Electronic Journal of Statistics
Vol. 8 (2014) 2264–2295
ISSN: 1935-7524
DOI: 10.1214/14-EJS952

Bayesian estimation for a parametric Markov Renewal model applied to seismic data

Ilenia Epifani

Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy
e-mail: ilenia.epifani@polimi.it

Lucia Ladelli

Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy
e-mail: lucia.ladelli@polimi.it

and

Antonio Pievatolo

IMATI-CNR, Via Bassini 15, 20133 Milano, Italy
e-mail: antonio.pievatolo@cnr.it

Abstract: This paper presents a complete methodology for Bayesian inference on a semi-Markov process, from the elicitation of the prior distribution, to the computation of posterior summaries, including a guidance for its implementation. The inter-occurrence times (conditional on the transition between two given states) are assumed to be Weibull-distributed. We examine the elicitation of the joint prior density of the shape and scale parameters of the Weibull distributions, deriving a specific class of priors in a natural way, along with a method for the determination of hyperparameters based on "learning data" and moment existence conditions. This framework is applied to data of earthquakes of three types of severity (low, medium and high size) that occurred in the central Northern Apennines in Italy and collected by the CPTI04 (2004) catalogue. Assumptions on two types of energy accumulation and release mechanisms are evaluated.

AMS 2000 subject classifications: Primary 60K20, 62F15, 62M05, 86A15; secondary 65C05.
Keywords and phrases: Bayesian inference, earthquakes, Gibbs sampling, Markov Renewal process, predictive distribution, semi-Markov process, Weibull distribution.

Received February 2013.

1. Introduction

Markov Renewal processes or their semi-Markov representation have been considered in the seismological literature as models which allow the distribution of the inter-occurrence times between earthquakes to depend on the last and next earthquake and is not necessarily exponential. The time predictable and the slip predictable models studied in Shimazaki and Nakata (1980), Grandori Guagenti and Molina (1986), Grandori Guagenti et al. (1988) and Betrò et al. (1989) are special cases of Markov Renewal processes. These models are capable of interpreting the predictable behavior of strong earthquakes in some seismogenic areas. In these processes the magnitude is a deterministic function of the inter-occurrence time. A stationary Markov Renewal process with Weibull inter-occurrence times has been studied from a classical statistical point of view in Alvarez (2005). The Weibull model allows for the consideration of monotonic hazard rates; it contains the exponential model as a special case which gives a Markov Poisson point process. In Alvarez (2005) the model parameters were fitted to the large earthquakes in the North Anatolian Fault Zone through maximum likelihood and the Markov Poisson point process assumption was tested. In order to capture a non monotonic behavior in the hazard, in Garavaglia and Pavani (2012) the model of Alvarez was modified and a Markov Renewal process with inter-occurrence times that are mixtures of an exponential and a Weibull distribution was fitted to the same Turkish data. In Masala (2012) a parametric semi-Markov model with a generalized Weibull distribution for the inter-occurrence times was adapted to Italian earthquakes. The semi-Markov model with generalized Weibull distributed times was first used in Foucher et al. (2009) to study the evolution of HIV infected patients. Votsi et al. (2012) considered a semi-Markov model for the seismic hazard assessment in the Northern Aegean sea and estimated the quantities of interest (semi-Markov kernel, Markov Renewal functions, etc.) through a nonparametric method.

While a wide literature concerning classical inference for Markov Renewal models for earthquake forecasting exists, to our knowledge comparatively little Bayesian work has been carried out. Patwardhan et al. (1980) considered a semi-Markov model with log-normal distributed discrete inter-occurrence times and applied it to the large earthquakes in the circum-Pacific belt. They stressed that it is relevant to use Bayesian techniques when prior knowledge is available and it is fruitful even if the sample size is small. Marín et al. (2005) also employed semi-Markov models in the Bayesian framework, applied to a completely different area: sow farm management. They used WinBugs to perform computations (but without giving details) and they elicited their prior distributions on parameters from knowledge on farming practices.

From a probabilistic viewpoint, a Bayesian statistical treatment of a semi-Markov process amounts to modeling the data as a mixture of semi-Markov processes, where the mixing measure is supported on the parameters, by means of their prior laws. A complete characterization of such a mixture is given in Epifani et al. (2002).

In this paper we develop a parametric Bayesian analysis for a Markov Renewal process modelling earthquakes in an Italian seismic region. The magnitudes are classified into three categories according to their severity: low, medium and high size, and these categories represent the states visited by the process. As in Alvarez (2005), the inter-occurrence times are assumed to be Weibull random variables. The "current sample" is formed by the sequences of earthquakes in a homogeneous seismic region and by the corresponding inter-occurrence times collected up to a time T. When T does not coincide with an earthquake, the last observed inter-occurrence time is censored. The prior distribution of the parameters of the model is elicited using a "learning dataset", i.e. data coming from a seismic region similar to that under analysis. The posterior distribution of the parameters is obtained through Gibbs sampling and the following summaries are estimated: transition probabilities, shape and scale parameters of the Weibull inter-occurrence times for each transition and the so-called cross state-probabilities (CSPs). The transition probabilities indicate whether the strength of the next earthquake is in some way dependent on the strength of the last one; the shape parameters of the inter-occurrence times indicate whether the hazard rate between two earthquakes of given magnitude classes is decreasing or increasing; the CSPs give the probability that the next earthquake occurs at or before a given time and is of a given magnitude, conditionally on the time elapsed since the last earthquake and on its magnitude.

The paper is organized as follows. In Section 2 we illustrate the dataset and we discuss the choice of the Weibull model in detail. Section 3 introduces the parametric Markov Renewal model. Section 4 deals with the elicitation of the prior. Section 5 contains the Bayesian data analysis with the estimation of the above-mentioned summaries. We also test a time predictable and a slip predictable model against the data. Section 6 concludes. Appendix A contains the detailed derivation of the full conditional distributions and the JAGS (Just Another Gibbs Sampler) implementation of the Gibbs sampler (Plummer, 2010).

2. A test dataset

We tested our method on a sequence of seismic events chosen among those examined in Rotondi (2010). The sequence collects events that occurred in a tectonically homogeneous macroregion, identified as MR_3 by Rotondi and corresponding to the central Northern Apennines in Italy. The subdivision of Italy into eight (tectonically homogeneous) seismic macroregions can be found in the DISS (2007) and the data are collected in the CPTI04 (2004) catalogue. The catalogue provides the moment magnitude of every seismic event. The moment magnitude, denoted by M_w, measures the size of earthquakes with respect to the energy released during the event. It is related to the seismic moment M_0 through the relationship: $M_w = \frac{2}{3}(\log_{10} M_0 - 16.05)$; the seismic moment measures the rigidity of the Earth multiplied by the average amount of slip on the fault and the size of the area that slipped. See Hanks and Kanamori (1979).

If one considers the earthquakes from CPTI04 (2004) with magnitude $M_w \geq 4.5$, the sequence is complete from year 1838: a lower magnitude would make the completeness of the series questionable, especially in its earlier part. The map of these earthquakes marked by dots appears in Figure 1. As a lower threshold for the class of strong earthquakes, we choose $M_w \geq 5.3$, as suggested by Rotondi (2010). Then a magnitude state space with three states

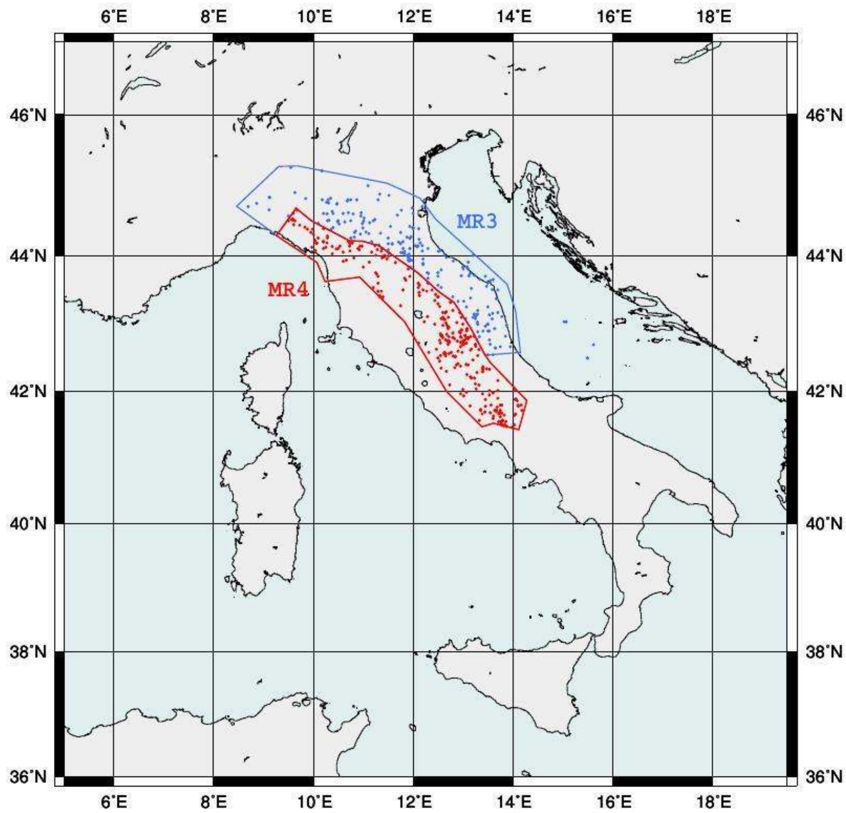

FIG 1. *Map of Italy with dots indicating earthquakes with magnitude* $M_w \geq 4.5$ *from* 1838 *to* 2002*, belonging to macroregions* MR_3 *(blue) and* MR_4 *(red) (Rotondi (2010)). Inclusion in the macroregions was based on the association between events and seismogenic sources; the contours of the regions have only an aesthetic function.*

is obtained by indexing an earthquake by 1, 2 or 3 if its magnitude belongs to intervals $[4.5, 4.9)$, $[4.9, 5.3)$, $[5.3, +\infty)$, respectively. Magnitude 4.9 is just the midpoint between 4.5 and 5.3 and the released energy increases geometrically as one moves through the endpoints, with a common ratio of 4: indeed $M_0(5.3)/M_0(4.9) = M_0(4.9)/M_0(4.5) = 10^{\frac{3}{2}0.4} \simeq 4$.

The energy released from an earthquake with $M_w = 4.9$ does not match the midpoint between seismic moments associated with magnitudes 4.5 and 5.3 (in fact, this correspondence holds if $M_w = 5.1$). However, there seem to be no general rule in the literature for splitting magnitude intervals. For example, Votsi et al. (2012) used cut-points 5.5, 5.7 and 6.1, so that $M_0(6.1)/M_0(5.7) \simeq 4$ and $M_0(5.7)/M_0(5.5) \simeq 2$, while the energy midpoint is at $M_w = 5.9$; following Altinok and Kolcak (1999), Alvarez (2005) uses cut-points 5.5, 6.0 and 6.5; Masala (2012) employed the magnitude classes $M_w < 4.7$, $M_w \in [4.7, 5)$, $M_w \geq 5$. All these authors do not give any special reason for their choices.

A more structured approach is attempted by Sadeghian (2012), who applied a statistical clustering algorithm to magnitudes, and again by Votsi et al. (2012) when they propose a different classification of states that combines both magnitude and fault orientation information. From a modelling viewpoint, this latter approach is certainly preferable, because it is likely to produce more homogeneous classes. However we do not have enough additional information to attempt this type of classification on our data in a meaningful way. An entirely different approach is that based on risk, in which cut-points would change with the built environment.

We now examine inter-occurrence times. Rotondi (2010) considers a nonparametric Bayesian model for the inter-occurrence times between strong earthquakes (i.e. $M_w \geq 5.3$), after a preliminary data analysis which rules out Weibull, Gamma, log-normal distributions among others frequently used. On the other hand, with a Markov Renewal model, the sequence of all the inter-occurrence times is subdivided into shorter ones according to the magnitudes, so that we think that a parametric distribution is a viable option. In particular, we focussed on the macroregion MR_3 because the Weibull distribution seems to fit the inter-occurrence times better than in other macroregions. This fact is based on qq-plots. The qq-plots for MR_3 are shown in Figure 2. The plot for transitions from 1 to 3 shows a sample quantile that is considerably larger than expected. The outlying point corresponds to a long inter-occurrence time of about 9 years, between 1987 and 1996, while 99 percent of the inter-occurence times are below 5 years. Obviously, the classification into macroregions influences the way the earthquake sequence is subdivided.

Given the Markov Renewal model framework, inter-occurrence time distributions other than the Weibull could be used, such as the inverse Gaussian, the log-normal or the Gamma. However, the inverse Gaussian qq-plots clearly indicate that this distribution does not fit the data. As for the log-normal, the outlying point in the qq-plot of the $(1, 3)$ transition becomes only a little less isolated, but at the expense of introducing an evident curvature in the qq-plot of the $(1, 1)$ transition, whereas the remaining qq-plots are unchanged. The Gamma qq-plots are indistinguishable from the Weibull qq-plots, but we prefer working with the Weibull in view of the existing literature on seismic data analysis where the Weibull is employed. In this respect, we could follow Masala (2012) and choose the generalized Weibull, which includes the Weibull, but the qq-plots are unchanged even with the extra parameter. From a Bayesian computational point of view, there is no special reason for preferring the (possibly generalized) Weibull to the Gamma, as neither of them possesses a conjugate prior distribution and numerical methods are needed in both cases for making inference.

In the existing literature, the Weibull distribution has been widely used to model inter-occurrence times between earthquakes from different areas and with different motivations, beyond those mentioned in Section 1. Abaimov et al. (2007) argued that the increase in stress caused by the motion of tectonic plates at plate boundary faults is adequately described by an increasing hazard function, such as the Weibull can have. Instead, other distributions have an inap-

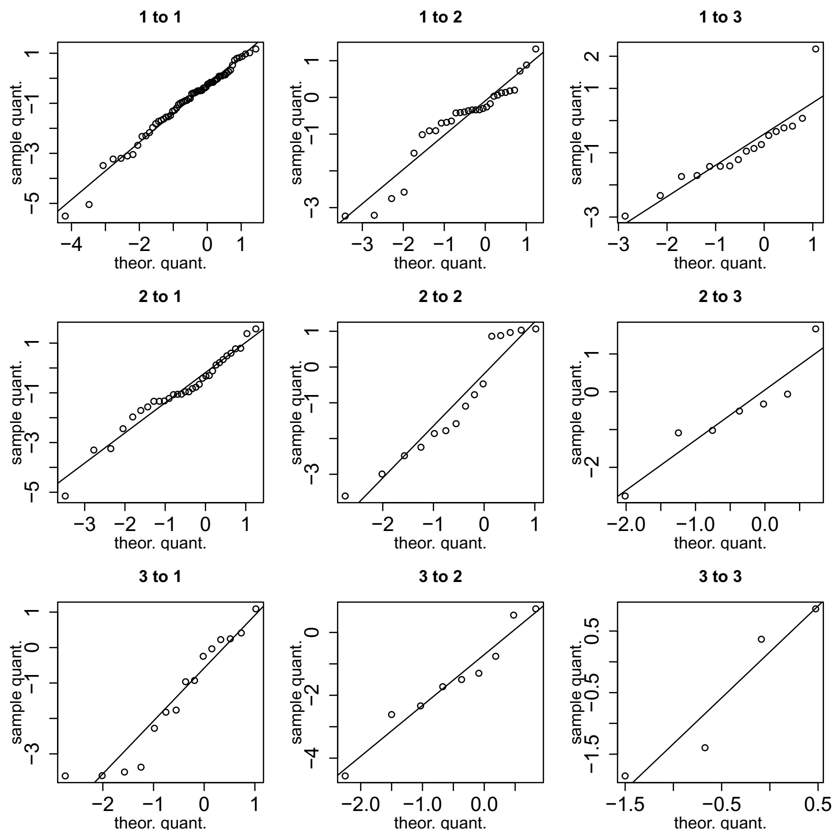

FIG 2. *Weibull qq-plots based on least squares fits of earthquake inter-occurrence times (central Northern Apennines) classified by transition between magnitude classes.*

propriate tail behaviour: the log-normal hazard tends to zero with time and the inverse Gaussian hazard tends to a constant. Goodness-of-fit checks for the recurrence times of slip events in the creeping section of the San Andreas fault in central California confirmed that the Weibull is preferable to the mentioned alternatives. Hristopulos and Mouslopoulou (2013) considered a Weibull model, for single faults (or fault systems with homogeneous strength statistics) and power law stress accumulation. They derived the Weibull model from a theoretical framework based on statistical mechanics of brittle fracture and they applied it to microearthquake sequences (small magnitudes) from the island of Crete and from a seismic area of Southern California, finding agreement with the data except for some deviations in the upper tail. Regarding tail behaviour, we can make a connection with Hasumi et al. (2009), who analyzed a catalogue of the Japan Meteorological Agency. These data support the hypothesis that the inter-occurrence times can be described by a mixture of a Weibull distribution and a log-Weibull distribution (which possesses a heavier tail); if only earthquakes with a magnitude exceeding a threshold are considered, the weight of the log-Weibull component becomes negligible as the threshold increases.

3. Markov Renewal model

Let us observe, over a period of time $[0, T]$, a process in which different events occur, with random inter-occurence times. Let us suppose that the possible states of the process are in a finite set $E = \{1, \dots, s\}$ and that the process starts from state j_0. Let us denote by τ the number of times the process changes states in the time interval $[0, T]$ and by t_i the time of the i-th change of state. Hence, $0 < t_1 < \dots < t_\tau \leq T$. Let $j_0, j_1, \dots, j_\tau$ be the sequence of states visited by the process and x_i the inter-occurrence time between states j_{i-1} and j_i. Then

$$x_i = t_i - t_{i-1} \qquad \text{for } i = 1, \dots, \tau$$

with $t_0 := 0$. Furthermore, let u_T be the time spent in j_τ

$$u_T = T - t_\tau,$$

so the time u_T is a right-censored time. Finally, our data are collected in the vector $(\mathbf{j}, \mathbf{x}, u_T)$, where $(\mathbf{j}, \mathbf{x}) = (j_n, x_n)_{n=1,\dots,\tau}$.

In what follows, we assume that the data $(\mathbf{j}, \mathbf{x}, u_T)$ are the result of the observation of a homogeneous Markov Renewal process $(J_n, X_n)_{n\geq 0}$ starting from j_0. This means that the sequence $(J_n, X_n)_{n\geq 0}$ satisfies

$$P(J_0 = j_0) = 1, \quad P(X_0 = 0) = 1 \tag{1}$$

and for every $n \geq 0$, $j \in E$ and $t \geq 0$

$$\begin{aligned} &P(J_{n+1} = j, X_{n+1} \leq t | (J_k, X_k)_{k\leq n}) \\ &\qquad = P(J_{n+1} = j, X_{n+1} \leq t | (J_n, X_n)) = p_{J_n j} F_{J_n j}(t). \end{aligned} \tag{2}$$

The transitions probabilities, p_{ij}, are collected in a transition matrix $\mathbf{p} = (p_{ij})_{i,j\in E}$ and $(F_{ij})_{i,j\in E}$ is an array of distribution functions on $\mathbb{R}_+ = (0, +\infty)$. For more details on Markov Renewal processes see, for example, Limnios and Oprisan (2001). We just recall that, under Assumptions (1) and (2):

- the process $(J_n)_{n\geq 0}$ is a Markov chain, starting from j_0, with transition matrix $\mathbf{p}$,
- the inter-occurrence times $(X_n)_{n\geq 0}$, conditionally on $(J_n)_{n\geq 0}$, form a sequence of independent positive random variables, with distribution function $F_{J_{n-1}\ J_n}$.

We assume that the functions F_{ij} are absolutely continuous with respect to the Lebesgue measure with density f_{ij}. Hence, the likelihood function of the data $(\mathbf{j}, \mathbf{x}, u_T)$ is

$$L(\mathbf{j}, \mathbf{x}, u_T) = \left(\prod_{i=0}^{\tau-1} p_{j_i j_{i+1}} f_{j_i j_{i+1}}(x_{i+1}) \right)^{\mathbb{1}(\tau>0)} \times \sum_{k\in E} p_{j_\tau k} \bar{F}_{j_\tau k}(u_T), \tag{3}$$

where, for every x, $\bar{F}_{ij}$ is the survival function

$$\bar{F}_{ij}(x) = 1 - F_{ij}(x) = P(X_{n+1} > x | J_n = i, J_{n+1} = j).$$

Furthermore, we assume that each inter-occurrence time has a Weibull density f_{ij} with shape parameter α_{ij} and scale parameter θ_{ij}, i.e.

$$f_{ij}(x) = \frac{\alpha_{ij}}{\theta_{ij}} \left(\frac{x}{\theta_{ij}}\right)^{\alpha_{ij}-1} \exp\left\{-\left(\frac{x}{\theta_{ij}}\right)^{\alpha_{ij}}\right\}, \quad x > 0,\ \alpha_{ij} > 0,\ \theta_{ij} > 0. \quad (4)$$

For conciseness, let $\boldsymbol{\alpha} = (\alpha_{ij})_{i,j\in E}$ and $\boldsymbol{\theta} = (\theta_{ij})_{i,j\in E}$.

In order to write the likelihood in a more convenient way, let us introduce the following natural statistics. We will say that the process visits the string (i, j) if a visit to i is followed by a visit to j and we denote by

- x_{ij}^{ρ} the time spent in state i at the ρ-th visit to the string (i, j),
- N_{ij} the number of visits to the string (i, j).

Then, assuming $\tau \geq 1$, Equations (3) and (4) yield the following representation of the likelihood function

$$\begin{aligned} L(\mathbf{j}, \mathbf{x}, u_T \,|\, \mathbf{p}, \boldsymbol{\alpha}, \boldsymbol{\theta}) &= \prod_{i,k\in E} p_{ik}^{N_{ik}} \\ &\times \prod_{i,k\in E} \left[\alpha_{ik}^{N_{ik}} \frac{1}{\theta_{ik}^{\alpha_{ik} N_{ik}}} \left(\prod_{\rho=1}^{N_{ik}} x_{ik}^{\rho}\right)^{\alpha_{ik}-1} \times \exp\left\{-\frac{1}{\theta_{ik}^{\alpha_{ik}}} \sum_{\rho=1}^{N_{ik}} (x_{ik}^{\rho})^{\alpha_{ik}}\right\}\right] \\ &\times \left(\sum_{k\in E} p_{j_\tau k} \exp\left\{-\left(\frac{u_T}{\theta_{j_\tau k}}\right)^{\alpha_{j_\tau k}}\right\}\right). \end{aligned} \quad (5)$$

Our purpose is to perform a Bayesian analysis for $\mathbf{p}, \boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ which allows us to introduce prior knowledge on the parameters. As shown in Appendix A, this analysis is possible via a Gibbs sampling approach.

4. Bayesian analysis

4.1. The prior distribution

Let us assume that a priori $\mathbf{p}$ is independent of $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$. In particular, the rows of $\mathbf{p}$ are s independent vectors with Dirichlet distribution with parameters $\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_s$. This means that, for $i = 1, \dots, s$, the prior density of the i-th row is

$$\pi_{1,i}(p_{i1}, \dots, p_{is}) = \frac{\Gamma(c_i)}{\prod_{j=1}^{s} \Gamma(\gamma_{ij})} \prod_{j=1}^{s} p_{ij}^{\gamma_{ij}-1} \quad (6)$$

on $T = \{(p_{i1}, \dots, p_{is}) |\ p_{ij} \geq 0,\ \sum_j p_{ij} = 1\}$ where $\boldsymbol{\gamma}_i = (\gamma_{i1}, \dots, \gamma_{is})$, with $\gamma_{ij} > 0$ and $c_i = \sum_{j=1}^{s} \gamma_{ij}$. Notice that the first subscript in $\pi_{1,i}$ indexes the density of $\mathbf{p}$. Similarly, the conditional density of $\boldsymbol{\theta}$, given $\boldsymbol{\alpha}$, and the density of $\boldsymbol{\alpha}$ will be indexed by 2 and 3, respectively. As far as $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ are concerned, the θ_{ij}'s, given the α_{ij}'s, are independent with generalized inverse Gamma densities

$$\pi_{2,ij}(\theta_{ij}|\boldsymbol{\alpha}) = \pi_{2,ij}(\theta_{ij}|\alpha_{ij}) = \frac{\alpha_{ij} b_{ij}(\alpha_{ij})^{m_{ij}}}{\Gamma(m_{ij})} \theta_{ij}^{-(1+m_{ij}\alpha_{ij})} \times \exp\left\{-\frac{b_{ij}(\alpha_{ij})}{\theta_{ij}^{\alpha_{ij}}}\right\}, \quad \theta_{ij} > 0, \quad (7)$$

where $m_{ij} > 0$ and

$$b_{ij}(\alpha_{ij}) = \left(t_{q_{ij}}^{ij}\right)^{\alpha_{ij}} [(1 - q_{ij})^{-1/m_{ij}} - 1]^{-1} \quad (8)$$

with $t_{q_{ij}}^{ij} > 0$ and $q_{ij} \in (0, 1)$. In other terms, $\theta_{ij}^{-\alpha_{ij}}$, given α_{ij}, has a prior Gamma density with shape m_{ij} and scale $1/b_{ij}(\alpha_{ij})$. In symbols, $\theta_{ij}|\alpha_{ij} \sim \mathcal{GIG}(m_{ij}, b_{ij}(\alpha_{ij}), \alpha_{ij})$. We borrow the expression of the $b_{ij}(\alpha_{ij})$'s in (8) from Bousquet (2010) and, as a consequence of this choice, $t_{q_{ij}}^{ij}$ turns out to be the marginal quantile of order q_{ij} of an inter-occurrence time between states i and j. Indeed, if $\pi_{3,ij}$ denotes the density of α_{ij} and X is such a random time, then

$$\begin{aligned} P(X > t) &= \int_0^{+\infty} \int_0^{+\infty} P(X > t|\alpha_{ij}, \theta_{ij}) \pi_{2,ij}(\theta_{ij}|\alpha_{ij}) \pi_{3,ij}(\alpha_{ij}) d\theta_{ij} d\alpha_{ij} \\ &= \int_0^{+\infty} \left[\frac{b_{ij}(\alpha_{ij})}{b_{ij}(\alpha_{ij}) + t^{\alpha_{ij}}}\right]^{m_{ij}} \pi_{3,ij}(\alpha_{ij}) d\alpha_{ij}, \qquad \forall t > 0. \end{aligned}$$

Hence, in view of (8), if $t = t_{q_{ij}}^{ij}$, we obtain $P(X > t_{q_{ij}}^{ij}) = 1 - q_{ij}$, for every proper prior density $\pi_{3,ij}$.

Finally, a priori, the components of $\boldsymbol{\alpha}$ are independent and have densities $\pi_{3,ij}$ such that

$$\pi_{3,ij}(\alpha_{ij}) \propto \alpha_{ij}^{m_{ij} - c_{ij}} (\alpha_{ij} - \alpha_{0,ij})^{c_{ij}-1} \exp\{-m_{ij} d_{ij} \alpha_{ij}\} \mathbb{1}(\alpha_{ij} \geq \alpha_{0,ij}), \quad \alpha_{0,ij} \geq 0,\ c_{ij} > 0,\ m_{ij} > 0,\ d_{ij} \geq 0. \quad (9)$$

As far as the prior $\pi_{3,ij}$ is concerned, it is easy to see that:

a) if $d_{ij} > 0$, then $\pi_{3,ij}$ is a proper prior;
b) if $\alpha_{0,ij} = 0$ and $d_{ij} > 0$, then $\pi_{3,ij}$ is a Gamma density;
c) if $c_{ij} = 1$, $\alpha_{0,ij} > 0$ and $d_{ij} > 0$, then $\pi_{3,ij}$ is a Gamma density truncated from below at $\alpha_{0,ij}$;
d) if $c_{ij} = m_{ij}$, $\alpha_{0,ij} > 0$ and $d_{ij} > 0$, then $\pi_{3,ij}$ is a Gamma density shifted by $\alpha_{0,ij}$;
e) if $c_{ij} = 1$ and $m_{ij} \to 0$, then $\pi_{2,ij}(\theta_{ij}|\alpha_{ij})\pi_{3,ij}(\alpha_{ij})$ approaches the Jeffreys prior for the Weibull model: $1/\theta_{ij} \mathbb{1}(\theta_{ij} > 0) \mathbb{1}(\alpha_{ij} \geq \alpha_{0,ij})$;
f) if $c_{ij} \geq 1$ and $m_{ij} \geq 1$, then $\pi_{3,ij}$ is a log-concave function.

The prior corresponding to the choices in c) was first introduced in Bousquet (2006) and Bousquet (2010). Beside, a complete discussion of the elicitation of Weibull priors can be found in Bousquet (2010) and the references therein. The author also devotes particular attention to the matter of incorporating expert opinions in the elicitation procedure. Readers interested in elicitation based on experts opinions are referred also to O'Hagan et al. (2006).

As discussed in Gilks and Wild (1992), the log-concavity of $\pi_{3,ij}$ is necessary in the implementation of the Gibbs sampler (see also Berger and Sun (1993)), although adjustments exist for the non-log-concave case (see Gilks et al. (1995)). Furthermore, we will show later that a support suitably bounded away from zero ensures the existence of the posterior moments of the θ_{ij}'s.

4.2. Elicitation of the hyperparameters

In this section we focus our attention on the prior of $(\alpha_{ij}, \theta_{ij})$, for fixed i, j. Adapting the approach developed by Bousquet to our situation, we give a statistical justification of the prior introduced in Subsection 4.1. An interpretation of the hyperparameters is also provided. For the sake of simplicity, let us drop the indices i, j in all the notation and quantities.

Suppose that a "learning dataset" $\mathbf{y}_m = (y_1, \dots, y_m)$ of m inter-occurrence times between visits to state i followed by a visit to state j is available from another seismic region similar to the one under analysis. Therefore the prior scheme defined by Equations (7)–(9) can be interpreted as a suitable modification of a posterior distribution of (α, θ), given the learning dataset $\mathbf{y}_m$. This approach allows us to elicit the hyperparameters in a data-based fashion.

More precisely, consider for (α, θ) the posterior density, conditionally on $\mathbf{y}_m$, when we start from the following improper prior:

$$\tilde{\pi}(\alpha, \theta) \propto \theta^{-1} \left(1 - \frac{\alpha_0}{\alpha}\right)^{c-1} \mathbb{1}(\theta \geq 0) \mathbb{1}(\alpha \geq \alpha_0), \tag{10}$$

for some suitable $c \geq 1$ and $\alpha_0 \geq 0$ (The condition $c \geq 1$ guarantees that $\tilde{\pi}(\alpha, \theta)$ is a log-concave function with respect to α). Consequently, the posterior density of θ, given α and $\mathbf{y}_m$, is

$$\tilde{\pi}_2(\theta | \mathbf{y}_m, \alpha) = \mathcal{GIG}(m, \tilde{b}(\mathbf{y}_m, \alpha), \alpha) \tag{11}$$

and the posterior density of α is

$$\tilde{\pi}_3(\alpha | \mathbf{y}_m) \propto \frac{\alpha^{m-c}(\alpha - \alpha_0)^{c-1}}{\tilde{b}^m(\mathbf{y}_m, \alpha)} \exp\{-m\beta(\mathbf{y}_m)\alpha\} \mathbb{1}(\alpha \geq \alpha_0), \tag{12}$$

with $\tilde{b}(\mathbf{y}_m, \alpha) = \sum_{i=1}^m y_i^\alpha$ and $\beta(\mathbf{y}_m) = \sum_{i=1}^m \ln y_i / m$.

Notice that the posterior we obtain has a simple hierarchical structure: $\tilde{\pi}_2(\theta | \mathbf{y}_m, \alpha)$ is a generalized inverse Gamma density and this provides both a justification of the form of the $\pi_{2,ij}(\theta | \alpha)$ in (7) and an interpretation of the first parameter m. Indeed m is equal to the size of the learning dataset $\mathbf{y}_m$ and so it is a measure of prior uncertainty.

Now, if we replace the function $\tilde{b}(\mathbf{y}_m, \alpha)$ in (11) and (12) by the simpler convex function of α introduced in (8), i.e. $b(\alpha) = t_q^\alpha[(1-q)^{-1/m} - 1]^{-1}$, with $t_q > 0$ and $q \in (0,1)$, then $\tilde{\pi}_3(\alpha|\mathbf{y}_m)$ takes the same form as in (9) with

$$d = \ln t_q - \frac{\sum_{i=1}^m \ln y_i}{m}. \tag{13}$$

In this way, we obtain a justification of the form of the prior density $\pi_{3,ij}$ in (9) and an easy way to elicit its parameter d_{ij} when the learning dataset is available. Furthermore, $b(\alpha)$ can be also elicited once the predictive quantile t_q is specified. Its specification can be accomplished, for example, in the two following different ways:

1. an empirical quantile $\hat{t}_q$ is estimated from the learning dataset;
2. an expert is asked about the chance, quantified by q, of an earthquake before an assigned t_q (but we will not follow this approach).

In the following, if a learning dataset of size $m \geq 2$ is available, we consider an empirical quantile $\hat{t}_q$ of order q such that

$$\ln \hat{t}_q - \frac{\sum_{i=1}^m \ln y_i}{m} > 0.$$

Therefore, letting $\hat{b}(\alpha)$ denote the value of $b(\alpha)$ corresponding to $\hat{t}_q$, we propose a Bayesian analysis based on the prior

$$\pi_2(\theta|\alpha) = \mathcal{GIG}(m, \hat{b}(\alpha), \alpha) \tag{14}$$

and

$$\pi_3(\alpha) \propto \alpha^{m-c}(\alpha - \alpha_0)^{c-1} \exp\left\{-m\left(\ln \hat{t}_q - \frac{\sum_{i=1}^m \ln y_i}{m}\right)\alpha\right\} \mathbb{1}(\alpha \geq \alpha_0), \tag{15}$$

where m is the size of the learning dataset. In addition, we choose $c = m$ so that $\pi_3(\alpha)$ is a shifted Gamma prior and consequently it is proper and log-concave.

The remaining hyperparameter α_0 is chosen so that the posterior second moment of θ is finite. If α is bounded away from zero, then

$$\mathbb{E}(\theta^2) = \mathbb{E}\left(\frac{\Gamma(m - 2/\alpha)}{\Gamma(m)}\left[\hat{b}(\alpha)\right]^{2/\alpha}\right) \leq \tilde{K}\mathbb{E}(\Gamma(m - 2/\alpha))$$

for a suitable constant $\tilde{K}$. As a consequence, if $\alpha_0 = 2/m$, then $\mathbb{E}(\theta^2) < +\infty$ and hence the posterior second moment of θ is also finite.

The choice $\alpha_0 = 2/m$ is suitable only if $m > 2$. If $m = 2$, then $\alpha_0 = 2/m = 1$ and decreasing hazard rates are ruled out. In the absence of additional specific prior information, this is an arbitrary restriction, so a value for α_0 smaller than 1 must be chosen. Then, the prior second moment of θ is not finite anymore.

On the other hand, for the posterior second moment to be finite, we need $\alpha > 2/(2+N)$, where N is the number of transitions between the two concerned states in the (current) sample. Thus the second moment of θ can stay non-finite, even a posteriori, if $2/(2+N) > \alpha_0$. This would show that the data add little information for that specific transition. To avoid this, we may let α_0 be the minimum between the value $2/3$, corresponding to the smallest learning sample size such that $\alpha_0 < 1$, and the value $2/(2+N)$, necessary for the finiteness of the posterior second moment. Therefore, $\alpha_0 = \min\{2/3, 2/(2+N)\}$.

Finally, if γ denotes the hyperparameter corresponding to the indexes i and j in the Dirichlet prior (6), then we select $\gamma = m+1$, i.e. γ is equal to the number of transitions from state i to state j in the learning dataset, plus one.

4.3. Scarce prior information

The construction of the prior distribution of (α, θ) must be modified for those pairs of states between which no more than one transition was observed in the learning dataset.

If $m = 1$, the single learning observation y_1 determines $\hat{b}(\alpha)$. As $\hat{t}_q = y_1$ for any q, it seems reasonable to use $q = 0.5$, so y_1 would represent the prior opinion on the median inter-occurrence time. Since $d = 0$ when $m = 1$, $\pi_3(\alpha)$ is improper for any $c > 0$. We make it proper by restricting its support to an interval (α_0, α_1). The value $\alpha_1 = 10$ is suitable for all practical purposes. As before, the choice $\alpha_0 = 2/m = 2$ would be too restrictive, so we again select $\alpha_0 = \min\{2/3, 2/(2+N)\}$. With regard to c, we put $c = 2$. Furthermore, the elicitation of the hyperparameter of the Dirichlet prior is again $\gamma = m + 1 = 2$, i.e. the number of transitions observed in the learning dataset (just one) plus one.

If $m = 0$, the prior information on the number of transitions is that there have been no transitions, but there is no information in the model on the inter-occurrence times. In this case we can represent in the model the absence of information, by choosing

$$q = 0.5 \quad \text{and} \quad \tilde{t}_{0.5} \sim U(t_1, t_2),$$

that is $\tilde{t}_{0.5}$ is uniformly distributed over a big time interval (t_1, t_2), independently from everything else. Hence, we use $q = 0.5$ and $\tilde{t}_{0.5}$ to obtain $\hat{b}(\alpha)$ and we obtain the previous case by substituting $m = 0$ with $m = 1$.

For clarity, in Table 1, we summarize the hyperparameter selection for priors (6), (14) and (15).

5. Analysis of the central Northern Apennines sequence

In this section we analyze the macroregion MR_3 sequence, using the proposed semi-Markov model.

We coded the Gibbs sampling algorithm in JAGS (Plummer, 2010), which is designed to work closely with the R (2012) package, in which all statistical

Table 1
Hyperparameter selection as the learning sample size m varies

	$m > 2$	$m = 2$	$m = 1$	$m = 0$
t_q	$\hat{t}_q$	$\hat{t}_q$	y_1	$\tilde{t}_q \sim U(t_1, t_2)$
c	m	m	2	2
α_0	$\frac{2}{m}$	$\min\left\{\frac{2}{3}, \frac{2}{2+N}\right\}$	$\min\left\{\frac{2}{3}, \frac{2}{2+N}\right\}$	$\min\left\{\frac{2}{3}, \frac{2}{2+N}\right\}$
γ	$m+1$	$m+1$	$m+1$	2

computations and graphics were performed. Details of the Gibbs sampler are in Appendix A. On the whole, 750,000 iterations for one chain were run for estimating the unknown parameters in the model, and the first 250,000 were discarded as burn-in. After the burn-in, one out of every 100 simulated values was kept for posterior analysis, for a total sample size of 5,000. Convergence diagnostics, such as those available in the R package CODA (Geweke, Heidelberger and Welch stationarity test, interval halfwidth test), were computed for all parameters, indicating that convergence had been achieved.

Model fitting, model validation and forecasting will involve the following steps:

1. the learning dataset for the elicitation of the prior distribution is chosen;
2. model fit is assessed by comparing observed inter-occurrence times (grouped by transition) to posterior predictive intervals;
3. cross state-probabilities are estimated, as an indication of the most likely magnitude and time to the next event, given information up to the present time;

Finally, in Subsection 5.1 we evaluate two special kinds of semi-Markov models against our data: the time predictable model and the slip predictable model. These models have already been considered in the seismological literature, see for example Grandori Guagenti and Molina (1986) and Grandori Guagenti et al. (1988).

1. For the elicitation of the prior distribution, the learning data are taken from MR_4 (represented in Figure 1), another macroregion among those considered by Rotondi (2010), who examines statistical summaries of the inter-occurrence times and suggests that MR_4 could be used as a learning set for the hyperparameters of MR_3. Peruggia and Santner (1996), in their analysis of the magnitudes and of the inter-occurence times of eartquakes from another Italian area, chose a subset of the incomplete older part of their series to elicit prior distributions. This procedure is justified in their case because the old and the new part of the series can be regarded as two different processes and the cut-point between them appears to be clearly identified. If we did the same with our series we would obtain different posterior distributions on changing the cut-point position. However, the use of subsetting is amply documented in the Bayesian literature in ways that can help overcome this ambiguity. See, for example, Berger and Pericchi (1996), O'Hagan (1995) and Yu et al. (2011).

TABLE 2

Summaries of datasets MR_3, tables (a) and (c), and MR_4, tables (b) and (d); MR_4 is the learning dataset used for hyperparameter elicitation. (a) and (b): number of observed transitions [and relative frequencies]; (c) and (d): median inter-occurrence times (in days)

(a)

	to 1	to 2	to 3
from 1	65 [0.580]	30 [0.268]	17 [0.152]
from 2	32 [0.585]	15 [0.283]	7 [0.132]
from 3	15 [0.536]	9 [0.321]	4 [0.143]

(b)

	to 1	to 2	to 3
from 1	114 [0.640]	51 [0.287]	13 [0.073]
from 2	56 [0.659]	25 [0.294]	4 [0.047]
from 3	8 [0.421]	8 [0.421]	3 [0.158]

(c)

	to 1	to 2	to 3
from 1	204	257	141
from 2	150	122	219
from 3	142	82	309

(d)

	to 1	to 2	to 3
from 1	105	61	193
from 2	104	99	76
from 3	209	117	78

Transition frequencies and median inter-occurrence times appear in Table 2 for both the MR_3 and the MR_4 datasets. The Dirichlet hyperparameters $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_s$ are set equal to the rows of Table 2(b) plus one. The medians are reported because we have selected $q = 0.5$ in Table 1: the medians in Table 2(d) are smaller than the medians in Table 2(c) in six entries out of nine, in some cases considerably smaller.

2. Let us consider the predictive check mentioned above. Figure 3 shows posterior predictive 95 percent probability intervals of the inter-occurrence times for every transition, with the observed inter-occurrence times superimposed. These are empirical intervals computed by generating stochastic inter-occurrence times from their relevant distributions at every iteration of the Gibbs sampler. Possible outliers, represented as triangles, are those times with posterior predictive tail probability less than 2.5 percent. In Table 3 we report the expected value (and standard deviation) of the inter-occurrence times. In Table 4 the numbers of upper and lower extreme points and their overall percentage are collected. While deviations from the nominal 95% coverage are acceptable for transitions with low absolute frequency, such as $(2, 3)$, $(3, 3)$, $(3, 2)$, the remaining transitions require attention. We see that the percentage of outliers higher than the nominal value is mostly due to the upper outliers, which occur as an effect of the difference between the prior opinion on the marginal median of the inter-occurrence times and the median of the observed sequence (compare Table 2(d) to Table 2(c)). A few really extreme inter-occurrence times, such as the small values observed at transitions $(1, 1)$, $(2, 1)$ and the large one at transition $(1, 3)$, match unsurprisingly the outlying points in the corresponding qq-plots in Figure 2. This fact could be regarded as a lack of fit of the Weibull model, but it

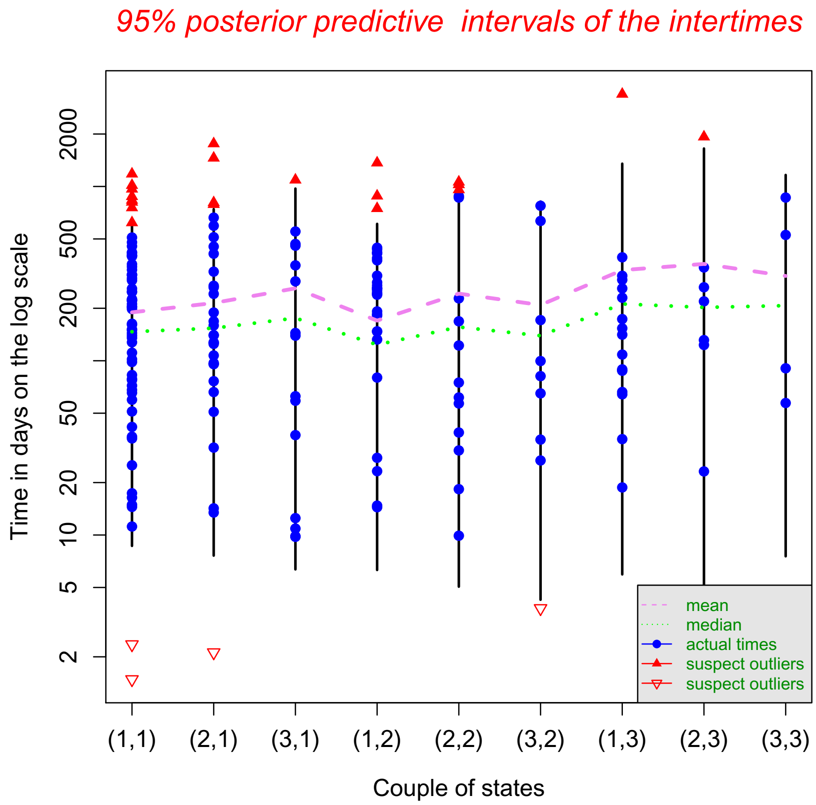

Fig 3. *Posterior predictive* 95 *percent credible intervals of the inter-occurrence times in days with actual times denoted by (blue) solid dots. Suspect outliers are denoted by (red)-pointing triangles. The (green) dotted line shows the posterior median and the (violet) dashed line the posterior mean. The prior distribution was elicited from the MR_4 learning set.*

Table 3
Predictive means (and standard deviations) of inter-occurrence times for each transition (in days); prior elicited from MR_4

	1	2	3
1	191 (12)	172 (18)	331 (70)
2	214 (22)	238 (43)	354 (145)
3	263 (58)	203 (55)	314 (134)

Table 4
Number of points having lower or upper posterior predictive tail probability less than 2.5 *percent and their percentage; prior elicited from MR_4*

	Upper outliers			Lower outliers			% of outliers		
	1	2	3	1	2	3	1	2	3
1	8	3	1	2	0	0	15.4%	10.0%	5.9%
2	4	3	1	1	0	0	15.6%	20.0%	14.3%
3	1	0	0	0	1	0	6.7%	11.1%	0.0%

TABLE 5
Posterior means (with standard deviations) of the shape parameter α in (a) and of the scale parameter θ in (b); prior elicited from MR_4

(a) Shape parameter α

	1	2	3
1	1.18 (0.06)	1.07 (0.08)	0.94 (0.10)
2	1.07 (0.07)	0.95 (0.10)	0.89 (0.14)
3	1.04 (0.16)	1.03 (0.16)	1.11 (0.21)

(b) Scale parameter θ

	1	2	3
1	201.7 (13.2)	175.6 (19.1)	317.8 (67.0)
2	219.2 (22.5)	231.0 (40.5)	327.3 (132.4)
3	262.7 (57.2)	201.9 (52.2)	320.1 (133.5)

TABLE 6
Summaries of the posterior distributions of the transition matrix **p**. *Posterior means (with standard deviations); prior elicited from MR_4*

	1	2	3
1	0.614 (0.028)	0.280 (0.026)	0.106 (0.018)
2	0.626 (0.041)	0.290 (0.038)	0.085 (0.023)
3	0.479 (0.070)	0.361 (0.067)	0.160 (0.051)

could also be due to an imperfect assignment of some events to the macroregion MR_3 or to an insufficient filtering of secondary events (i.e. aftershocks and foreshocks): earthquakes incorrectly assigned to MR_3 and aftershocks or foreshocks can give rise to very short inter-occurrence times; on the other hand, earthquakes which should be in MR_3 but which were attributed to other macroregions can produce very long inter-occurrence times.

The shape parameters α_{ij} are particularly important as they reflect an increasing hazard if larger than 1, a decreasing hazard if smaller than 1 and a constant hazard if equal to 1. Table 5(a) displays the posterior means of these parameters (along with their posterior standard deviations). Finally, Table 6 shows the posterior means of the transition probabilities. Notice that the last row departs from the other two; we will return to this below.

3. Cross state-probability plots are an attempt at predicting what type of event and when it is most likely to occur. A cross state-probability (CSP) $P^{ij}_{t_0|\Delta x}$ represents the probability that the next event will be in state j within a time interval Δx under the assumption that the previous event was in state i and t_0 time units have passed since its occurrence:

$$P^{ij}_{t_0|\Delta x} = P\left(J_{n+1} = j,\ X_{n+1} \leq t_0 + \Delta x \middle|\ J_n = i,\ X_{n+1} > t_0\right)$$
$$= \frac{p_{ij}\left(\bar{F}_{ij}(t_0) - \bar{F}_{ij}(t_0 + \Delta x)\right)}{\sum_{k \in E} p_{ik} \bar{F}_{ik}(t_0)}. \quad (16)$$

Figure 4 displays the CSPs with time origin at 31 December 2002, the closing date of the CPTI04 (2004) catalogue. At this time, the last recorded event had

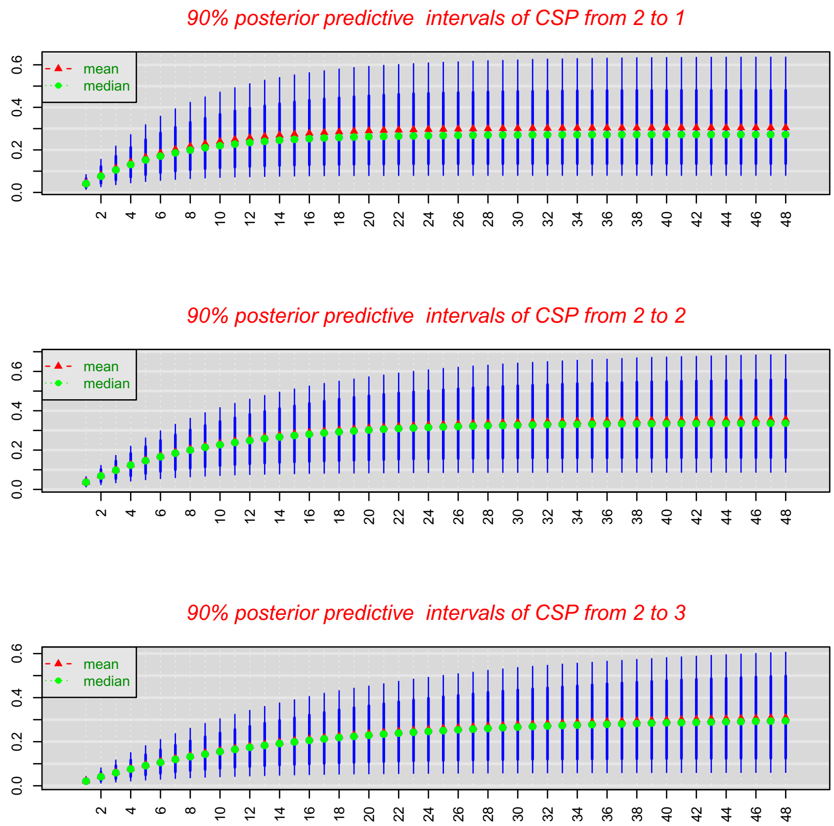

Fig 4. *Posterior mean and median of CSPs with time origin on 31 December 2002 up to 48 months ahead, along with 90 percent posterior credible intervals. Transitions are from state 2 to state 1, 2 or 3 (first to third panel, respectively). Months since 31 December 2002 are along the x-axis. The learning set is* MR_4.

Table 7

CSPs with time origin on 31 December 2002, as represented in Figure 4; prior elicited from MR_4

	1 Month	2 Months	3 Months	4 Months	5 Months	6 Months	1 Year	2 Years	3 Years	4 Years
to 1	0.045	0.080	0.113	0.140	0.164	0.184	0.256	0.296	0.303	0.304
to 2	0.038	0.069	0.099	0.125	0.149	0.169	0.257	0.327	0.348	0.356
to 3	0.023	0.041	0.061	0.078	0.095	0.109	0.180	0.256	0.291	0.309

been in class 2 and had occurred 965 days earlier (so t_0 is about 32 months). From these plots we can read out the probability that an event of any given type will occur before a certain number of months. For example, after 24 months, the sum of the mean CSPs in the three graphs indicates that the probability that at least one event will have occurred is around 88%, with a larger probability assigned to an event of type 2, followed by type 1 and type 3. The posterior means of the CSPs are also reported in Table 7.

TABLE 8

CSPs as the end of the catalogue shifts back by one-year steps. The numbers in boxes are the probability that the next observed event has occurred at or before the time when it occurred and is of the type that has been observed. The prior was elicited from MR_4

end of catalogue: 31/12/2001; previous event type:2; inter-occurrence time: 600 days											
	1 Month	2 Months	3 Months	4 Months	5 Months	6 Months	1 Year	392 days	2 Years	3 Years	4 Years
to 1	0.069	0.122	0.173	0.215	0.251	0.282	0.392	**0.401**	0.451	0.461	0.462
to 2	0.038	0.068	0.097	0.123	0.146	0.166	0.248	0.256	0.310	0.328	0.333
to 3	0.015	0.027	0.039	0.050	0.061	0.070	0.113	0.118	0.158	0.178	0.187
end of catalogue: 31/12/2000; previous event type: 2; inter-occurrence time: 235 days											
	1 Month	2 Months	3 Months	4 Months	5 Months	6 Months	1Year	2 Years	757 days	3 Years	4 Years
to 1	0.085	0.152	0.217	0.271	0.318	0.358	0.506	0.586	**0.587**	0.598	0.600
to 2	0.035	0.063	0.090	0.113	0.133	0.151	0.222	0.273	0.274	0.286	0.289
to 3	0.009	0.017	0.024	0.031	0.037	0.042	0.066	0.089	0.090	0.099	0.103
end of catalogue: 31/12/1999; previous event type: 1; inter-occurrence time: 177 days											
	1 Month	2 Months	3 Months	4 Months	130 days	5 Months	6 Months	1 Year	2 Years	3 Years	4 Years
to 1	0.103	0.186	0.261	0.323	0.339	0.376	0.420	0.565	0.622	0.626	0.627
to 2	0.042	0.075	0.104	0.127	**0.133**	0.147	0.163	0.216	0.238	0.240	0.240
to 3	0.012	0.023	0.033	0.042	0.044	0.050	0.057	0.088	0.116	0.126	0.130
end of catalogue: 31/12/1998; previous event type: 3; inter-occurrence time: 280 days											
	1 Month	2 Months	3 Months	4 Months	5 Months	6 Months	188 days	1 Year	2 Years	3 Years	4 Years
to 1	0.057	0.103	0.147	0.183	0.217	0.245	**0.251**	0.356	0.431	0.451	0.457
to 2	0.046	0.081	0.114	0.142	0.166	0.185	0.189	0.257	0.299	0.308	0.311
to 3	0.024	0.043	0.062	0.078	0.094	0.107	0.109	0.162	0.204	0.217	0.222
end of catalogue: 31/12/1997; previous event type: 3; inter-occurrence time: 442 days											
	1 Month	2 Months	85 days	3 Months	4 Months	5 Months	6 Months	1 Year	2 Years	3 Years	4 Years
to 1	0.061	0.110	0.150	0.157	0.196	0.232	0.263	0.383	0.468	0.492	0.499
to 2	0.044	0.077	0.104	0.109	0.135	0.158	0.177	0.247	0.291	0.301	0.304
to 3	0.021	0.038	**0.052**	0.055	0.069	0.081	0.092	0.136	0.169	0.180	0.184
end of catalogue: 31/12/1996; previous event type: 3; inter-occurrence time: 77 days											
	1 Month	2 Months	3 Months	4 Months	5 Months	6 Months	1 Year	450 days	2 Years	3 Years	4 Years
to 1	0.061	0.108	0.154	0.192	0.227	0.256	0.370	0.398	0.444	0.462	0.467
to 2	0.055	0.097	0.138	0.171	0.200	0.224	0.312	0.332	0.360	0.369	0.372
to 3	0.018	0.033	0.048	0.060	0.072	0.081	0.120	**0.130**	0.146	0.153	0.155
end of catalogue: 31/12/1995; previous event type: 1; inter-occurrence time: 3100 days											
	1 Month	2 Months	3 Months	4 Months	5 Months	6 Months	288 days	1 Year	2 Years	3 Years	4 Years
to 1	0.082	0.141	0.190	0.227	0.256	0.278	0.324	0.340	0.359	0.360	0.360
2	0.118	0.208	0.286	0.348	0.399	0.440	0.535	0.573	0.630	0.637	0.639
3	0.000	0.000	0.001	0.001	0.001	0.001	**0.001**	0.001	0.001	0.001	0.001
end of catalogue: 31/12/1994; previous event type: 1; inter-occurrence time: 2735 days											
	1 Month	2 Months	3 Months	4 Months	5 Months	6 Months	1 Year	653 days	2 Years	3 Years	4 Years
to 1	0.085	0.144	0.196	0.23 5	0.267	0.291	0.358	0.377	0.379	0.380	0.380
to 2	0.114	0.198	0.274	0.334	0.385	0.425	0.555	0.607	0.611	0.618	0.619
to 3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.001**	0.001	0.001	0.001
end of catalogue: 31/12/1993; previous event type: 1; inter-occurrence time: 2370 days											
	1 Month	2 Months	3 Months	4 Months	5 Months	6 Months	1 Year	2 Years	1018 days	3 Years	4 Years
to 1	0.090	0.153	0.209	0.251	0.285	0.310	0.384	0.406	0.407	0.407	0.408
to 2	0.108	0.188	0.260	0.317	0.366	0.404	0.530	0.584	0.590	0.591	0.592
to 3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	**0.001**	0.001	0.001
end of catalogue: 31/12/1992; previous event type: 1; inter-occurrence time: 2005 days											
	1 Month	2 Months	3 Months	4 Months	5 Months	6 Months	1 Year	2 Years	3 Years	1383 days	4 Years
to 1	0.082	0.140	0.192	0.231	0.262	0.286	0.357	0.378	0.380	0.380	0.380
to 2	0.106	0.185	0.258	0.317	0.367	0.407	0.544	0.608	0.617	0.618	0.618
to 3	0.000	0.000	0.000	0.001	0.001	0.001	0.001	0.001	0.001	**0.001**	0.001

The predictive capability of our model can be assessed by marking the time of the next event on the relevant CSP plot. In our specific case, the first event in 2003 which can be assigned to the macroregion MR_3 happened in the Forlì area on 26 January and was of type 1, with a CSP of 4.5%. This is a low probability, but a single case is not enough to judge our model, which would be a bad one if repeated comparisons did not reflect the pattern represented by the CSPs. Therefore, we repeated the same comparison by re-estimating the model using only the data up to 31 December 2001, 31 December 2000, and so on backwards down to 1992. The results are shown in Table 8. The boxed numbers correspond to the observed events and it is a good sign that they do not always correspond to very high or very low CSPs (relative to the remaining numbers on the same line) as this would indicate that events occur too late or too early compared to the estimated model. Correspondingly, if we were to plot the conditional densities obtained by differentiating the CSPs with respect to Δx, marking the observed inter-occurrence times on the x-axis, we would observe that few of them appear in the tails, with the exception of the last three arrays in the table. This exception is related to the period from the last event in July 1987 to the event in October 1996, during which no events with $M_w \geq 4.5$ occurred: this is the longest inter-occurrence time in the whole series and it is the outlying point in the top right panel of Fig. 2, so it is not surprising that the boxed CSPs are small. With regard to the existence of occasional very long interoccurrence times, a reviewer noted that it is very difficult to find a parametric model that would explain them and that perhaps an outlier accommodation mechanism is needed, although this might disrupt the Markov renewal assumption.

5.1. Time predictable and slip predictable models

The examination of the posterior distributions of transition probabilities and of the predictive distributions of the inter-occurrence times can give some insight into the type of energy release and accumulation mechanism. We consider the two mechanisms mentioned above: the time predictable model (TPM) and the slip predictable model (SPM).

In the TPM, it is assumed that when a maximal energy threshold is reached, some fraction of it (not always the same) is released and an earthquake occurs. The consequence is that the time until the next earthquake increases with the severity of the last earthquake. Thus, the inter-occurrence time distribution depends on the current event type, but not on the next event type, that is, we expect $F_{ij}(t) = F_{i\cdot}(t)$, $j = 1, 2, 3$. The strength of an event does not depend on the strength of the previous one, because every time the same energy level has to be reached for the event to occur. So we expect $p_{ij} = p_{\cdot j}$, $j = 1, 2, 3$, that is, a transition matrix with equal rows. If this is the case, the CSPs (16) would simplify as follows,

$$P^{ij}_{t_0|\Delta x} = \frac{p_{ij}\left(\bar{F}_{ij}(t_0) - \bar{F}_{ij}(t_0 + \Delta x)\right)}{\sum_{k\in E} p_{ik}\bar{F}_{ik}(t_0)} = \frac{p_{\cdot j}\left(\bar{F}_{i\cdot}(t_0) - \bar{F}_{i\cdot}(t_0 + \Delta x)\right)}{\bar{F}_{i\cdot}(t_0)}, \quad (17)$$

so that, under the TPM assumption, given i, they are proportional to each other as $j = 1, 2, 3$ for any Δx, and the ratio $P^{ij}_{t_0|\Delta x}/P^{ik}_{t_0|\Delta x}$ equals $p_{\cdot j}/p_{\cdot k}$ for any pair (j, k).

In the SPM, after an event, energy falls to a minimal threshold and increases until the next event, where it starts to increase again from the same threshold. The consequence is that the energy of the next earthquake increases with time since the last earthquake. So, the magnitude of an event depends on the length of the inter-occurrence time, but not on the magnitude of the previous one, because energy always accumulates from the same threshold. In this case $p_{ij} = p_{\cdot j}$, but

TABLE 9
Posterior means (with standard deviations) of the transition matrix **p** *with the noninformative prior*

	1	2	3
1	0.567 (0.046)	0.271 (0.041)	0.162 (0.033)
2	0.567 (0.064)	0.283 (0.059)	0.150 (0.046)
3	0.500 (0.085)	0.323 (0.079)	0.177 (0.064)

$F_{ij}(t) = F_{\cdot j}(t)$ again, so

$$P^{ij}_{t_0|\Delta x} = \frac{p_{\cdot j}\left(\bar{F}_{\cdot j}(t_0) - \bar{F}_{\cdot j}(t_0 + \Delta x)\right)}{\sum_{k \in E} p_{\cdot k} \bar{F}_{\cdot k}(t_0)}. \tag{18}$$

Then, under the SPM assumption, the CSP are equal to each other as $i = 1, 2, 3$ for any Δx, given j.

An additional feature that can help discriminate between TPM and SPM is the tail of the inter-occurrence time distribution: for a TPM, the tail of the inter-occurrence time distribution is thinner *after* a weak earthquake than *after* a strong one; for an SPM, the tail of the inter-occurrence time is thinner *before* a weak earthquake than *before* a strong one.

In the present case the posterior mean of the third row of **p**, see Table 6, is clearly different from the other two rows, unlike the empirical transition matrix derived from Table 2(a), because of the prior information from MR$_4$. So, with this prior, both the TPM and the SPM are excluded.

On the other hand, things change with the noninformative prior elicited without a learning set. In this case, we let all the Dirichlet hyperparameters γ_{ij}'s be equal to 2. Following Section 4.3, the missing learning set for each string (i, j) is substituted by a unique fictitious observation $\tilde{t}^{ij}_{0.5}$ uniformly distributed over $(1, 5000)$ days, and m_{ij} is set to one; this establishes the prior for θ_{ij}. The prior of α_{ij} derived from Equation (15) with $m_{ij} = 1$ and $c_{ij} = 2$ (taken from Table 1) is

$$\pi_3(\alpha_{ij}) \propto \left(1 - \frac{\alpha_{ij}}{\alpha_{0,ij}}\right) \mathbb{1}(\alpha_{0,ij} \leq \alpha_{ij} \leq \alpha_{1,ij})$$

with $\alpha_{0,ij} = 2/(2 + N_{ij})$ (see Table 2(a) for the N_{ij}'s) and $\alpha_{1,ij} = 10$. Note that on our current sample, the lower limit $\alpha_{0,ij}$ is always smaller than $2/3$.

With this prior specification, the posterior distributions of the rows of the transition matrix do not differ significantly, as seen from Table 9, so we can assume $p_{ij} = p_{\cdot j}$ for all indexes i and examine the ratios of CSPs to verify the TPM and the SPM hypotheses.

Figures 5 and 6 display the posterior means of the ratios of the CSPs as a function of Δx for $t_0 = 0$, with the noninformative prior. For the TPM the generic ratio of two CSPs indexed by (i, j) and (i, k) should be approximately constant and close to $p_{\cdot j}/p_{\cdot k}$, where the $p_{\cdot j}$ represents the common values of the entries in the j-th column of **p**, under the TPM. The horizontal lines in Figure 5 are the posterior expectations of p_{ij}/p_{ik}, which would estimate $p_{\cdot j}/p_{\cdot k}$ if the TPM assumption were true. For the SPM, the ratio of CPSs, now indexed

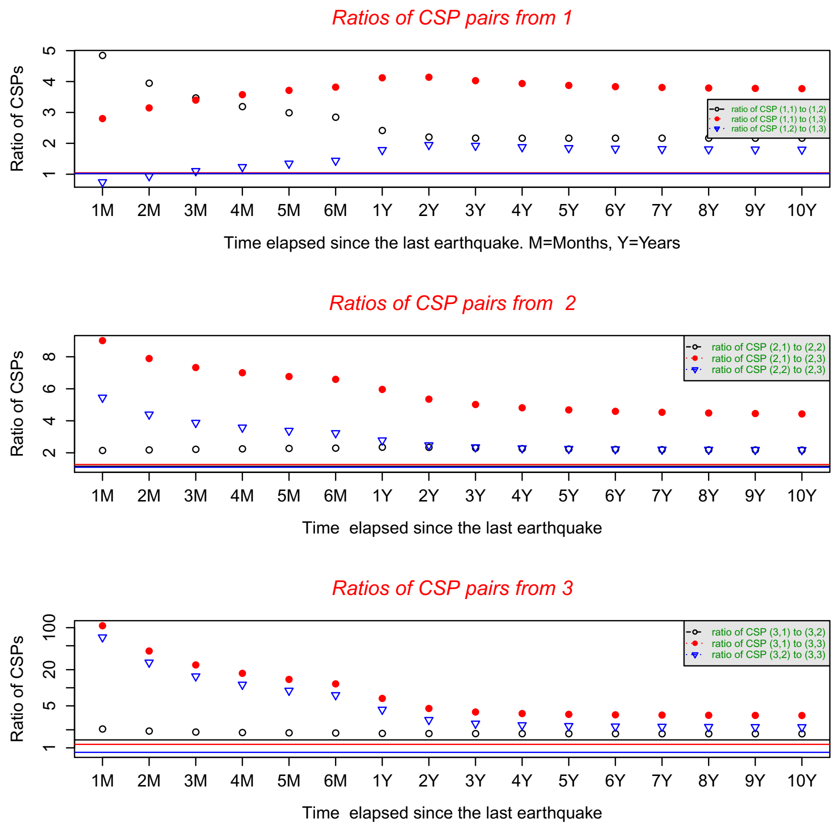

FIG 5. *Checking the TPM: posterior means of ratios of CSPs* $P^{ij}_{0|\Delta x}/P^{ik}_{0|\Delta x}$*, with time origin at 0, up to 10 years ahead. Transitions are from state 1, 2 and 3 to state 1, 2 or 3. Horizontal lines indicate the theoretical values of the ratios for the TPM. The prior distribution is noninformative.*

by (i, j) and (k, j), should be close to one. The plots indicate that it is not so, therefore neither the TPM nor the SPM are supported by the data, even with the noninformative prior.

As for the TPM, this finding is confirmed by the examination of the posterior probabilities that $\alpha_{ij} < \alpha_{ik}$ and $\theta_{ij} < \theta_{ik}$, for any given i and $j \neq k$: $\Pr(\alpha_{ij} < \alpha_{ik}|\mathbf{j}, \mathbf{x}, u_T)$ is 0.57 for string $(2, 1)$ versus string $(2, 3)$ and 0.53 for $(3, 1)$ versus $(3, 2)$, but it is either larger than 0.74 or smaller than 0.37 for all the other strings; $\Pr(\theta_{ij} < \theta_{ik}|\mathbf{j}, \mathbf{x}, u_T)$ is 0.59 for $(1, 2)$ versus $(1, 3)$ and is 0.55 for $(3, 1)$ versus $(3, 2)$, but it is lower than 0.39 for all the other strings. As for the SPM, we have examined $\Pr(\alpha_{ij} < \alpha_{kj}|\mathbf{j}, \mathbf{x}, u_T)$ and $\Pr(\theta_{ij} < \theta_{kj}|\mathbf{j}, \mathbf{x}, u_T)$ for any j and $i \neq k$: $\Pr(\alpha_{ij} < \alpha_{kj}|\mathbf{j}, \mathbf{x}, u_T)$ is 0.53 for $(2, 2)$ versus $(3, 2)$, but it is either larger than 0.74 or smaller than 0.37 for all the other strings; $\Pr(\theta_{ij} < \theta_{kj}|\mathbf{j}, \mathbf{x}, u_T)$ is between 0.44 and 0.62 for three comparisons but it is either larger than 0.72 or smaller than 0.30 for the remaining ones.

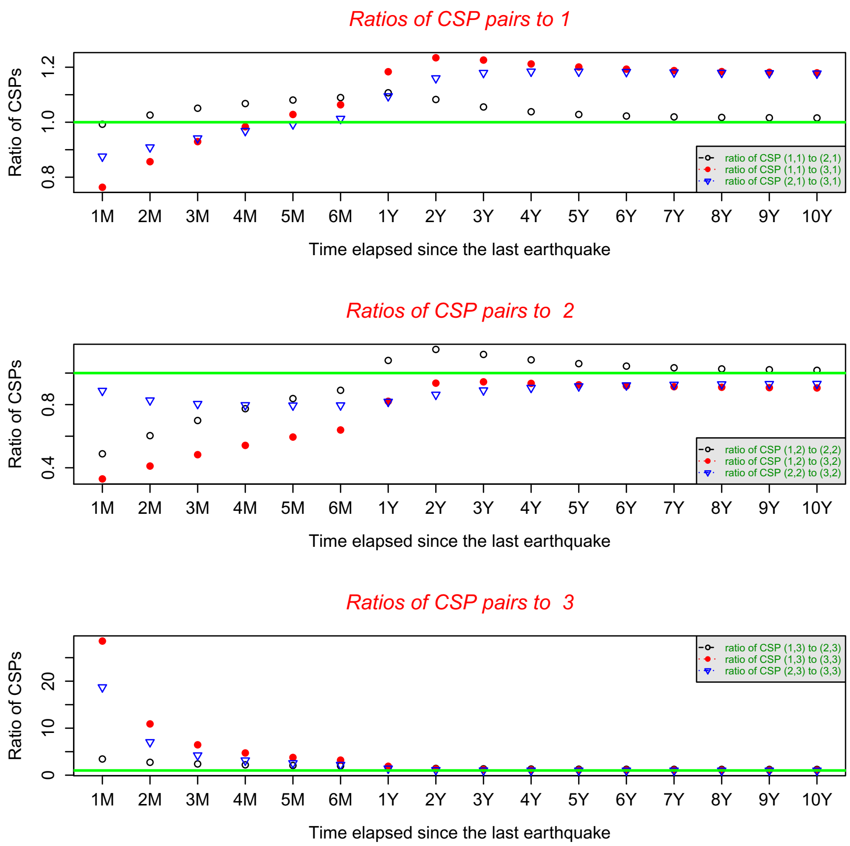

FIG 6. *Checking the SPM: posterior means of ratios of CSPs* $P^{ij}_{0|\Delta x}/P^{kj}_{0|\Delta x}$*, with time origin at 0, up to 10 years ahead. Transitions are from state 1, 2 and 3 to state 1, 2 or 3. Green horizontal lines at one indicate the theoretical value of the ratios for the SPM. The prior distribution is noninformative.*

6. Concluding remarks

We have presented a complete Bayesian methodology for the inference on semi-Markov processes, from the elicitation of the prior distribution, to the computation of posterior summaries including predictive probabilities of future events. A guidance for its JAGS implementation is given too. In particular, we have examined in detail the elicitation of the joint prior density of the shape and scale parameters of the Weibull-distributed inter-occurrence times (conditional on the transition between two given states), deriving a specific class of priors in a natural way, along with a method for the determination of hyperparameters based on "learning data" and moment existence conditions. This framework has been applied to the analysis of seismic data, but it can be adopted for inference on any system for which a Markov Renewal process is plausible.

With regard to the seismic data analysis, other uses of our model can be envisaged. The model can be applied to areas with a less complex tectonics, such as Turkey, by replicating for example Alvarez's analysis. Outliers, such as those appearing in Figure 3, could point at events whose assignment to a specific seismogenic source should be re-discussed. The analysis of earthquake occurrence can support decision making related to the risk of future events. We have not examined this issue here, but a methodology is outlined by Cano et al. (2011).

In our analysis we do not take into account the spatio-temporal interactions between earthquakes. The inclusion of the spatial information in a more advanced stochastic model would certainly improve the prediction capabilities of earthquakes. A first step in this direction could be to use the data of the entire catalogue, organized in homogeneous subregions, and build a Bayesian hierarchy of semi-Markov models which enable sharing of informations across regions, while acknowledging existing local differences.

A final note concerns the more recent Italian seismic catalogue CPTI11 (2011), including events up to the end of 2006. Every new release of the catalogue involves numerous changes in the parameterization of earthquakes; as the DISS event classification by macroregion is not yet available for events in this catalogue we cannot use this more recent source of data.

Appendix A: Gibbs sampling

Here we derive the full conditional distributions involved in the Gibbs sampling, we give indications on its JAGS implementation and we reproduce the JAGS code.

A.1. Full conditional distributions

Let the last inter-occurrence time be censored i.e. $u_T > 0$. Hence, in order to obtain some simple full conditional distributions and then an efficient Gibbs sampling, we introduce the auxiliary variable $j_{\tau+1}$ which represents the unobserved state following the last visited state j_τ. Moreover, let $\mathbf{t_q} = (t^{ij}_{q_{ij}},\ i,j \in E)$. Each hyperparameter $t^{ij}_{q_{ij}}$ may be either a known constant or $t^{ij}_{q_{ij}}$ is uniformly distributed over an interval (t_1, t_2). Moreover, all of them are independent of each other. Thus, the state space of the Gibbs sampler is $(\mathbf{p}, \boldsymbol{\alpha}, \boldsymbol{\theta}, j_{\tau+1}, \mathbf{t_q})$ and the following full likelihood derived from (5):

$$L(\mathbf{j}, \mathbf{x}, u_T, j_{\tau+1} | \mathbf{p}, \boldsymbol{\alpha}, \boldsymbol{\theta}) = \prod_{i,k \in E} p_{ik}^{N_{ik}} \times \prod_{i,k \in E} \left[\alpha_{ik}^{N_{ik}} \frac{1}{\theta_{ik}^{\alpha_{ik} N_{ik}}} \left(\prod_{\rho=1}^{N_{ik}} x_{ik}^{\rho} \right)^{\alpha_{ik}-1} \times \exp\left\{ -\frac{1}{\theta_{ik}^{\alpha_{ik}}} \sum_{\rho=1}^{N_{ik}} (x_{ik}^{\rho})^{\alpha_{ik}} \right\} \right] \times \left(p_{j_\tau j_{\tau+1}} \exp\left\{ -\left(\frac{u_T}{\theta_{j_\tau j_{\tau+1}}} \right)^{\alpha_{j_\tau j_{\tau+1}}} \right\} \right) \quad (19)$$

is multiplied by the prior and used to determine the full conditionals. For every i and j, let

$$\begin{aligned}
\mathbf{p}_{(-i)} &= \text{the transition matrix } \mathbf{p} \text{ without the } i\text{-th row},\\
\boldsymbol{\alpha}_{(-ij)} &= (\alpha_{hk},\ \ h,k \in E, \quad (h,k) \neq (i,j)),\\
\boldsymbol{\theta}_{(-ij)} &= (\theta_{hk},\ \ h,k \in E, \quad (h,k) \neq (i,j)),\\
\mathbf{t}_{\mathbf{q}(-ij)} &= (t^{hk}_{q_{hk}},\ \ h,k \in E, \quad (h,k) \neq (i,j)),\\
\tilde{N}_{ij} &= N_{ij} + \mathbb{1}((j_\tau, j_{\tau+1}) = (i,j)), \quad \tilde{\mathbf{N}}_i = \left(\tilde{N}_{ij},\ \ j = 1,\dots,s\right),\\
\tilde{M}_{ij}(\alpha_{ij}) &= \sum_{\rho=1}^{N_{ij}} (x^\rho_{ij})^{\alpha_{ij}} + u_T^{\alpha_{ij}} \mathbb{1}((j_\tau, j_{\tau+1}) = (i,j)),\\
\mathcal{C}_{ij} &= \prod_{\rho=1}^{N_{ij}} x^\rho_{ij}.
\end{aligned}$$

The following result on the full conditional distributions of the Gibbs sampling holds.

Proposition A.1. *Let the prior on* $(\mathbf{p},\ \boldsymbol{\alpha},\ \boldsymbol{\theta},\ \mathbf{t_q})$ *be the following*

i) $\mathbf{p}$ *is independent of* $\boldsymbol{\alpha}$ *and* $\boldsymbol{\theta}$ *and the rows of* $\mathbf{p}$ *are* s *independent vectors with Dirichlet distribution with parameters* $\boldsymbol{\gamma}_1,\dots,\boldsymbol{\gamma}_s$ *and total mass* $c_1,\dots c_s$*, respectively,*

ii) the θ_{ij}*'s, given the* α_{ij}*'s and the* $t^{ij}_{q_{ij}}$*'s, are independent with* $\theta_{ij}|\alpha_{ij} \sim \mathcal{G}IG(m_{ij}, b_{ij}(t^{ij}_{q_{ij}}, \alpha_{ij}), \alpha_{ij})$*, where*

$$b_{ij}(t^{ij}_{q_{ij}}, \alpha_{ij}) = \left(t^{ij}_{q_{ij}}\right)^{\alpha_{ij}} [(1-q_{ij})^{-1/m_{ij}} - 1]^{-1},$$

and $t^{ij}_{q_{ij}}$ *is either a known constant or* $t^{ij}_{q_{ij}}$ *is uniformly distributed over* (t_1, t_2)*,*

iii) $\pi_{3,ij}(\alpha_{ij}) \propto \alpha_{ij}^{m_{ij}-c_{ij}} (\alpha_{ij} - \alpha_{0,ij})^{c_{ij}-1} \exp\{-m_{ij} d_{ij} \alpha_{ij}\} \mathbb{1}(\alpha_{ij} \in I_{ij})$, $m_{ij} > 0, c_{ij} > 0$, $\alpha_{0,ij} > 0$ *and* $d_{ij} \geq 0$ *where*

$$I_{ij} = \begin{cases} (\alpha_{0,ij}, \alpha_{1,ij}) & \text{if } d_{ij} = 0 \\ (\alpha_{0,ij}, \infty) & \text{if } d_{ij} > 0. \end{cases}$$

Then,

a) the conditional distribution of $\mathbf{p}_i$*, given* $\mathbf{j},\ \mathbf{x},\ u_T,\ j_{\tau+1},\ \mathbf{p}_{(-i)},\ \boldsymbol{\alpha},\ \boldsymbol{\theta}$ *and* $\mathbf{t_q}$ *is a Dirichlet distribution with parameter* $\tilde{\mathbf{N}}_i + \boldsymbol{\gamma}_i$*;*

b) the conditional distribution of $\theta_{ij}^{\alpha_{ij}}$*, given* $\mathbf{j},\ \mathbf{x},\ u_T,\ j_{\tau+1},\ \mathbf{p},\ \boldsymbol{\alpha},\ \boldsymbol{\theta}_{(-ij)}$ *and* $\mathbf{t_q}$ *is an inverse Gamma distribution with shape* $m_{ij}+N_{ij}$ *and rate* $b_{ij}(t^{ij}_{q_{ij}}, \alpha_{ij}) + \tilde{M}_{ij}(\alpha_{ij})$*;*

c) the conditional density of α_{ij}*, given* $\mathbf{j},\ \mathbf{x},\ u_T,\ j_{\tau+1},\ \mathbf{p},\ \boldsymbol{\alpha}_{(-ij)},\ \boldsymbol{\theta}$ *and* $\mathbf{t_q}$ *is proportional to*

$$\alpha_{ij}^{N_{ij}+1+m_{ij}-c_{ij}}\,(\alpha_{ij}-\alpha_{0,ij})^{c_{ij}-1} \times \exp\left\{-\left(m_{ij}d_{ij}-\log\frac{\mathcal{C}_{ij}(t^{ij}_{q_{ij}})^{m_{ij}}}{\theta_{ij}^{N_{ij}+m_{ij}}}\right)\alpha_{ij}\right\} \times \exp\left\{-\frac{b_{ij}(t^{ij}_{q_{ij}},\alpha_{ij})+\tilde{M}_{ij}(\alpha_{ij})}{\theta_{ij}^{\alpha_{ij}}}\right\}\mathbb{1}(\alpha_{ij}\in I_{ij}), \quad (20)$$

and it is log-concave if $c_{ij}\geq 1$*;*

d) the conditional density of the unseen state $J_{\tau+1}$*, given* $\mathbf{j},\ \mathbf{x},\ u_T,\ \mathbf{p},\ \boldsymbol{\alpha},\boldsymbol{\theta}$ *and* $\mathbf{t_q}$*, is*

$$\frac{p_{j_\tau j}\exp\left\{-\left(\frac{u_T}{\theta_{j_\tau j}}\right)^{\alpha_{j_\tau j}}\right\}}{\sum_{k\in E}p_{j_\tau k}\exp\left\{-\left(\frac{u_T}{\theta_{j_\tau k}}\right)^{\alpha_{j_\tau k}}\right\}};$$

e) if $t^{ij}_{q_{ij}}$ *is uniformly distributed over* (t_1,t_2)*, then its conditional distribution given* $\mathbf{j},\ \mathbf{x},\ u_T,\ \mathbf{p},\ \boldsymbol{\alpha},\ \boldsymbol{\theta}$ *and* $\mathbf{t}_{\mathbf{q}(-ij)}$ *is a doubly-truncated at* (t_1,t_2) *generalized Gamma with parameters* $a=\theta_{ij}[(1-q_{ij})^{-1/m_{ij}}-1]^{1/\alpha_{ij}}$, $d=\alpha_{ij}m_{ij}+1$ *and* $p=\alpha_{ij}$*, i.e.*

$$\pi(t^{ij}_{q_{ij}}|\mathbf{j},\mathbf{x},u_T,\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\theta},\mathbf{t}_{\mathbf{q}(-ij)}) = \frac{p/a^d}{\Gamma(d/p)}\left(t^{ij}_{q_{ij}}\right)^{d-1}\exp\left\{-\left(t^{ij}_{q_{ij}}/a\right)^p\right\}\mathbb{1}(t_1<t^{ij}_{q_{ij}}<t_2).$$

Proof. As the row $\mathbf{p}_i$ is independent of $(\mathbf{p}_{(-i)},\boldsymbol{\alpha},\boldsymbol{\theta},\mathbf{t_q})$, conditionally on the data and $j_{\tau+1}$, then

$$\pi(\mathbf{p}_i|\mathbf{j},\mathbf{x},u_T,\ j_{\tau+1},\mathbf{p}_{(-i)},\boldsymbol{\alpha},\boldsymbol{\theta},\mathbf{t_q}) \propto L(\mathbf{j},\mathbf{x},u_T,\ j_{\tau+1}|\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\theta})\times\pi_{1,i}(\mathbf{p}_i)\propto\prod_{j\in E}p_{ij}^{\tilde{N}_{i,j}}\times\prod_{j\in E}p_{ij}^{\gamma_{ij}-1},$$

where $\pi_{1,i}$ denotes the Dirichlet prior of $\mathbf{p}_i$. Hence point $a)$ of Proposition A.1 follows.

As regards the full conditional distribution of θ_{ij}, we have

$$\begin{aligned}\pi(\theta_{ij}\mid\mathbf{j},\mathbf{x},u_T,\ j_{\tau+1},\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\theta}_{(-ij)},\mathbf{t_q}) &\propto L(\mathbf{j},\mathbf{x},u_T,\ j_{\tau+1}\ |\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\theta})\times\pi_{2,ij}(\theta_{ij}|\alpha_{ij},t^{ij}_{q_{ij}})\\ &\propto\prod_{i,k\in E}\left[\alpha_{ik}^{N_{ik}}\frac{\mathcal{C}_{ik}^{\alpha_{ik}-1}}{\theta_{ik}^{\alpha_{ik}N_{ik}}}\times\exp\left\{-\frac{\tilde{M}_{ik}(\alpha_{ik})}{\theta_{ik}^{\alpha_{ik}}}\right\}\right]\times\exp\left\{-\frac{b_{ij}(t^{ij}_{q_{ij}},\alpha_{ij})}{\theta_{ij}^{\alpha_{ij}}}\right\}\\ &\times\theta_{ij}^{-[1+\alpha_{ij}m_{ij}]}\propto\ \theta_{ij}^{-[1+\alpha_{ij}(m_{ij}+N_{ij})]}\exp\left\{-\frac{b_{ij}(t^{ij}_{q_{ij}},\alpha_{ij})+\tilde{M}_{ij}(\alpha_{ij})}{\theta_{ij}^{\alpha_{ij}}}\right\}.\end{aligned}$$

As one can see, the last function is the kernel of an inverse Gamma distribution with parameters $m_{ij}+N_{ij}$ and $b_{ij}(t^{ij}_{q_{ij}},\alpha_{ij})+\tilde{M}_{i,j}(\alpha_{ij})$ and point $b)$ follows. A similar reasoning yields a full conditional distribution for α_{ij} proportional to (20). Furthermore, concerning its log-concavity, notice that the function in (20) can be written as the product of the following four log-concave functions:

$$\alpha_{ij}^{N_{ij}+m_{ij}},\ \Big(1-\frac{\alpha_{ij}}{\alpha_{0,ij}}\Big)^{c_{ij}-1},\ \exp\Big\{-\big(m_{ij}d_{ij}-\log\frac{\mathcal{C}_{ij}(t^{ij}_{q_{ij}})^{m_{ij}}}{\theta_{ij}^{N_{ij}+m_{ij}}}\big)\alpha_{ij}\Big\},$$
$$\exp\Big\{-\frac{b_{ij}(t^{ij}_{q_{ij}},\alpha_{ij})+\tilde{M}_{ij}(\alpha_{ij})}{\theta_{ij}^{\alpha_{ij}}}\Big\}$$

In particular, the second function is log-concave for $c_{ij}\geq 1$ and the last term is a product of log-concave functions of the kind $\alpha_{ij}\mapsto\exp\{-z^{\alpha_{ij}}\}$. Hence the log-concavity follows from the property that the product of log-concave functions is also log-concave.

Regarding point d), it is enough to observe that Equations (2) and (19) imply that

$$\begin{aligned}P(J_{\tau+1}=j|\,\mathbf{j},\mathbf{x},u_T,\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\theta}) &= \frac{P(J_{\tau+1}=j,\ X_{\tau+1}>u_T|\,\mathbf{j},\mathbf{x},\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\theta})}{\sum_{k\in E}P(J_{\tau+1}=k,\ X_{\tau+1}>u_T|\,\mathbf{j},\mathbf{x},\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\theta})}\\ &= \frac{p_{j_\tau j}\exp\left\{-\left(\frac{u_T}{\theta_{j_\tau j}}\right)^{\alpha_{j_\tau j}}\right\}}{\sum_{k\in E}p_{j_\tau k}\exp\left\{-\left(\frac{u_T}{\theta_{j_\tau k}}\right)^{\alpha_{j_\tau k}}\right\}}.\end{aligned}$$

Finally, if $t^{ij}_{q_{ij}}$ is uniformly distributed over (t_1,t_2), then

$$\begin{aligned}\pi(t^{ij}_{q_{ij}}\mid\mathbf{j},\mathbf{x},u_T,\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\theta},\mathbf{t}_{\mathbf{q}(-ij)}) &\propto \pi_2(\theta_{ij}|\alpha_{ij},t^{ij}_{q_{ij}})\mathbb{1}(t_1\leq t^{ij}_{q_{ij}}\leq t_2)\\ &\propto b^{m_{ij}}(t^{ij}_{q_{ij}},\alpha_{ij})\exp\left\{-\frac{b(t^{ij}_{q_{ij}},\alpha_{ij})}{\theta_{ij}^{\alpha_{ij}}}\right\}\mathbb{1}(t_1\leq t^{ij}_{q_{ij}}\leq t_2)\\ \propto \left(t^{ij}_{q_{ij}}\right)^{\alpha_{ij}m_{ij}}\exp&\left\{-\left(\frac{t^{ij}_{q_{ij}}}{\theta_{ij}[(1-q_{ij})^{-1/m_{ij}}-1]^{1/\alpha_{ij}}}\right)^{\alpha_{ij}}\right\}\mathbb{1}(t_1\leq t^{ij}_{q_{ij}}\leq t_2)\end{aligned}$$

The last equation is the kernel of a generalized Gamma, doubly-truncated at (t_1,t_2), as introduced in Stacy (1962), so point e) follows. □

A.2. JAGS implementation

Proposition A.1 implies that JAGS should be able to run an exact Gibbs sampler. The model description we have adopted in the JAGS language is based on the full likelihood (19). It is not important that the model description matches the actual model which generated the data as long as the full conditional distributions, which are determined by the joint distribution of the data and the parameters, remain unchanged. In detail, we consider the following joint distribution:

$$L_1(\mathbf{j},\mathbf{x},u_T|j_{\tau+1},\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\theta})\pi(j_{\tau+1}|\mathbf{p}_{j_\tau})\pi(\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\theta}|\mathbf{t}_\mathbf{q})\pi(\mathbf{t}_\mathbf{q})$$

where $\pi(\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\theta}|\mathbf{t}_\mathbf{q})$ is the joint prior of $(\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\theta})$, given the matrix of hyperparameters $\mathbf{t}_\mathbf{q}=(t^{ij}_{q_{ij}},\ \ i,j\in E)$ as derived in Section 4 using Equations (6), (14) and (15),

$$\pi(j_{\tau+1}|\mathbf{p}_{j_\tau}) = p_{j_\tau j_{\tau+1}}$$
$$L_1(\mathbf{j},\mathbf{x},u_T|j_{\tau+1},\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\theta}) = L(\mathbf{j},\mathbf{x},u_T,j_{\tau+1}|\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\theta})/p_{j_\tau j_{\tau+1}}.$$

and the density $\pi(\mathbf{t_q})$ of $\mathbf{t_q}$ is such that the $t^{ij}_{q_{ij}}$'s are independent and either uniformly distributed over (t_1,t_2) or concentrated on the suitable empirical quantile $\hat{t}^{ij}_{q_{ij}}$ obtained from the learning dataset (see Subsections 4.2 and 4.3).

The factors of the likelihood L_1, to be extracted from Equation (19), are modelled in JAGS as follows. For any value of i, the factor

$$\prod_{k\in E} p_{ik}^{N_{ik}}$$

is contributed by a multinomial likelihood with probability vector $\mathbf{p}_i$ and $\sum_k N_{ik}$ trials. The factors in square brackets in (19) are contributed by the uncensored Weibull inter-occurrence times for every string (i,k) and are obtained in JAGS as Weibull densities with parameters α_{ik} and θ_{ik}, using a "for" loop sweeping the strings. The last factor, which accounts for the censored inter-occurrence time u_T, is handled by a special instruction, by which u_T is first declared to be a right censored time with upper censoring point $T-t_\tau$, and then is assigned a Weibull distribution with parameters $\alpha_{j_\tau j_{\tau+1}}$ and $\theta_{j_\tau j_{\tau+1}}$.

The factor $\pi(j_{\tau+1}|\mathbf{p}_{j_\tau})\pi(\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\theta}|\mathbf{t_q})$, representing the prior associated with L_1, is handled as follows. The additional prior on $j_{\tau+1}$ is a discrete distribution on the integers 1, 2, 3, with probabilities taken from the row of $\mathbf{p}$ indexed by j_τ. Every row $\mathbf{p}_i$ of $\mathbf{p}$ is assigned a Dirichled distribution directly. Whenever $m_{ik}\geq 2$, as $c_{ik}=m_{ik}$, the shape parameters α_{ik} have a shifted Gamma prior, see Equation (15), obtainable by defining in JAGS a new non-shifted Gamma variable with shape c_{ik} and rate $m_{ik}d_{ik}$, which, after summing the shift, is assigned to α_{ik}; for the value of d_{ik} see Equation (13). The generalized Gamma for θ_{ik} is defined conditionally on α_{ik}: first a Gamma prior with shape m_{ij} and rate $\hat{b}_{ik}(\alpha_{ik})$ is assigned to a new random variable a_{ik} and then $a_{ik}^{-1/\alpha_{ik}}$ is assigned to θ_{ik}.

In case there is either just one observation or no learning dataset $\mathbf{y}_m$, some special instructions in the JAGS code are needed. In particular, if there is no learning dataset, then the missing learning dataset is substituted, for every string (i,k), by the fictitious observation $\tilde{t}_{ik}$ drawn from a uniform distribution over (1,5000) days (so $m_{ik}=1$ for all strings). Then, the priors of the θ_{ik} retain the same form, whereas the priors of the α_{ik}'s, derived from Equation (15) with $m_{ik}=1$ and $c=2$ (a value taken from Table 1), are

$$\pi_3(\alpha_{ik}) \propto \left(1-\frac{\alpha_{0,ik}}{\alpha_{ik}}\right)\mathbb{1}(\alpha_{0,ik}\leq\alpha_{ik}\leq\alpha_{1,ik})$$

with assigned $\alpha_{0,ik}$ and $\alpha_{1,ik}$. This latter distribution is coded using the so-called zeros trick: a fictitious zero observation from a Poisson distribution with mean $\phi_{ik}=-\ln\pi_3(\alpha_{ik})$ is introduced; then a uniform prior over $[\alpha_{0,ik},\alpha_{1,ik}]$ is assigned to α_{ik}. The effect on the formula of the joint distribution is that the likelihood L_1 gets multiplied by the factor $\exp(-\phi_{ik})$, contributed by the zero observation; the multiplication by the uniform density gives back the correct factor accounting for the prior of α_{ik}.

A.3. JAGS code

```
# JAGS CODE defining the model (likelihood and prior distributions)
# The code is tailored for 3 states, but can be adapted to a
    different number of states
#
# VARIABLES and CONSTANTS appearing in the script
#
# k: number of states (here k=3)
# transitions[,]: a 3 by 3 matrix containing observed frequencies
# P[,]: a 3 by 3 matrix containing the transition probabilities
# N.from[]: a length three vector; N.from[i] is the number of
    transitions which occurred in the data from state i, i = 1,2,3
# dirichlet.parameters[,]: a 3 by 3 matrix; the i-th row contains
    the hyperparameters of the Dirichlet distribution of the i-th
    row of P
# l: indexes the transition; it ranges from 1 to tau, which is the
    number of the last observed transition
# x.exact[]: a length tau vector of the observed inter-occurrence
    times
# ctf[] is a length tau vector; any entry can take value from 1 to
    9, indexing possible transitions in the following order: (1,1),
     (2,1), (3,1), (1,2), (2,2), (3,2), (1,3), (2,3), (3,3)
# alpha[] and a[] are length nine vectors of the shape and rate
    parameters of the Weibull inter-occurrence times: positions in
    the vectors are matched to the transition with the same rule
    used for ctf[] just above
# j.state[]: a length tau vector of the sequence of observed states
    , except for the initial state
# j.star: the unobserved destination state of the transition from j
    .state[tau]
# t.cen: the last right-censored inter-occurrence time
# x.star: the unobserved uncensored inter-occurrence time between
    transition tau and transition tau+1
# M[]: length nine vector of m's, the frequency of every transition
     in the learning data; this is a vector of ones if there is no
    learning data (see Sections 4.2 and 4.3)
# CC[]: length nine vector with generic entry [(1-q)^{-1/m
    }-1]^{-1}; q and m depend on the transition (see Sections 4.2),
     whereas they take value 0.5 and 1 thoughout if there is no
    learning data (see Section 4.3)
# Q.LAMBDA[]: length nine vector with the quantiles of the nine
    inter-occurrence time distributions (the t_q's in Section 4.2
    and 4.3)

# SCRIPT starts

# Semi Markov Model
model   {
# Likelihood and prior for P
        for (i in 1:k)
```

```
                {
                        transitions[i,1:k] ~ dmulti( P[i,1:k],N.
                            from[i] )
                        P[i,1:k] ~ ddirch( dirichlet.parameters[i
                            ,1:k] )
                }
# Weibull model for inter-occurrence times.
# The relationship between JAGS and our parameterization is: shape
    = alpha, rate = a,  theta = 1/a^(1 / alpha)
        for (l in 1:tau)
                {
                        x.exact[l] ~ dweib(alpha[ctf[l]] , a[ctf[l
                            ]])
                }
# Mixture of Weibulls for the last censored inter-occurrence time
        j.star ~ dcat( P[j.state[tau],1:k] )
        is.censored ~ dinterval( x.star,t.cen )
        x.star ~ dweib( alpha[j.state[tau]+3*(j.star-1)] , a[j.
            state[tau]+3*(j.star-1)] )
#
# Prior on (alpha, a) Weibull parameters when there are m>=2
    transitions for all nine pairs (i,j) in the learning data
# alpha.trick[] is a lenght nine vector which, shifted by LOWER,
    gives alpha[]
# BM[]: length nine vector with the rate parameters of the priors
    of the alpha's  (the factor multiplying alpha in eq. (15))
# The next 6 lines of code should be commented out if there is no
    learning data
#
        alpha <- alpha.trick+LOWER
        theta <- 1 / (a^(1 / alpha))
        for (r in 1:(k^2)){
                alpha.trick[r] ~ dgamma(M[r], BM[r])
                a[r] ~ dgamma(M[r], CC[r]*(Q.LAMBDA[r])^alpha[r])
        }

#
# Prior on (alpha, a) Weibull parameters when there are no learning
     data (see Section 4.3)
# The zeros trick is used for the prior of alpha
# t[]: lenght two vector with the common upper and lower bounds for
     the entries in Q.LAMBDA[]
# LOWER[], UPPER[]: length nine vectors with the endpoints alpha_0
    and alpha_1 of the support of the shape parameters
# The next 8 lines of code should be commented out if learning data
     is available
#
        theta <- 1 / (a^(1 / alpha))
        for (r in 1:(k^2)){
                        a[r] ~ dgamma(M[r], CC[r]*(Q.LAMBDA[r])^
                            alpha[r])
                        Q.LAMBDA[r]~ dunif(t[1], t[2])
                        zero[r] ~ dpois(phi[r])
                        phi[r] <- -log( 1 - LOWER[r] / alpha[r]  )
                        alpha[r] ~ dunif(LOWER[r], UPPER[r])
        }
}
```

Acknowledgments

We are grateful to Renata Rotondi for providing us with data and the map of Italy along with her very helpful comments. We would like to thank the anonymous reviewers and an associate editor for their suggestions and remarks.

References

Altinok, Y. and Kolcak, D. (1999). An application of the semi-Markov model for earthquake occurrences in North Anatolia, Turkey. *Journal of the Balkan Geophysical Society*, Vol. **2**, pp. 90–99.

Alvarez, E. E. (2005). Estimation in stationary Markov Renewal processes, with application to earthquake forecasting in Turkey. *Methodol. Comput. Appl. Probab.*, Vol. **7**, pp. 119–130. MR2153577

Abaimov, S. G., Turcotte, D. L., and Rundle, J. B. (2007). Recurrence-time and frequency-slip statistics of slip events on the creeping section of the San Andreas fault in central California. *Geophys. J. Int.*, Vol. **170**, pp. 1289–1299.

Berger, J. O. and Pericchi, L. R. (1996). The intrinsic Bayes factor for model selection and prediction. *J. Amer. Statist. Assoc.*, Vol. **91**, pp. 109–122. MR1394065

Berger, J. O. and Sun, D. (1993). Bayesian analysis for the Poly-Weibull distribution. *J. Amer. Statist. Assoc.*, Vol. **88**, pp. 1412–1418. MR1245378

Betrò, B., Garavaglia, E., Grandori Guagenti, E., Rotondi, R., and Tagliani, A. (1989). Sulla distribuzione dei tempi di intercorrenza fra eventi sismici in alcune zone italiane (On the distribution of the inter-occurence times between seismic events in some Italian areas). Proceedings del 4 Conv. Naz. "L'Ingegneria Sismica in Italia", ANIDIS – Politecnico di Milano, Milano, Ed. Patron Bologna, Vol. **1**, pp. 135–144.

Bousquet, N. (2006). A Bayesian analysis of industrial lifetime data with Weibull distributions. *Rapport de recherche INRIA n.6025*, pp. 1–21.

Bousquet, N. (2010). Elicitation of Weibull priors. `http://arxiv.org/abs/1007.4740v2`. MR2679337

Cano, J., Moguerza, J. M., and Ríos Insua, D. (2011). Bayesian analysis for semi-Markov processes with applications to reliability and maintenance. *Technical report*, Madrid, Universidad Rey Juan Carlos.

CPTI Working Group (2004). Catalogo Parametrico dei Terremoti Italiani, version 2004 (CPTI04), Ist. Naz. di Geofis. e Vulcanol., Bologna, Italy. `http://emidius.mi.ingv.it/CPTI04/`.

DISS Working Group (2007). Database of Individual Seismogenic Sources (DISS), version 3.0.2: A compilation of potential sources for earthquakes larger than M 5.5 in Italy and surrounding areas. `http://diss.rm.ingv.it/diss/`.

Epifani, I., Fortini, S., and Ladelli, L. (2002). A characterization for mixtures of Semi Markov processes. *Statist. Probab. Lett.*, Vol. **60**, pp. 445–457. MR1947184

Foucher, Y., Mathieu, E., Saint-Pierre, P., Durand, J. F., and Daurès, J. P. (2009). A semi-Markov model based on generalized Weibull distribution with an illustration for HIV disease. *Biom. J.*, Vol. **47**, pp. 825–833. MR2221921

Garavaglia, E. and Pavani, R. (2012). About earthquake forecasting by Markov renewal processes. *Methodol. Comput. Appl. Probab.*, Vol. **13**, pp. 155–169. MR2755137

Gilks, W. R., Best, N. G., and Tan, K. K. C. (1995). Adaptive rejection Metropolis sampling. *Appl. Statist.*, Vol. **44**, pp. 455–472.

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Appl. Statist.*, Vol. **41**, pp. 337–348.

Grandori Guagenti, E. and Molina, C. (1986). Semi-Markov processes in seismic risk analysis. In *Semi-Markov Models: Theory and Applications*, Ed. Janssen, J., pp. 487–503, Plenum Press, New York. MR0874009

Grandori Guagenti, E., Molina, C., and Mulas, G. (1988). Seismic risk analysis with predictable models. *Earthquake Eng. Struct. Dynam.*, Vol. **16**, pp. 343–359.

Hanks, T. C. and Kanamori, H. (1979). A moment magnitude scale. *J. Geophys. Res.*, Vol. **84**, pp. 2348–2350.

Hasumi, T., Akimoto, T., and Aizawa, Y. (2009). The Weibull–log Weibull distribution for interoccurrence times of earthquakes. *Physica A*, Vol. **388**, pp. 491–498.

Hristopulos, D. T. and Mouslopoulou, V. (2013). Strength statistics and the distribution of earthquake interevent times. *Physica A*, Vol. **392**, pp. 485–496.

Limnios, N. and Oprisan, G. (2001). *Semi-Markov Model Processes*. Birkhauser, Boston. MR1843923

Masala, G. (2012). Earthquakes occurrences estimation through a parametric semi-Markov approach. *J. Appl. Statist.*, Vol. **39**, pp. 81–96. MR2872329

Marín, J., Plà, L., and Ríos Insua, D. (2005). Forecasting for some stochastic process models related to sow farm management. *J. Appl. Statist.*, Vol. **32**, pp. 797–812. MR2214310

O'Hagan, A. (1995). Fractional Bayes factors for model comparison. *J. R. Stat. Soc. Ser. B*, Vol. **57**, pp. 99–138. MR1325379

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., and Rakow, T. (2006). *Uncertainty Judgements: Eliciting Experts' Probabilities*. Wiley, Chichester.

Patwardhan, A. S., Kulkarni, R. B., and Tocher, D. (1980). A semi-Markov model for characterizing recurrence of great earthquakes. *Bull. Seismol. Soc. Am.*, Vol. **70**, pp. 323–347.

Peruggia, M. and Santner, T. (1996). Bayesian analysis of time evolution of earthquakes. *J. Amer. Statist. Assoc.*, Vol. **91**, pp. 1209–1218.

Plummer, M. (2010). *JAGS Version 3.1.0: Just another Gibbs sampler*, URL `http://mcmc-jags.sourceforge.net/`.

R Development Core Team (2012). R: A language and environment for statistical computing, Vienna, Austria. ISBN 3-900051-07-0, URL `http://www.R-project.org/`.

Rotondi, R. (2010). Bayesian nonparametric inference for earthquake recurrence time distributions in different tectonic regimes. *J. Geophys. Res.*, Vol. **115**, B01302, DOI: 10.1029/2008JB006272.

Rovida, A., Camassi, R., Gasperini, P., and Stucchi, M. (Eds.) (2011). CPTI11, the 2011 version of the Parametric Catalogue of Italian Earthquakes. Milano, Bologna, Italy. `http://emidius.mi.ingv.it/CPTI`.

Sadeghian, R. (2012). Forecasting time and place of earthquakes using a semi-Markov model (with case study in Tehran province). *J. Ind. Eng. Int.*, Vol. **8**, p. 20.

Stacy, E. W. (1962). A generalization of the gamma distribution. *Ann. Math. Statist.*, Vol. **33**, pp. 847–1226. MR0143277

Shimazaki, K. and Nakata, T. (1980). Time-predictable recurrence model for large earthquakes. *Geophys. Res. Lett.*, Vol. **7**, pp. 279–282.

Votsi, I., Limnios, N., Tsaklidis, G., and Papadimitriou, E. (2012). Estimation of the expected number of earthquake occurrences based on semi-Markov models. *Methodol. Comput. Appl. Probab.*, Vol. **14**, pp. 685–703. MR2966315

Yu, Q., MacEachern, S. N., and Peruggia, M. (2011). Bayesian synthesis: Combining subjective analyses, with an application to ozone data. *Ann.Appl. Stat.*, Vol. **5**, pp. 1678–1698. MR2849791