete state-space models through particle filtering*, CREST working paper 2001-34 (2001) pp. 1-23.

[CCF99] CARPENTER J., CLIFFORD P., FEARNHEAD P., *An improved particle filter for non-linear problems*, Radar, Sonar and Navigation, IEE Proceedings 146,1 (1999) 2-7.

[CR96] CASELLA G., ROBERT C., *Rao-Blackwellisation of sampling schemes*, Biometrika, 1 (1996) 91-94.

[D78] DAVISON C. *The founder of seismology*, Arno Press, New York (1978).

[DFG01] DOUCET A., DE FREITAS N., GORDON N., *Sequential Monte Carlo methods in practice*, Springer-Verlag, New York (2001).

[DMZ88] DE NATALE G., MUSMECI F., ZOLLO A., *A linear intensity model to investigate the casual relation between Calabrian and North-Aegean earthquake sequences*, Geophysical Journal, 95 (1988) 285-293.

[DVj88] DALEY D.J., VERE-JONES D., *An introduction to the theory of point processes*, Springer-Verlag, New York (1988).

[EK86] ETHIER S.N., KURTZ T.G., *Markov processes. Characterization and convergence* John Wiley & Sons, Inc. (1986).

[G89] GEWEKE J., *Bayesian inference in econometric models using Monte Carlo integration*, Econometrica, 57, 6 (1989) 1317-1339.

[GH03] GREEN P.J., HJORT N.L., *Highly structured stochastic systems*, Oxford University Press (2003).

[GO97] GUO Z., OGATA Y., *Statistical relations between the parameters of aftershocks in time, space, and magnitude*, Journal of Geophysical Research, 102, B2 (1997) 2857-2873.

[GS90] GRIMMET G.R., STIRZAKER D.R., *Probability and Random Processes*, Oxford Science Publications (1990).

[GSS93] GORDON N.J., SALMOND D.J., SMITH A.F.M., *Novel approach to nonlinear/non-Gaussian Bayesian state estimation*, IEEE Proceedings-F, 40, 2 (1993) 107-113.

[GT01] GERARDI A., TARDELLI P., *Finite state and discrete time approximation for filters* Nonlinear Analysis 47, 2485-2495 (2001).

[H03] HAJEK B., *Notes for ECE 467. Analysis of Computer Networks*, http://tesla.csl.uiuc.edu/ hajek/Papers/networkanalysis.html (2003).

[HA73] HAWKES A., ADAMOPOULOS L., *Cluster models for earthquakes - regional comparisons*, Bulletin of the International Statistical Institute, 45, 3 (1973) 454-461.

[HW95] HALL-WALLACE M.K. *Can earthquakes be predicted?*, Journal of Geoscience Education 46 (1998) 439-449.

[K71] KNOPOFF L., *A stochastic model for the occurrence of main sequence events*, Reviews of Geophysics and Space Physics, 9 (1971) 175-188.

[K91] KARR A.F. *Point processes and their statistical inference*, Marcel Dekker (1991).

[K93] KISSLINGER C., *The stretched exponential function as an alternative model for aftershock decay rate*, Journal of Geophysical Research, 98 (1993) 1913-1921.

[K98] KITAGAWA G., *A self-organizing state-space model*, Journal of the American Statistical Association, 93, 443 (1998) 1203-1215.

[KA84] KIREMIDJIAN A., ANAGNOS T., *Stochastic slip-predictable model for earthquake occurrences*, Bulletin of the Seismological Society of America, 74 (1984) 739-755.

[KKM90] KLIEMANN W., KOCH G., MARCHETTI F. *On the unnormalized solution of filtering problem with counting process observations*, IEEE Transactions on Information Theory, 36, 6 (1990) 1415-1425.

[KLW94] KONG A., LIU J.S., WONG W.H., *Sequential imputations and bayesian missing data problems*, Journal of the American Statistical Association, 89, 425 (1994) 278-288.

[L66] LOMNITZ C., *Magnitude stability in earthquake sequences*, Bulletin of Seismological Society of America, 56 (1966) 247-249.

[L96] LIU J.S., *Metropolized independent sampling with comparisons to rejection sampling and importance sampling*, Statistics and Computing, 6 (1996) 113-119.

[L98] LOMBARDI A.M., *Equazione del filtraggio di un processo di salto con osservazioni di conteggio* Tesi di laurea, Università degli Studi di Roma "La Sapienza" (1997-1998).

[LC95] LIU J.S., CHEN R., *Blind deconvolution via sequential imputations*, Journal of the American Statistical Association, 90, 430 (1995) 567-576.

[LC98] LIU J.S., CHEN R., *Sequential Monte Carlo methods for dynamic systems*, Journal of the American Statistical Association, 93, 443 (1998) 1022-1031.

[LCSVj99] LIU J.S., CHEN Y., SHI Y., VERE-JONES D., *Coupled stress release model for time-dependent seismicity*, Pure and Applied Geophysics, 155 (1999) 649-667.

[LHB99] LU C., HARTE D., BEBBINGTON M., *A linked stress release model for historical Japanese earthquakes: coupling among major seismic regions*, Earth Planets Space, 51 (1999) 907-916.

[LS76] LEWIS P.A., SHEDLER G.S., *Simulation of nonhomogeneous Poisson processes with log linear rate function*, Biometrika, 63, 3 (1976) 501-505.

[LS01.I] LIPTSER R.S., SHIRYAEV A.N., *Statistics of random processes - I General Theory* second edition, Springer (2001).

[LS01.II] LIPTSER R.S., SHIRYAEV A.N., *Statistics of random processes - II Application* second edition, Springer (2001).

[O82] OGATA Y., AKAIKE H., KATSURA K., *The application of linear intensity models to the investigation of casual relations between a point process and another stochastic process*, Annals of the Institute of Statistical Mathematics, 34 B (1982) 373-387.

[O83] OGATA Y., *Estimation of the parameters in the modified Omori formula for aftershock frequencies by the maximum likelihood procedure*, Journal of Physical Earth, 31 (1983) 115-124.

[O88] OGATA Y., *Statistical moldels for earthquake occurrences and residual analysis for point processes*, Journal of the American Statistical Association, 83, 401 (1988) 9-27.

[O95] OGATA Y., *The centenary of the Omori formula for a decay law of aftershock activity*, Journal of Physical Earth, 43 (1995) 1-33.

[O98] OGATA Y., *Space-time point-process models for earthquake occurrences*, Annals of the Institute of Statistical Mathematics, 50, 2 (1998) 379-402.

[O99] OGATA Y., *Seismicity analysis through Point-process modeling: A Review*, Pure and Applied Geophysics, 155 (1999) 471-507.

[OKT03] OGATA Y., KATSURA K., TANEMURA M., *Modelling of heterogeneous space-time earthquake occurrences and its residual analysis*, Journal of the Royal Statistical Society, series C, 52, 4 (2003) 499-510.

[OMK93] OGATA Y., MATSU'URA R.S., KATSURA K., *Fast likelihood computation of epidemic type aftershock-sequence model*, Geophysical Research Letters, 20, 19 (1993) 2143-2146.

[Ot87] OTSUKA M., *A simulation of earthquake occurrence. Part 8. On Omori's law to express aftershock seismicity*, Zisin, 2nd series, 40 (1987) 65-75 (in Japanese).

[R11] REID H., *The elastic rebound theory of earthquakes*, University California Publication. Bulletin Department of Geology, 6 (1911) 412-444.

[R76] ROHATGI V.K., *An introduction to probability theory and mathematical statistics*, Wiley, New York (1976).

[R01] ROBERT C.P., *The Bayesian Choice. A Decision Theoretic Motivation*, Springer-Verlag, New York (2001).

[RV03] ROTONDI R., VARINI E., *Bayesian analysis of a marked point process: application in seismic assessment*, Statistica Methods & and Applications 12 (2003) 79-92.

[RV03.1] ROTONDI R., VARINI E., *Bayesian analysis of linked stress release models*, Proceedings of the Meeting on: Modelli Complessi e Metodi Computazionali Intensivi per la Stima e la Previsione (Complex models and intensive computational methods for estimation and forecasting), 2003 September 4-6, Treviso (Italy), 356-361.

[RV04] ROTONDI R., VARINI E., *Valutazione della pericolosità sismica nella regione Sannio-Matese-Ofanto-Irpinia attraverso modelli di rilascio di sforzo* (*Seismic hazard assessment in the Sannio-Matese-Ofanto-Irpinia region through stress release models*), Proceedings of the XI National Meeting "L'ingegneria Sismica in Italia" (Seismic engineering in Italy), 2004 January 25-29, Genoa (Italy), pp.8.

[S00] SCHOENBERG F.P., *Point processes*, http://web.stat.ucla.edu/ frederic/papers/pointprocesses.pdf (2000).

[S02] STORVIK G., *PArticle filters for state space models with the presence of unknown static parameters*, IEEE Transactions on Signal Processing, 50 (2002) 281-289.

[SN80] SHIMAZAKI K., NAKATA T. *Time-predictable model for large earthquakes*, Geophysical Research Letters, 7 (1980) 279-282.

[SP95] SHEDLOCK K.M., PAKISER L.C. *Earthquakes*, USGS General Interest Publications (1995) http://pubs.usgs.gov/gip/earthq1/

[U61] UTSU T., *A statistical study on the occurrence of aftershocks*, Geophysical Magazine, 30 (1961) 521-605.

[UOM95] UTSU T., OGATA Y., MATSU'URA R.S., *The centenary of the Omori formula for a decay law of aftershock activity*, Journal of Physical Earth, 43 (1995) 1-33.

[US55] UTSU T., SEKI A., *Relation between the area of aftershock region and the energy of the mainshock*, Zisin, 2nd series, 7 (1955) 233-240 (in Japanese).

[Vj78] VERE-JONES D., *Earthquake prediction - a statistician's view*, Journal of Physical Earth, 26 (1978) 129-146.

[Vj95] VERE-JONES D., *Forecasting earthquakes and earthquake risk*, International Journal of Forecasting, 11 (1995) 503-538.

[VjD66] VERE-JONES D., DAVIES R.B., *A statistical survey of earthquakes in the main seismic region of New Zealand. Part 2. Time series analyses.*, New Zealand Journal of Geology and Geophysics, 9 (1966) 251-284.

[VjY88] VERE-JONES D., Yongiu D., *A point process analysis of historical earthquakes from North China*, Earthquake Research in China, 2, 2 (1988) 165-181.

[WVjZ91] WANG A., VERE-JONES D., ZHENG X., *Simulation and estimation procedures for stress release models*, in Beckmann M.J., Gopalan M.N., Subramanian R. *Stochastic Processes and their Applications. Proceedings of the Symposium held in honour of Professor S. K. Srinivasan at the Indian Institute of Technology, Bombay, India, December 27-30, 1990* (Lecture Notes in Economics and Mathematical Systems) Springer-Verlag, Berlin (1991).

[ZOVj02] ZHUANG J., OGATA Y., VERE-JONES D., *Stochastic declustering of state-space earthquake occurrences*, Journal of American Statistical Association, 97, 458 (2002) 369-380.

[ZVj91] ZHENG X., VERE-JONES D., *Application of stress release models to historical earthquakes from North China*, Pure and Applied Geophysics, 135, 4 (1991) 559-576.

[ZVj94] ZHENG X., VERE-JONES D., *Further applications of the stochastic stress release model to historical earthquake data*, Tectonophysics, 229 (1994) 101-121.