



Estimation of the Expected Number of Earthquake Occurrences Based on Semi-Markov Models

Irene Votsi, Nikolaos Limnios, George Tsaklidis, Eleftheria Papadimitriou

► To cite this version:

Irene Votsi, Nikolaos Limnios, George Tsaklidis, Eleftheria Papadimitriou. Estimation of the Expected Number of Earthquake Occurrences Based on Semi-Markov Models. *Methodology and Computing in Applied Probability*, 2012, 14 (3), pp.685 - 703. 10.1007/s11009-011-9257-4 . hal-01635178

HAL Id: hal-01635178

<https://hal.science/hal-01635178>

Submitted on 7 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Estimation of the expected number of earthquake occurrences based on semi-Markov models

I. VOTSI*, N. LIMNIOS†, G. TSAKLIDIS‡, E. PAPADIMITRIOU§

Abstract

The present paper aims at the introduction of the semi-Markov model in continuous time as a candidate model for the description of seismicity patterns in time domain in the Northern Aegean Sea (Greece). Estimators of the semi-Markov kernels, Markov renewal functions and transition functions are calculated through a nonparametric method. Moreover, the hitting times for spatial occurrence of the strongest earthquakes as well as the confidence intervals of certain important indicators are estimated. Firstly, the classification of model states is based on earthquakes magnitude. The instantaneous earthquake occurrence rate between the states of the model as well as the total earthquake occurrence rate are calculated. In order to increase the consistency between the model and the process of earthquake generation, seismotectonic features have been incorporated as an important component in the model. Therefore, a new classification of states is proposed which combines both magnitude and fault orientation states. This model which takes into account seismotectonic features contributes significantly to the seismic hazard assessment in the region under study. The model is applied to earthquake catalogues for the Northern Aegean Sea, an area that accommodates high seismicity, being a key structure from the seismotectonic point of view.

Keywords. Semi-Markov process, semi-Markov kernels, earthquake occurrence rates, expected number of earthquake occurrences, hitting times.

1 Introduction

Several analytical models have been proposed to represent the process of earthquake occurrence. Some of them are based on empirical observations of precursory phenomena, others on physical modelling of the earthquake process and a third class on statistical analysis of patterns of seismicity. The most common statistical model is the Poisson model, which assumes temporal independence of earthquakes. The simplest time-dependent model is the non-homogeneous Poisson model, which is not appropriate for seismic hazard assessment in long time intervals, because the hazard should be updated when a new earthquake occurs.

A semi-Markov model that considers the non-random character of earthquake magnitude and recurrence time was proposed by Cluff et al. (1980). In this model a parametric method

*Postal address: Department of Mathematics, Aristotle University of Thessaloniki, 54124, Thessaloniki, Greece and Laboratoire de Mathématiques Appliquées, Université de Technologie de Compiègne, 60205, Compiègne, France.

†Postal address: Laboratoire de Mathématiques Appliquées, Université de Technologie de Compiègne, 60205, Compiègne, France.

‡Postal address: Department of Mathematics, Aristotle University of Thessaloniki, 54124, Thessaloniki, Greece.

§Postal address: Department of Geophysics, Aristotle University of Thessaloniki, 54124, Thessaloniki, Greece. Email address: evotsi@math.auth.gr, nikolaos.limnios@utc.fr, tsaklidi@math.auth.gr, ritsa@geo.auth.gr.

was applied for recurrence intervals estimation. More recent studies include the application of Weibull semi-Markov process through a new parametric estimation method (Alvarez, 2005). A similar application based on a mixture of exponential and Weibull distribution for sojourn times was later presented by Garavaglia and Pavani (2009).

Parametric methods provide estimators with several attractive asymptotic properties; however, these estimators present inconvenience when the sample size is small. Frequently, the robustness of the forecasting results obtained via parametric methods is limited due to the absence of sufficient data. Since applications of parametric methods presuppose certain conditions with respect to the sample size, this difficulty could be overcome through the application of nonparametric methods. Altinok and Kolcak (1999) adopted a nonparametric method to a semi-Markov chain applied to an earthquake catalogue for the North Anatolian region of Turkey in order to estimate the transition probabilities between magnitude or region states.

The problem of statistical inference for semi-Markov processes is of increasing interest in literature. There is a growing literature concerning inference problems for continuous-time semi-Markov processes. For instance, Moore and Pyke (1968) studied empirical and maximum likelihood estimators for semi-Markov kernels; Lagakos et al. (1978) obtained the nonparametric maximum likelihood estimator for the kernel of a finite state semi-Markov process with some absorbing states; Akritas and Roussas (1979) studied the asymptotic local normality; Gill (1980) constructed an estimator for the kernel of a finite state semi-Markov kernel, using counting processes; Ouhbi and Limnios (1999) studied empirical estimators for non-linear functionals of finite semi-Markov kernels.

In the present paper we propose a semi-Markov model in continuous time for seismic hazard assessment in the Northern Aegean Sea. The semi-Markov kernel, the Markov renewal functions, the transition probabilities and the distributions of sojourn times for every state are estimated through a nonparametric method. Moreover, the confidence intervals of important indicators are calculated. Firstly, the classification of states is derived for different ranges of magnitudes; however, observations highlight the need for a seismic hazard analysis model that includes seismotectonic features. Therefore the states are classified via a combination of magnitudes and faulting properties (Rhoades et al., 2010). This division is introduced in order to improve the forecasting of the semi-Markov model. Stochastic models allow us to estimate the probability of occurrence of a given outcome, namely to provide forecasts. In this study we explicitly deal with forecasts, not predictions.

The organization of the paper is as follows: In Section 2, notation and preliminaries of the theory of semi-Markov processes are presented. Section 3 includes the development of the semi-Markov model for the description of the earthquake generation mechanism in the area described in Section 3.1. Section 3.2 includes the estimation of the semi-Markov kernels, conditional transition functions as well as the estimation of transition probabilities. In Section 3.3 the stationary distribution of the semi-Markov process is described and estimated nonparametrically. Section 3.4 specializes the results for the expected number of earthquake occurrences, transition functions and hitting times for the occurrence of the strongest earthquakes. The estimated earthquake recurrence rates are exhibited and estimated in Section 3.5. In Section 4 a new classification of states of the semi-Markov model is suggested which further includes the spatial component. Moreover the estimated hitting times for each subarea is presented. Finally, in Section 5, we give some concluding remarks.

2 Notation and Preliminaries of Semi-Markov Processes

Firstly, we briefly recall the main definitions from the theory of semi-Markov processes which are directly useful for our study (see, e.g., Limnios and Oprisan, 2001). Let us consider a Markov Renewal Process (MRP), $(J, S) = (J_n, S_n)_{n \geq 0}$, defined on a complete probability space, where $(J_n)_{n \geq 0}$ is a Markov chain with values in the state space of the process, $E = \{1, 2, \dots, s\}$, and $(S_n)_{n \geq 0}$ are the jump times which take values in $\mathbb{R}^+ = [0, \infty)$. $J_0, J_1, \dots, J_n, \dots$ are the consecutive states to be visited by the MRP and $X_0 = S_0 = 0$, X_1, X_2, \dots defined by $X_n = S_n - S_{n-1}$ for $n \geq 1$, are the sojourn times in these states. Define also the Semi-Markov Process (SMP) $(Z_t)_{t \in \mathbb{R}^+}$ by $Z_t = J_{N(t)}$, $t \geq 0$, where $N(t)$ is the counting process of the SMP up to time t . The stochastic behavior of the SMP is determined completely by its initial law, $P(J_0 = k) = a(k)$, and its semi-Markov kernel,

$$Q_{ij}(x) = P(J_{n+1} = j, X_{n+1} \leq x | J_0, J_1, \dots, J_n = i, X_1, X_2, \dots, X_n), \quad (1)$$

for all $x \in \mathbb{R}^+$ and $i, j \in E$. The probabilities $p_{ij} = \lim_{t \rightarrow \infty} Q_{ij}(t) = Q_{ij}(\infty)$ are the transition probabilities of the Embedded Markov Chain (EMC), $(J_n)_{n \geq 0}$. It is worth noticing that in the present study we consider $Q_{ii}(t) \neq 0$, for all $i \in E$.

Let us now consider the distribution function associated with the sojourn time in state i before going to state j , $F_{ij}(x) = P(X_{n+1} \leq x | J_n = i, J_{n+1} = j)$, and the sojourn time distribution in state i , $H_i(x) = P(X_{n+1} \leq x | J_n = i)$. For a fixed time T ($T \geq 0$), let $N_i(T)$ be the number of visits of $(J_n)_{n \geq 0}$ to state $i \in E$ up to time T , and let $N_{ij}(T)$ be the number of transitions from state i to state j up to time T , that is

$$N_i(T) := \sum_{n=1}^{N(T)} \mathbf{1}_{\{J_n=i\}} = \sum_{n=1}^{\infty} \mathbf{1}_{\{J_n=i, S_n \leq T\}} \quad (2)$$

and

$$N_{ij}(T) := \sum_{n=1}^{N(T)} \mathbf{1}_{\{J_{n-1}=i, J_n=j\}} = \sum_{n=1}^{\infty} \mathbf{1}_{\{J_{n-1}=i, J_n=j, S_n \leq T\}}, \quad (3)$$

where $\mathbf{1}$ stands for the indicator function. The observation of a sample path of a SMP in the time interval $[0, T]$ is described as $H_T = \{J_0, J_1, \dots, J_{N(T)}, X_1, X_2, \dots, X_{N(T)}\}$.

Here the aim is to estimate the considerable quantities of a finite state space SMP by observing a sample path in the time interval under consideration, in order to contribute to the seismic hazard assessment in Northern Aegean Sea. We aim to treat the SMP from a nonparametric perspective, by proposing and calculating empirical estimators of the most important indicators in the respective theory.

In the sequel, the empirical estimators of the aforementioned functions are presented. Let us define the following empirical estimator of the semi-Markov kernel (Moore and Pyke, 1968; Ouhbi and Limnios, 1999):

$$\hat{Q}_{ij}(x, T) := \frac{1}{N_i(T)} \sum_{n=1}^{N(T)} \mathbf{1}_{\{J_{n-1}=i, J_n=j, X_n \leq x\}}. \quad (4)$$

The empirical estimator of the semi-Markov kernel is strongly consistent and asymptotically normal. From this definition we obtain $\hat{Q}_{ij}(x, T) = \hat{p}_{ij}(T) \cdot \hat{F}_{ij}(x, T)$ where

$$\hat{p}_{ij}(T) := \frac{N_{ij}(T)}{N_i(T)} \quad \text{and} \quad \hat{F}_{ij}(x, T) := \frac{1}{N_{ij}(T)} \sum_{n=1}^{N(T)} \mathbf{1}_{\{J_{n-1}=i, J_n=j, X_n \leq x\}}, \quad (5)$$

are the empirical estimators of the transition probabilities and the conditional transition functions, respectively. The conditional transition mechanism describes the probability function of the process moving into each possible new state, given the old state and the new one.

At this point, we proceed to the definition of the convolution operation, a crucial operation for the aims of the paper. If we denote by $\phi(i, t)$, $i \in E$, $t \geq 0$, a real valued measurable function, the Stieltjes convolution of ϕ by the semi-Markov kernel Q , is defined as

$$Q * \phi(i, t) = \sum_{k \in E} \int_0^t Q_{ik}(ds) \phi(k, t - s). \quad (6)$$

Moreover, the quantity $P_i(J_n = j, S_n \leq t)$, denoted by the n -fold convolution of Q_{ij} by itself in the Stieltjes convolution sense, $Q_{ij}^{(n)}(t)$, is defined by the recursive formula

$$Q_{ij}^{(n)}(t) = \begin{cases} \sum_k \int_0^t Q_{ik}(ds) Q_{kj}^{(n-1)}(t - s) & \text{if } n \geq 2 \\ Q_{ij}(t) & \text{if } n = 1 \\ \delta_{ij} \mathbf{1}_{\{t \geq 0\}} & \text{if } n = 0 \end{cases},$$

where δ_{ij} is the Kronecker's delta symbol and $\mathbf{1}_{\{t \geq 0\}}$ the indicator function

$$\mathbf{1}_{\{t \geq 0\}} = \begin{cases} 1 & \text{if } t \geq 0 \\ 0 & \text{elsewhere} \end{cases}.$$

For reader's convenience a list of symbols is given (p. 16-17).

3 Modelling Earthquakes in Northern Aegean Sea

In the present paper, we model the earthquake occurrence mechanism by a continuous-time, finite state space semi-Markov process; that is, the process Z_t evolves by jumping between the states, where it obeys a certain probabilistic distribution of time. The semi-Markov model is fitted nonparametrically to earthquakes observed in Northern Aegean Sea of Greece.

3.1 Data Selection

The region of the Northern Aegean Sea (Greece) and its surrounding area has attracted the interest of many researchers because it accommodates high seismic activity, being a key structure from the seismotectonic point of view. The study area has experienced several destructive earthquakes ($M \geq 6.5$) as is indicated by both instrumental data and historical information. It constitutes the northern boundary of the south Aegean plate (Papazachos et al., 1998) and is a continuation of the western part of the North Anatolian fault. Strike-slip dextral faulting dominates this region as the North Anatolian fault prolongates into the north Aegean area, where it bifurcates into two main branches of NE-SW trend. Conjugate NW-SE striking sinistral strike slip faults, running parallel to the coastline of the continental Greece, also accommodate strong earthquakes. Parallel secondary faults are also recognized from seismicity and fault-plane solutions of recent strong earthquakes.

The Northern Aegean Sea region is selected for this investigation because it has an adequate number of strong ($M \geq 6.4$) earthquakes, along with an adequate number of moderate ($M \geq 5.5$) events since 1953. In this study, we consider a wide region of the Northern Aegean Sea, limited by the rectangle of coordinates 23.5° - 26.5° E and 38.3° - 40.5° N. All earthquakes with $M \geq 5.5$ in the Aristotle University of Thessaloniki (AUTH) catalogue since 1953 are included (Table 6). The catalogue is characterized by accuracy, homogeneity and completeness.

3.2 Classification of States

The data set comprises events of magnitude $M \geq 5.5$ that have occurred during the period [1953, 2010]. As the model refers to main events only, aftershocks were carefully identified and removed from the data before application begins, by means of Reasenbergs's (1985) declustering algorithm. Concerning states classification, according to previous studies, the states of the SMP can be considered to be either the magnitudes or energy release levels of earthquakes. The continuous magnitude scale is divided into appropriate intervals to specify discrete states of the system. Initially, we define three states corresponding to magnitudes: *State 1*: [5.5, 5.6], *State 2*: [5.7, 6.0] and *State 3*: [6.1, 7.2], namely the state space is $E = \{1, 2, 3\}$. The model uses only earthquake data, with no explicit use of geologic, tectonic or geodetic information.

In the sequel, the semi-Markov model in continuous time is applied to the above mentioned earthquake catalogue and the number of observed transitions in the dataset as well as the the empirical estimators of transition probabilities from each state i to each state j ($i, j \in E$), are presented as elements of the matrices T and $\hat{P} = (\hat{p}_{ij})$, respectively

$$T = \begin{pmatrix} 6 & 6 & 3 \\ 5 & 2 & 2 \\ 4 & 1 & 3 \end{pmatrix} \text{ and } \hat{P} = \begin{pmatrix} 0.3750 & 0.4375 & 0.1875 \\ 0.5556 & 0.2222 & 0.2222 \\ 0.5000 & 0.1250 & 0.3750 \end{pmatrix}.$$

The empirical estimators of semi-Markov kernels for transitions from state i to state j ($i, j \in E$), $\hat{Q}_{ij}(x, T)$, are exhibited in Figure 1.

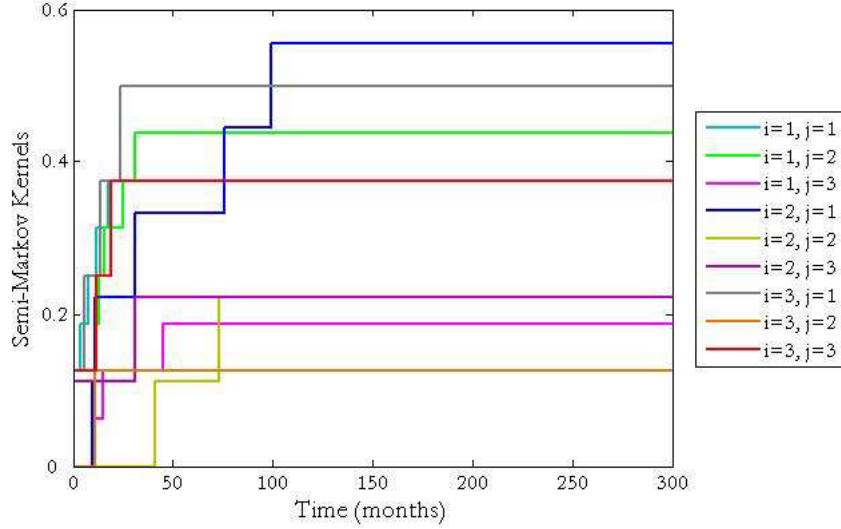


Figure 1: Empirical Estimators of Semi-Markov Kernels, $\hat{Q}_{ij}(x, T)(i, j \in E)$.

The estimated distribution functions associated with the sojourn time in each state $i \in E$ before going to each state $j \in E$, $\hat{F}_{ij}(x, T)$, are presented in Figure 2.

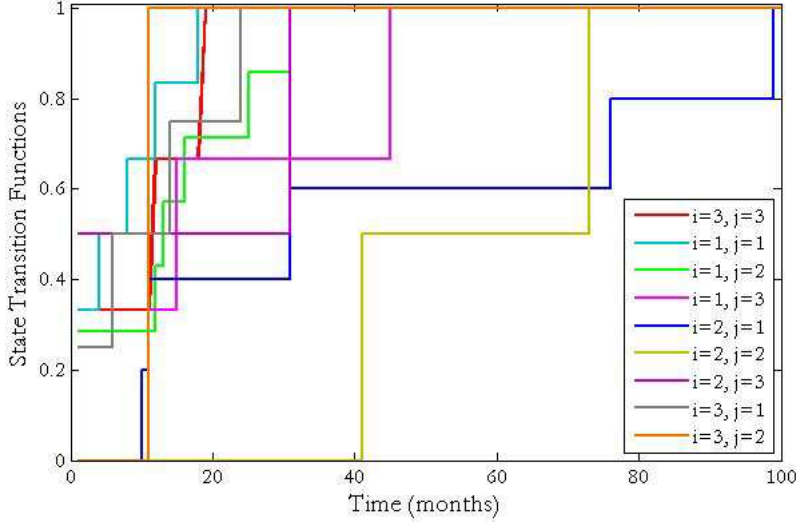


Figure 2: Empirical Estimators of Conditional Transition Functions, $\hat{F}_{ij}(x, T)$ ($i, j \in E$).

3.3 Stationary Distribution of the SMP

If the EMC $(J_n)_{n \geq 0}$ is irreducible with invariant distribution ν_i and the mean sojourn time in state i , m_i , is finite for all $i \in E$, the empirical estimator of the stationary distribution of the SMP is expressed in terms of the stationary distribution of the EMC.

The following estimator for the stationary distribution of the SMP is determined by the formula (Limnios et al., 2005)

$$\hat{\pi}_i(T) := \frac{\hat{\nu}_i(T) \hat{m}_i(T)}{\sum_{k=1}^s \hat{\nu}_k(T) \hat{m}_k(T)}, \quad (7)$$

where $\hat{\nu}_i(T) = \frac{N_i(T)}{N(T)}$, for $i \in E$, is the empirical estimator of the stationary distribution of the EMC and $\hat{m}_i(T) = \int_0^\infty (1 - \hat{H}_i(t, T)) dt$ is the estimated mean sojourn time in state i .

Let us now denote by $(S_n^i)_{n \geq 0}$ the renewal function of successive times of visits to state i and by μ_{ii} the mean recurrence time of (S_n^i) , namely $\mu_{ii} = E[S_2^i - S_1^i]$, $i \in E$. The empirical estimator of the mean recurrence time is

$$\hat{\mu}_{ii} = \frac{1}{\hat{\nu}_i(T)} \sum_{p \in E} \hat{\nu}_p(T) \hat{m}_p(T). \quad (8)$$

In Table 1 the empirical estimators of the quantities given in equations (7), (8) as well as the estimated mean recurrence times are exhibited.

Table 1: Estimated Stationary Distribution of the SMP and Estimated Mean Recurrence Times.

State i	$\hat{\nu}_i(T)$	$\hat{m}_i(T)$	$\hat{\pi}_i(T)$	$\hat{\mu}_{ii}(T)$
1	0.4848	18.7208	0.4620	40.5292
2	0.2727	24.1481	0.3351	72.0519
3	0.2425	16.4500	0.2029	81.0583

3.4 Hitting Times for Earthquake Occurrences, Expected Number of Earthquake Occurrences and Transition Functions

Using the preceding notation from the theory of SMPs, we estimate indicators of great importance in terms of the empirical estimators of semi-Markov kernels. It is worth estimating the expected number of earthquake occurrences from one state to another up to any time. The Markov renewal matrix $\Psi(t) = (\psi_{ij}(t))$, where $\psi_{ij}(t) = E_i[N_j(t)] = \sum_{n=0}^{\infty} P_i(J_n = j, S_n \leq t) = \sum_{n=0}^{\infty} Q_{ij}^{(n)}(t)$, $(i, j) \in E^2, t \geq 0$ provides the relevant knowledge. The estimator of the (i, j) th element of the Markov renewal matrix, $\hat{\psi}_{ij}(x, T)$, has the following form

$$\hat{\psi}_{ij}(x, T) = \sum_{n=0}^{\infty} \hat{Q}_{ij}^{(n)}(x, T).$$

The previous results allow us to construct confidence intervals for the matrix-valued function ψ . At this point, the expected number of earthquake occurrences from each state $i \in E$ to the third state that includes the stronger earthquakes in which we are more interested, as well as the corresponding 95% confidence interval is estimated (Fig. 3). The estimated quantities and their confidence intervals do not differ significantly. The width of the confidence intervals indicates that for every $i \in E$ the quantity $\psi_{i3}(x, T)$ can be predicted accurately throughout the entire data range (with regard to the x-axes).

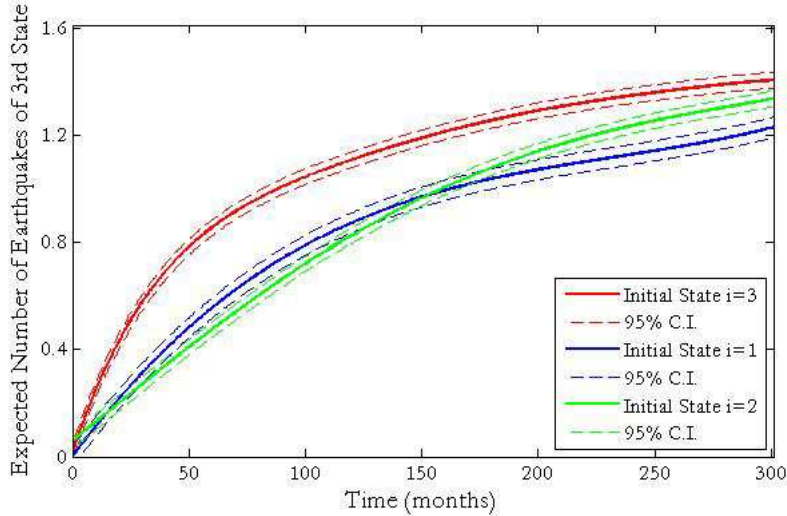


Figure 3: The 95% confidence interval of the expected number of earthquake occurrences into state 3, given that the initial state is state $i \in E$, $\hat{\psi}_{i3}(x, T)$.

Here, it is acknowledged that the estimation of the transition functions is an outstanding issue for the determination of the semi-Markov model. We proceed to the calculation of the transition functions $P_{ij}(t) = P(Z_t = j | Z_0 = i)$ through the empirical estimator (in matrix form) (Limnios, 1997):

$$\hat{P}(x, T) = (I - \hat{Q}(x, T))^{(-1)} * (I - \text{diag}(\hat{Q}(x, T) \cdot \mathbf{1})), \quad (9)$$

where I stands for the identity matrix and $(I - \widehat{Q}(x, T))^{(-1)}$ stands for the inverse of a matrix $I - \widehat{Q}(x, T)$ in the Stieltjes convolution sense (Limnios and Oprisan, 2001).

For $t \in \mathbb{R}^+$, $A(t) = \begin{pmatrix} 1 & C(t) \\ D(t) & 1 \end{pmatrix}$ we have $(A(t))^{(-1)} = (1 - C * D(t))^{(-1)} * \begin{pmatrix} 1 & -C(t) \\ -D(t) & 1 \end{pmatrix}$,

where $C(t)$, $D(t)$ are sub-distribution functions and $(1 - C * D(t))^{(-1)} = \sum_{n=0}^{\infty} (C * D)^{(n)}(t)$ is a standard renewal function. Furthermore, $diag(\cdot)$ denotes a diagonal matrix with its i th entry equal to $\sum_{j=1}^s Q_{ij}(t)$, and $\mathbf{1}$ denotes the column-vector of ones. The transition function $P_{ij}(t)$ denotes the probability that the SMP will visit state j at time t given that the initial state of the SMP is state i . Both estimators of the transition functions and the expected number of earthquake occurrences are strongly consistent and asymptotically normal (Ouhbi and Limnios, 1999).

The distribution of the hitting time for an earthquake occurrence of the third state is expressed by the formula $W(t) = 1 - \sum_i a(i) \cdot W_i(t)$, with

$$W_i(t) = \sum_{j=1}^2 (\psi_{ij} * \delta_{ij} \overline{H}_i)(t), \quad (10)$$

where $i \in \{1, 2\}$ and δ_{ij} is the Kronecker's delta symbol.

The following matrix plug of the hitting time's estimator is proposed:

$$\widehat{W}(x, T) = 1 - \widehat{a}_1 (I - \widehat{Q}_{11}(x, T))^{(-1)} * \widehat{H}_1(x, T), \quad (11)$$

where $\overline{H}_1(x, T) = (\overline{H}_i(x, T); i \in \{1, 2\})'$ and $\overline{H}_2(x, T) = (\overline{H}_3(x, T))$.

Furthermore, index 1 represents the restriction of the corresponding vector or matrix to the set of states $\{1, 2\}$ and index 2 represents the restriction to the third state. Figure 4 shows the estimated transition functions $\widehat{P}_{ij}(x, T)$.

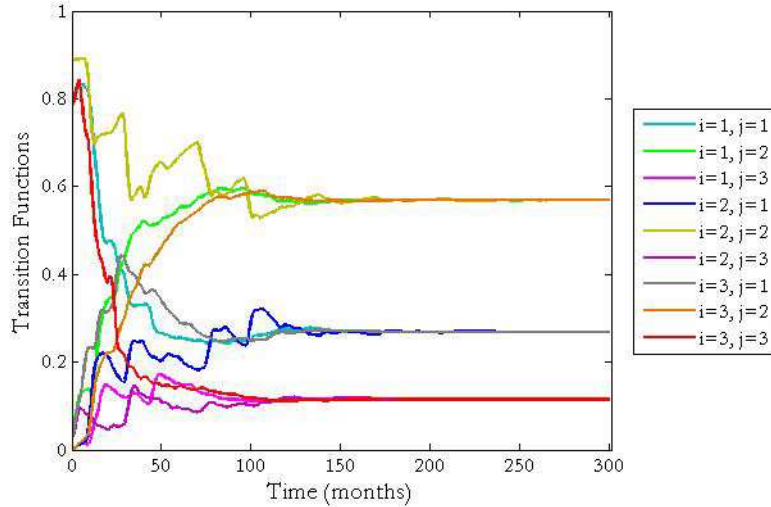


Figure 4: Estimated Transition Functions from State i to State j , $\widehat{P}_{ij}(x, T)$ ($i, j \in E$).

The estimated hitting time of an earthquake occurrence with magnitude $M \geq 6.1$ is presented in Figure 5.

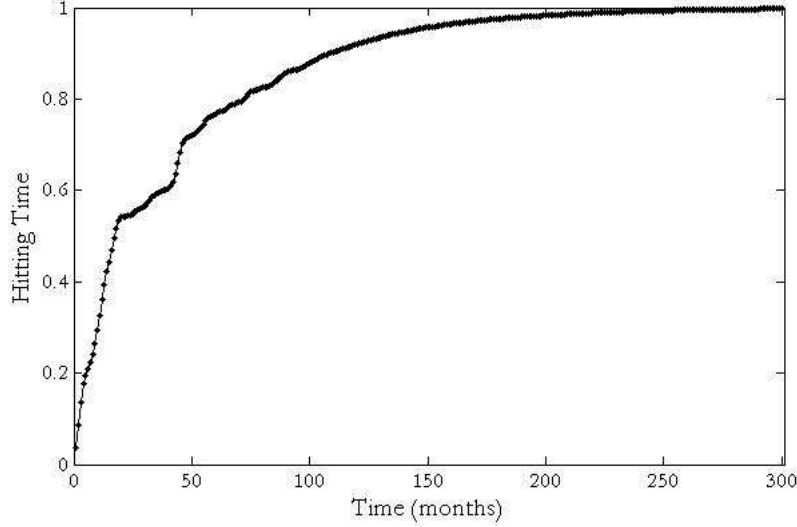


Figure 5: Estimated Distribution of Hitting Time of State 3 ($M \geq 6.1$), $\widehat{W}(x, T)$.

3.5 Estimating Earthquake Occurrence Rates

In the last several decades the application of stochastic models theory to earthquake forecast has been the focus of considerable research activity. Forecasting results can be feasible to be obtained through the calculation of earthquake occurrence rates.

Given that the last earthquake occurrence was in state i and at least a time interval of length t has already elapsed, the probability of an earthquake occurrence of state j in the next time interval of length Δ is denoted by $\lambda_{ij}(t)\Delta$. The term *Instantaneous Earthquake Occurrence Rate at state j in the next step conditional on the starting state i* is used for the description of the probability $\lambda_{ij}(t)\Delta$, which is expressed by means of the semi-Markov kernels via the formula

$$\lambda_{ij}(t) = \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \cdot \frac{Q_{ij}(t + \Delta) - Q_{ij}(t)}{\overline{H}_i(t)} \quad (12)$$

and

$$\lambda_{ij}(t) \cdot \Delta = \frac{Q_{ij}(t + \Delta) - Q_{ij}(t)}{\overline{H}_i(t)} + o(\Delta). \quad (13)$$

Table 2 shows the estimated instantaneous earthquake occurrence rate for each type of transitions. It provides forecasting results for an earthquake occurrence in the next time interval of length Δ , knowing that the last earthquake occurred before at least one semester and assuming different values for Δ ($\Delta = 1/2, 1, 2, 3, 4$ years).

Under the condition that the last event was in state $i = 1$ and one, two or three years have elapsed with no earthquake occurrences, the probability that an earthquake will occur into

Table 2: Estimated Instantaneous Earthquake Occurrence Rates.

Δ	$t = 1 \text{ semester}$								
	$\lambda_{11}(t)\Delta$	$\lambda_{12}\Delta$	$\lambda_{13}\Delta$	$\lambda_{21}\Delta$	$\lambda_{22}\Delta$	$\lambda_{23}\Delta$	$\lambda_{31}\Delta$	$\lambda_{32}\Delta$	$\lambda_{33}\Delta$
1 semester	0.182	0.091	0.909	0.250	0	0	0	0.2	0.2
1 year	0.273	0.273	0.182	0.250	0	0	0.2	0.2	0.2
2 years	0.273	0.364	0.182	0.250	0	0	0.4	0.2	0.4
3 years	0.273	0.456	0.182	0.375	0.125	0.125	0.4	0.2	0.4
4 years	0.273	0.456	0.271	0.375	0.125	0.125	0.4	0.2	0.4

the ensuing semester or years is exhibited in Table 3. Conditioning on the events $\{J_n = 2\}$, $\{J_n = 3\}$ the estimated instantaneous earthquake occurrence rates are reported in Tables 4, 5, respectively.

Table 3: Estimated Instantaneous Earthquake Occurrence Rates – Starting State 1.

Δ	$t = 12 \text{ months}$			$t = 24 \text{ months}$			$t = 36 \text{ months}$		
	$\lambda_{11}(t)\Delta$	$\lambda_{12}\Delta$	$\lambda_{13}\Delta$	$\lambda_{11}\Delta$	$\lambda_{12}\Delta$	$\lambda_{13}\Delta$	$\lambda_{11}\Delta$	$\lambda_{12}\Delta$	$\lambda_{13}\Delta$
6 months	0.143	0.286	0.143	0	0.333	0	0	0	1
12 months	0.143	0.286	0.143	0	0.667	0	0	0	1
24 months	0.143	0.572	0.143	0	0.667	0.333	0	0	1
36 months	0.143	0.572	0.285	0	0.667	0.333	0	0	1
48 months	0.143	0.572	0.285	0	0.667	0.333	0	0	1

Table 4: Estimated Instantaneous Earthquake Occurrence Rates – Starting State 2.

Δ	$t = 12 \text{ months}$			$t = 24 \text{ months}$			$t = 36 \text{ months}$		
	$\lambda_{21}(t)\Delta$	$\lambda_{22}\Delta$	$\lambda_{23}\Delta$	$\lambda_{21}\Delta$	$\lambda_{22}\Delta$	$\lambda_{23}\Delta$	$\lambda_{21}\Delta$	$\lambda_{22}\Delta$	$\lambda_{23}\Delta$
6 months	0	0	0	0	0	0	0	0.250	0
12 months	0	0	0	0.167	0	0.167	0	0.250	0
24 months	0.167	0	0.167	0.167	0.167	0.167	0	0.250	0
36 months	0.167	0.167	0.167	0.167	0.167	0.167	0	0.250	0
48 months	0.167	0.167	0.167	0.167	0.167	0.167	0.250	0.500	0

Table 5: Estimated Instantaneous Earthquake Occurrence Rates – Starting State 3.

Δ	$t = 12 \text{ months}$			$t = 24 \text{ months}$			$t = 36 \text{ months}$		
	$\lambda_{31}(t)\Delta$	$\lambda_{32}\Delta$	$\lambda_{33}\Delta$	$\lambda_{31}\Delta$	$\lambda_{32}\Delta$	$\lambda_{33}\Delta$	$\lambda_{31}\Delta$	$\lambda_{32}\Delta$	$\lambda_{33}\Delta$
6 months	0.667	0	0	0	0	0	0	0	0
12 months	0.667	0	0.333	0	0	0	0	0	0
24 months	0.667	0	0.333	0	0	0	0	0	0
36 months	0.667	0	0.333	0	0	0	0	0	0
48 months	0.667	0	0.333	0	0	0	0	0	0

Tables 3, 4 and 5 lead to the next prospective conclusion: assuming that the last event was in state $i \in E$ and at least a constant time interval of length t has elapsed ($t = 12, 24, 36$ months), the probability of an earthquake occurrence of a given state $j \in E$ increases as time elapses, namely as Δ increases ($\Delta = 6, 12, 24, 36, 48$ months).

Focused on the strongest earthquakes of the catalogue which are classified to the third state of the semi-Markov model, we proceed to the estimation of the *Total Earthquake Occurrence Rate*. The total earthquake occurrence rate is defined as

$$\lambda_a(t) = \lim_{h \rightarrow 0} \frac{1}{h} P_a(Z_{t+h} = 3 | Z_u \in \{1, 2\}; \forall u \leq t). \quad (14)$$

The probability that in the next infinitesimal time interval of length h an earthquake with magnitude $M \geq 6.1$ will occur, given that in the time interval $(0, t)$ only earthquakes of states 1 and 2 have occurred, has the form

$$\lambda_a(t) \cdot h + o(h) = P_a(Z_{t+h} = 3 | Z_u \in \{1, 2\}; \forall u \leq t). \quad (15)$$

The empirical estimator of the total earthquake occurrence rate is given by the formula (Limnios and Oprisan, 2001)

$$\hat{\lambda}_a(t, T) = \frac{a_1 \cdot \hat{\psi}_{11} * \hat{H}'_1(t, T) \cdot \mathbf{1}}{a_1 \cdot \hat{\psi}_{11} * \hat{H}_1(t, T) \cdot \mathbf{1}} \quad (16)$$

and is presented in Figure 6.

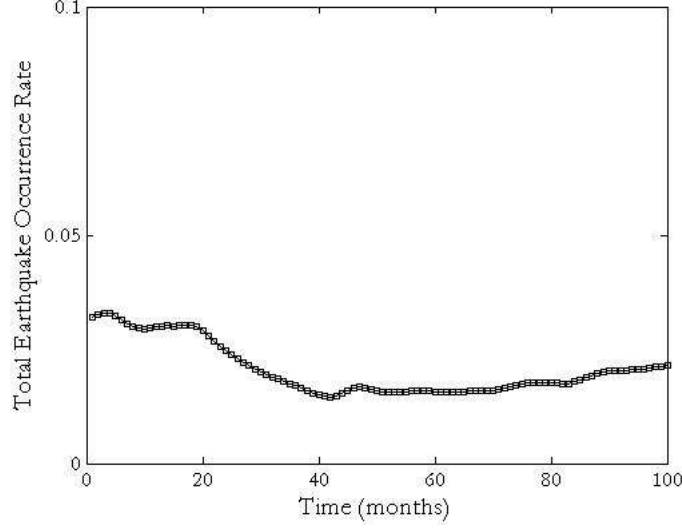


Figure 6: Estimated Total Earthquake Occurrence Rate, $\hat{\lambda}_a(t, T)$.

It can be observed by Fig. 6 that the estimated total earthquake occurrence rate takes values smaller than 0.05 for each $t \in [0, 100]$; this conclusion holds even if $t > 100$. In the sequel we calculate the probability that an earthquake with magnitude $M \geq 6.1$ (3rd state), which is not necessarily the first, will occur in the next time interval. The term *Rate of Occurrence of Earthquakes* with $M \geq 6.1$ is used for the description of the probability, which is denoted by $ro(t)$. In order to calculate the function $ro(t)$ a few assumptions are to be made. If we assume that the semi-Markov kernel $Q(t)$ is absolutely continuous with respect

to Lebesgue measure on \mathbb{R}^+ , with derivative $q(t) = (q_{ij}(t))_{\{i,j\} \in E}$ and

$$\psi'(t) = \sum_{n=0}^{\infty} [Q^{(n)}]'(t) < \infty$$

for any fixed $t \in \mathbb{R}^+$, then the function $ro(t)$ takes the form (Ouhbi and Limnios, 2002)

$$ro(t) = \sum_{i \in \{1,2\}} \sum_{l=1}^s a_l \int_0^t \psi_{li}(du) \cdot q_{i3}(t-u), \quad (17)$$

with corresponding empirical estimator

$$\widehat{ro}(t, T) = \sum_{i \in \{1,2\}} \sum_{l=1}^s a_l \cdot \widehat{\psi}_{li} * \widehat{q}_{i3}(t-u). \quad (18)$$

Under the aforementioned assumptions the estimator (18) is uniformly strongly consistent and asymptotically normal for the function $ro(t)$ (Ouhbi and Limnios, 2002). Figure 7 exhibits the empirical estimator of the function $ro(t)$ and its 95% confidence interval.

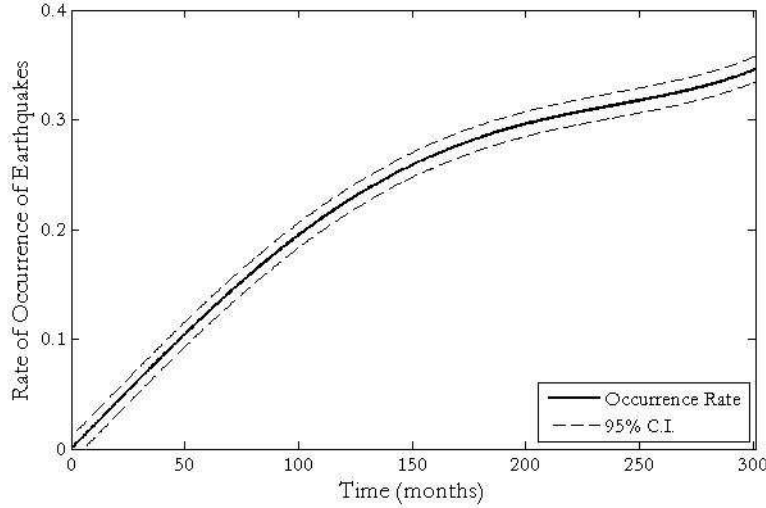


Figure 7: Rate of Occurrence of 3rd State's Earthquakes ($M \geq 6.1$), $\widehat{ro}(t, T)$.

4 New Classification of States

The aforementioned classification of states was uniquely determined by earthquake magnitudes. In order to achieve more reliable predictions we should also incorporate reliable seismotectonic information. For this reason a new classification of states is proposed combining both magnitude and fault orientation states. As the incorporation of the triggering feature into the model is desirable, we used all the data of magnitude $M \geq 5.2$ since 1964 (Table 7). From this time, information for focal mechanisms significantly contributes to the seismic risk mitigation.

When searching for a potential correlation between static stress changes and seismicity changes, one approach is to calculate these changes for the nodal planes of the subset of shocks with known focal mechanisms (Stein, 1999). Since the stress field depends on the fault orientation, it is necessary to calculate the stress field for a representative set of fault orientation classes, which cover all the earthquakes in the catalogue. Therefore, the study area was divided in smaller subareas on the basis of faulting similarity following Rhoades and his colleagues (2010). They considered the distribution of strike angles, dip angles and rake angles in the available focal mechanisms, and then divided the strike angles into 5 groups, the dip angles into 3 groups, and the rake angles into five groups. All the known focal mechanisms were found to be contained in only 15 of the 75 resulting possible classes for combinations of strike angle, dip angle and rake angle groups. For each fault orientation class, the faulting type is represented by average values of the strike, rake and dip angles. In an effort to balance between faulting details and adequate data sample in each subarea, we merged these 15 spatial clusters into four subareas, which are shown as polygons in Figure 8.

The states of the model are derived from a combination of their magnitude states (*State* 1: [5.2, 5.5], *State* 2: [5.6, 7.2]) with the corresponding subareas (Fig. 8). Through the incorporation of the seismotectonic criteria into the classification of states, the hitting time for an earthquake occurrence with $M \geq 5.6$ into each one of the subareas is presented in Figure 9. We are concerned for the events with $M \geq 5.6$ as these events cause disastrous damages and even human losses, and therefore assessment of their future occurrence significantly contributes to the seismic risk mitigation.

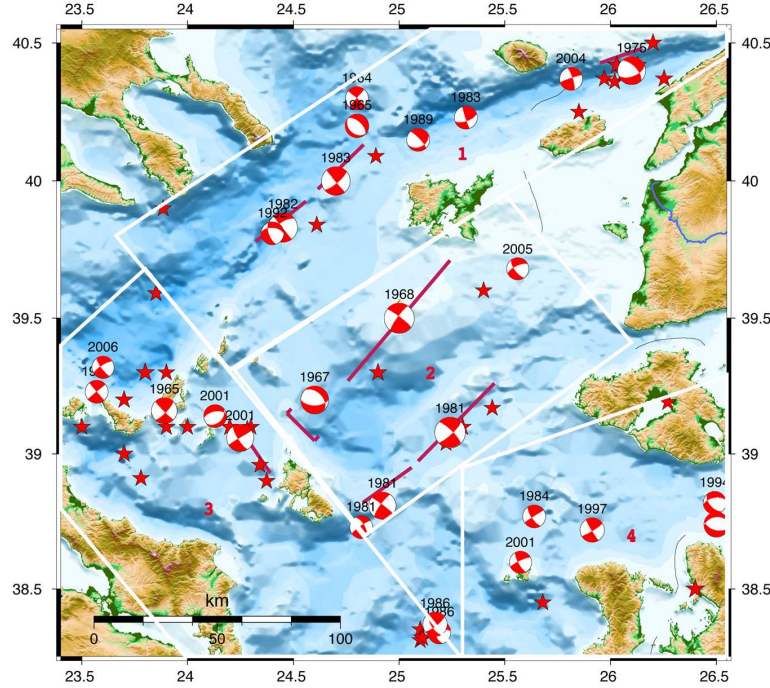


Figure 8: Map of the Northern Aegean study region, showing locations of 67 earthquakes with $M \geq 5.2$ since 1964 and focal mechanisms where available as well as the defined subareas.

Comparing the distributions of the hitting times for an earthquake occurrence with $M \geq 5.6$ between the four subareas (Fig. 9), we conclude that all the subareas present a similar behavior with regard to the particular estimated quantity.

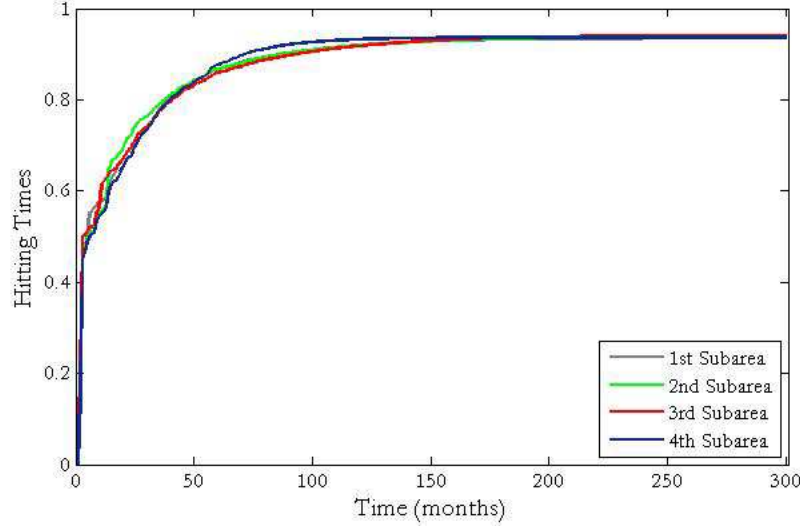


Figure 9: Estimated Distributions of Hitting Times of Earthquakes with $M \geq 5.6$.

5 Conclusions

It has been widely recognized that statistical methodologies can serve as a mathematical tool for the achievement of earthquake forecasting objectives. The seismic hazard modelling approach developed on the basis of statistical methodologies is expected to provide a useful contribution to real time earthquake hazard assessment responding on the social demand for adopting earthquake countermeasures.

It is of great importance to apply the semi-Markov models in order to achieve forecasting results in one of the most seismically active and deforming regions in the world, the area of Northern Aegean Sea. All the relevant quantities including the semi-Markov kernels, the transition functions and the conditional transition functions were estimated. The stationary distribution along with the mean recurrence times to each one of the states were calculated. Moreover, for the proposed semi-Markov model the earthquake occurrence rates including the instantaneous earthquake occurrence rate, the total earthquake occurrence rate and the rate of occurrence of earthquakes with $M \geq 6.1$ were calculated. In addition, the transition probabilities and the distributions of the hitting times to the strongest earthquakes were calculated. Finally, in order to include the spatial component into our model, a new data set was used and the distribution of the hitting time of an earthquake in each one of the defined subareas was estimated.

Further research subjects could include the study of the sensitivity of the models in the determination of the regional division and in the slight perturbation of earthquake magnitudes. For providing more accurate forecasting results one more way is the inclusion of uniquely defined tectonic features. Thus, current work concerns the determination and application of a semi-Markov model based on both coseismic stress changes associated with the occurrence of large earthquakes and slow tectonic stress accumulation, following the procedure of Deng and Sykes (1997).

Acknowledgements

We are grateful to the anonymous referees for their valuable comments. This research has been co-financed by the European Union (European Social Fund-ESF) and Greek national funds through the Operational Program “Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF)- Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund.

References

- [1] Akritas, M. G. and Roussas, G. G. (1979): Asymptotic expansion of the log-likelihood function based on stopping times defined on a Markov process. *Ann. Inst. Statist. Math.*, *31*, 21–38.
- [2] Altinok, Y. and Kolcak, D. (1999): An application of the semi-Markov model for earthquake occurrences in North Anatolia, Turkey. *J. Balkan Geophys. Soc.*, *2*, 90–99.
- [3] Alvarez, E.E. (2005): Estimation in Stationary Markov Renewal Processes, with Application to Earthquake Forecasting in Turkey. *Methodol. Comput. Appl. Probab.*, *7*, 119–130.
- [4] Cluff, L.S., Patwardhan, A.S. and Coppersmith, K.J. (1980): Estimating the probability of occurrences of surface faulting earthquakes on the Wasatch fault zone, Utah. *Bull. Seism. Soc. Am.*, *70*, 1463–1478.
- [5] Deng, J. and Sykes, L.R. (1997): Evolution of the stress field in southern California and triggering of moderate-size earthquakes: a 200–year perspective. *J. Geophys. Res.*, *102*, 9859–9886.
- [6] Garavaglia, E. and Pavani, R. (2009): About Earthquake Forecasting by Markov Renewal Processes. *Methodol. Comput. Appl. Probab.*, doi:10.1007/s11009-009-9137-3.
- [7] Gill, R.D. (1980): Nonparametric estimation based on censored observations of Markov renewal process. *Z. Wahrsch. Verw. Geb.*, *53*, 97–116.
- [8] Lagakos, S.W., Sommer, C.J. and Zelen, M. (1978): Semi-Markov models for partially censored data. *Biometrika*, *65*, 311–317.
- [9] Limnios, N. (1997): Dependability analysis of semi-Markov systems. *Reliab. Eng. Syst. Saf.*, *55*, 203–207.
- [10] Limnios, N. and Oprisan, G. (2001): Semi-Markov Processes and Reliability, Birkhauser, Boston.
- [11] Limnios, N. and Ouhbi, B. (2006): Nonparametric estimation of some important indicators in reliability for semi-Markov processes. *Stat. Methodol.*, *3*, 341–350.
- [12] Limnios, N., Ouhbi, B. and Sadek, A. (2005): Empirical estimator of stationary distribution for semi-Markov processes. *Commun. Stat. - Theor. Methods*, *34*, 287–995.
- [13] Moore, E.H. and Pyke, R. (1968): Estimation of the transition distributions of a Markov renewal process. *Ann. Inst. Statist. Math.*, *20*, 411–424.
- [14] Ouhbi, B. and Limnios, N. (1999): Non-parametric Estimation for Semi-Markov Processes Based on its Hazard Rate Functions. *Statist. Infer. Stoch. Processes*, *2*, 151–173.

- [15] Ouhbi, B. and Limnios, N. (2002): The rate of occurrence of failures for semi-Markov processes and estimation. *Statist. Probab. Lett.*, 59, 245–255.
- [16] Papadimitriou, E.E. and Sykes, L.R. (2001): Evolution of the stress field in the northern Aegean Sea (Greece). *Geophys. J. Int.*, 146, 747–759.
- [17] Papazachos, B.C., Papadimitriou, E.E., Kiratzi, A.A., Papazachos, C.B. and Louvari, E.K. (1998): Fault plane solutions in the Aegean Sea and the surrounding area and their tectonic implications. *Boll. Geof. Teor. Appl.*, 39, 199–218.
- [18] Reasenbergs, P. (1985): Second-order moment of Central California seismicity, 1969–1982. *J. Geophys. Res.*, 90, 5479–5495.
- [19] Rhoades, D.A., Papadimitriou, E.E., Karakostas, V.G., Console, R. and Murru, M. (2010): Correlation of Static Stress Changes and Earthquake Occurrence in the North Aegean Region. *Pure Appl. Geophys.*, doi:10.1007/s00024-010-0092-2.
- [20] Stein, R.S. (1999): The role of stress transfer in earthquake occurrence. *Nature*, 402, 605–609.

List of Symbols

- (E, \mathcal{E}) system state space: general measurable space
- $(J_n)_{n \in \mathbb{N}}$ embedded Markov chain (EMC)
- $(Z_t)_{t \in \mathbb{R}^+}$ semi-Markov process (SMP)
- $(S_n)_{n \in \mathbb{N}}$ jump times of the SMP
- $(S_n^i)_{n \in \mathbb{N}}$ recurrence times for state $i \in E$
- $(X_n)_{n \in \mathbb{N}}$ sojourn/inter-arrival times of the SMP
- $(J_n, S_n)_{n \in \mathbb{N}}$ Markov renewal process (MRP)
- $N(t)$ number of jumps in the time interval $(0, t]$ for the SMP
- $N_i(t)$ number of visits to the state i in the time interval $(0, t]$ for the SMP
- $Q_{ij}(t)$ semi-Markov kernel: discrete state space case; $i \in E, j \in E, t \in \mathbb{R}^+$
- p_{ij} transition function of the Markov chain $(J_n)_{n \geq 0}$: discrete-time case
- $H_i(t)$ distribution function of the sojourn time in state $i, i \in E$: discrete state space case
- $\psi_{ij}(t)$ Markov renewal function: discrete state space case; $i \in E, j \in E, t \in \mathbb{R}^+$
- a initial law
- $Q_1 * Q_2$ Stieltjes convolution of two semi-Markov kernels on (E, \mathcal{E})
- $Q^{(n)}$ n th fold Stieltjes convolution of the semi-Markov kernel $Q, n \in \mathbb{N}$
- μ_{ii} mean recurrence time of $(S_n^i)_{n \in \mathbb{N}}, i \in E$

m_i mean sojourn time in state i : discrete case
 ν stationary distribution for the EMC
 π stationary distribution for the SMP
 $\mathbf{1}$ column vector $(1, \dots, 1)'$
 $\mathbf{1}_A$ indicator (or characteristic) function of a subset A
 $*$ Stieltjes convolution product
 \mathbb{N} set of natural numbers: $\{0, 1, 2, \dots\}$
 \mathbb{R}^+ set of nonnegative real numbers: $[0, \infty)$
 \hat{Q} corresponding estimator of Q

Table 6: Earthquake Catalogue I

<i>Date</i>	<i>Time</i>	<i>Long.</i>	<i>Lat.</i>	<i>M</i>	<i>Date</i>	<i>Time</i>	<i>Long.</i>	<i>Lat.</i>	<i>M</i>
02-05-1953	18:37	38.700	26.500	5.6	14-06-1979	11:44	38.740	26.500	5.9
03-08-1954	18:18	40.100	24.500	5.9	19-12-1981	14:10	39.000	25.260	7.2
02-06-1955	23:34	40.400	25.800	5.5	18-01-1982	19:27	39.780	24.500	7.0
06-01-1956	12:15	40.400	26.300	5.5	06-08-1983	15:43	40.000	24.700	6.8
16-01-1958	04:18	39.500	25.400	5.7	05-10-1984	20:58	39.100	25.300	5.6
11-04-1964	16:00	40.300	24.800	5.5	25-03-1986	01:41	38.340	25.190	5.5
29-04-1964	04:21	39.200	23.700	5.6	29-03-1986	18:36	38.370	25.170	5.8
09-03-1965	17:57	39.160	23.890	6.1	24-05-1994	02:05	38.828	26.492	5.5
23-08-1965	14:08	40.500	26.200	5.6	24-05-1994	02:18	38.845	26.568	5.5
20-12-1965	00:08	40.200	24.800	5.6	04-05-1995	00:34	40.570	23.660	5.8
04-03-1967	17:58	39.200	24.600	6.6	14-11-1997	21:38	38.729	25.913	5.6
19-02-1968	22:45	39.500	25.000	7.1	26-07-2001	00:21	38.995	24.382	6.3
10-03-1968	07:10	39.100	24.200	5.5	06-07-2003	19:10	40.376	26.254	5.5
06-04-1969	03:49	38.500	26.400	5.9	15-06-2004	12:02	40.374	25.813	5.5
17-03-1975	05:35	40.380	26.100	5.8	21-12-2006	18:30	39.355	23.586	5.7
27-03-1975	05:15	40.400	26.100	6.6	09-11-2007	01:43	38.757	25.735	5.5
11-02-1976	07:35	40.530	24.500	5.7					

Table 7: Earthquake Catalogue II

<i>Date</i>	<i>Time</i>	<i>Long.</i>	<i>Lat.</i>	<i>M</i>	<i>Subarea</i>	<i>Date</i>	<i>Time</i>	<i>Long.</i>	<i>Lat.</i>	<i>M</i>	<i>Subarea</i>
23-02-1964	22:41	39.200	23.700	5.4	3	06-05-1984	09:12	38.770	25.638	5.4	4
11-04-1964	16:00	40.300	24.800	5.5	1	29-07-1974	01:58	40.370	25.970	5.2	1
29-04-1964	04:21	39.200	23.700	5.6	3	05-10-1984	20:58	39.100	25.300	5.6	2
29-04-1964	17:00	39.100	23.500	5.2	3	25-03-1986	01:41	38.340	25.190	5.5	3
09-03-1965	17:57	39.160	23.890	6.1	3	29-03-1986	18:36	38.370	25.170	5.8	3
09-03-1965	17:59	39.300	23.800	5.7	3	03-04-1986	23:32	38.350	25.100	5.2	3
09-03-1965	18:37	39.300	23.900	5.2	3	03-06-1986	06:16	38.310	25.100	5.3	3
09-03-1965	19:46	39.100	23.900	5.2	3	17-06-1986	17:54	38.320	25.110	5.4	3
13-03-1965	04:08	39.100	24.000	5.3	3	06-08-1987	06:21	39.190	26.270	5.2	4
13-03-1965	04:09	39.000	23.700	5.5	3	08-08-1987	22:15	40.090	24.890	5.3	1
23-08-1965	14:08	40.500	26.200	5.6	1	27-08-1987	16:46	38.910	23.780	5.2	3
20-12-1965	00:08	40.200	24.800	5.6	1	30-05-1988	16:47	40.250	25.850	5.2	1
04-03-1967	17:58	39.200	24.600	6.6	2	19-03-1989	05:36	39.230	23.570	5.4	3
19-02-1968	22:45	39.500	25.000	7.1	2	05-10-1989	06:52	40.150	25.090	5.4	1
20-02-1968	02:21	39.600	25.400	5.2	2	23-07-1992	20:12	39.810	24.400	5.4	1
10-03-1968	07:10	39.100	24.200	5.5	3	24-05-1994	02:05	38.820	26.492	5.5	4
24-04-1968	08:18	39.300	24.900	5.5	2	16-07-1997	13:06	39.040	25.222	5.2	2
06-04-1969	03:49	38.500	26.400	5.9	4	14-11-1997	21:38	38.720	25.913	5.8	4
17-03-1975	05:11	40.360	26.020	5.3	1	11-04-1998	09:29	39.900	23.884	5.2	1
17-03-1975	05:17	40.390	26.060	5.4	1	22-08-2000	03:35	39.590	23.850	5.2	3
17-03-1975	05:35	40.380	26.100	5.8	1	10-06-2001	13:11	38.600	25.574	5.6	4
27-03-1975	05:15	40.400	26.100	6.6	1	26-07-2001	00:21	39.060	24.248	6.4	3
29-04-1975	02:06	40.420	26.030	5.7	1	26-07-2001	00:34	39.050	24.267	5.3	3
14-06-1979	11:44	38.740	26.500	5.9	4	26-07-2001	02:06	38.960	24.342	5.2	3
12-11-1980	16:04	39.100	24.300	5.3	3	26-07-2001	02:09	38.900	24.373	5.3	3
19-12-1982	14:10	39.000	25.260	7.2	2	30-07-2001	15:24	39.140	24.130	5.4	3
21-12-1981	14:13	39.170	25.431	5.2	2	29-10-2001	20:21	39.090	24.283	5.4	3
27-12-1981	17:39	38.810	24.941	6.5	2	06-07-2003	19:10	40.370	26.254	5.5	1
29-12-1981	08:00	38.700	24.836	5.4	2	06-07-2003	20:10	40.420	26.128	5.2	1
18-01-1982	19:27	39.780	24.500	7.0	1	15-06-2004	12:02	40.370	25.813	5.2	1
18-01-1982	19:31	39.440	24.610	5.6	1	22-11-2004	19:13	38.450	25.679	5.2	4
10-04-1982	04:50	39.940	24.610	5.2	2	24-08-2005	03:06	39.680	25.560	5.2	2
06-08-1983	15:43	40.000	24.700	6.8	1	21-12-2006	18:30	39.320	23.600	5.3	3
10-10-1983	10:17	40.230	25.316	5.4	1						