Karaelmas Fen ve Müh. Derg. 11(1):44-53, 2021

Karaelmas Science and Engineering Journal
Journal home page: https://dergipark.org.tr/tr/pub/karaelmasfen
DOI: 10.7212/karaelmasfen.889013

Research Article
Received / Geliş tarihi : 01.10.2020
Accepted / Kabul tarihi : 26.10.2020

Fitting Hidden Markov Model to Earthquake Data: A Case Study in the Aegean Sea

Saklı Markov Modelinin Deprem Verilerine Uygulanması: Ege Denizinde Bir Örnek Olay

Ozgur Danisman, Umay Uzunoglu Kocer

Dokuz Eylul University, Department of Statistics, Izmir, Turkey

Abstract

Studies about stochastic modeling of earthquake data have increased considerably in recent years. It is a well-known fact that earthquakes occur as a result of unobservable changes in underground stress levels. The hidden Markov model provides a suitable framework for modeling earthquake data due to its assumptions. We present a hidden Markov model to examine hidden changes in the underground stress level and to make some probabilistic earthquake forecasts in the Aegean Sea. The Aegean region is selected for the modeling because of the active nature of earthquake occurrences. A hidden Markov chain is defined in which the corresponding states are stress levels of the ground. Four models with different numbers of hidden states are constructed and compared according to the Akaike and Bayesian information criteria. The proposed model is capable of forecasting the short-term probabilities of both earthquake magnitudes and also locations. Baum-Welch algorithm, which is an iterative expectation-maximization algorithm, is used for the estimation of model parameters. The traditional Baum-Welch algorithm considers only one variable as an observation for the iterations. In this paper, a naive and quite simple approach is used for the Baum-Welch algorithm to estimate the model parameters with more than one observation. It is possible to obtain the marginal and joint probability distributions of multiple observations with this approach.

Keywords: Hidden Markov model, Baum-Welch algorithm, Multiple observations, Earthquake, Forecasting, Aegean Sea

Öz

Deprem verilerinin olasılıksal modellemesi ile ilgili çalışmalar son yıllarda giderek artmaktadır. Depremlerin yer altındaki gerilim düzeyindeki gözlenemeyen değişimler sonucu oluştukları bilinen bir gerçektir. Saklı Markov modelleri varsayımlarından dolayı deprem verilerini modellemek için uygun bir çerçeve sunar. Yeraltı gerilim düzeyindeki gözlenemeyen değişimleri göz önünde bulundurmak ve Ege denizinde bazı olasılıksal deprem tahmini yapmak için gizli Markov modeli sunmaktayız. Ege bölgesi, deprem oluşumu bakımından aktif bir bölge olduğu için seçilmiştir. Gizli durumları yeraltı stres düzeyi olan bir gizli Markov modeli tanımlanmıştır. Gizli durum sayıları farklı olan dört ayrı model tanımlanmış ve bu modeler Akaike ve Bayesian bilgi kriterlerine göre karşılaştırılmıştır. Önerilen model deprem büyüklükleri ve bölgelerine ilişkin kısa dönem olasılık tahminleri verebilmektedir. Model parametrelerini tahmin etmek için iteratif bir algoritma olan Baum-Welch algoritması kullanılmıştır. Geleneksel Baum-Welch algoritması iterasyonlarda yalnız bir gözlem değişkeni kullanır. Bu çalışmada, Baum-Welch algoritmasının birden fazla gözlem değişkeni bulunduğunda kullanımı için oldukça kolay ve anlaşılır bir bakış açısı önerilmiştir. Bu yaklaşımla çoklu gözlem değişkenlerinin marjinal ve ortak olasılık fonksiyonlarını elde etmek mümkün olmaktadır.

Anahtar Kelimeler: Saklı Markov modeli, Baum-Welch algoritması, Çoklu gözlem, Deprem, Tahmin, Ege Denizi

1. Introduction

Analyzing seismic data and developing some statistical methods to forecast earthquake occurrences or earthquake times have been a challenging task for decades. The elastic rebound theory (Reid 1910) is one of the classical models for earthquake mechanisms. It states that stress in a region accumulated due to the movement of tectonic plates, and is released when it exceeds some value of strength. The stochastic version of the elastic rebound theory, called *stress release model*, is proposed (Vere-Jones 1978). Besides, researchers who are interested can find detailed information about the seismicity-based earthquake forecasting techniques (Tiampo and Shcherbakov 2012).

There have been some studies aiming to model the general structure of seismic data by using the Markov process or Poisson process. However, earthquakes are affected by various

*Corresponding author: ozgur.danisman@deu.edu.tr

Ozgur Danisman orcid.org/0000-0001-5855-6413
Umay Uzunoglu Kocer orcid.org/0000-0001-5044-6236

natural factors such as the structure of the fault and previous earthquake occurrences, and these events may change the magnitude and/or the occurrence time of an earthquake. Some of these natural events are observable and some of them are not. Since there is a dependency between these natural factors and the earthquakes, a modeling approach that characterizes this dependency is required. Since it is known that the Poisson process has memorylessness property and it assumes that different time intervals for a specific region are independent, the Poisson process usually does not provide an adequate modeling setting.

The dependence between the previous and the most recent earthquake magnitudes encouraged researchers to study modeling the seismic data using the Markov process. When the Markov process is used for earthquake forecasting, the observed data becomes the states of the system and unobservable natural events are not included in the model. See Anagnos and Kiremidjian (1988) and its references for the first applications of the Poisson process and Markov process for seismic studies. Sojourn times are assumed to be exponential for Markov processes; on the other hand, it is possible to assume different sojourn time distributions such as Weibull or Gamma for semi-Markov processes. There also exists some studies about the modeling of earthquake data with the semi-Markov process. A semi-Markov process is proposed (Alvarez 2005) where a mixture of exponential and Weibull distribution is assumed for waiting times. Construction of a semi-Markov process is proposed (Sadeghian 2012) to predict the probability of the time and place of occurrence of earthquakes. In another study that assumed sojourn times as Weibull distribution (Masala 2012) proposed both homogeneous and non-homogeneous semi-Markov process and he allowed dependence in both space and time, in their model. A semi-Markov model for the estimation of the expected number of earthquakes is considered and a seismic hazard assessment was presented for Northern Aegean Sea (Votsi et al. 2012). The semi-Markov kernel and the distributions of sojourn times were estimated by the help of a nonparametric method. In addition, a detailed information about the seismology from the statistical point of view, about discrete HMM and hidden Markov renewal model was presented (Votsi 2019).

The existence of the unobservable factors that affect the earthquake occurrence led to the start of using the hidden Markov model (HMM) to characterize the earthquake occurrences. HMM is a special type of Markov process in which states are unobservable. An unobservable Markov chain leads the system and observations occur after each transition. This type of a model is more suitable to model the seismic data because unobservable natural events are assumed to be states of a hidden Markov chain and some of the outputs such as magnitude or occurrence time are the observations occurring after each transition. Hence, the HMM is studied instead of the Markov process and Poisson process for the earthquake analysis since the early 2000s.

HMM is first introduced with the definition as “emitted probabilistic functions of unobservable finite-state Markov chains” (Baum et al. 1966). For modeling seismic data, the very first study (Granat and Donellan 2002) used the GPS seismicity data collected in the Southern California region. Hidden states were the changes in stress within a fault system and earthquake forecasts were presented. It is claimed that the Poisson process which is a traditional model for earthquakes is insufficient for forecasting future events and HMM is proposed (Chambers et al. 2003). They assumed each hidden state leads to an exponential distribution with different rates and overall time distribution is a mixture of exponential distributions. They used 110 interevent times (in days) between 1975 and 2000 and related earthquakes have magnitudes with $M \geq 2.7$ and assumed two hidden states; rapid response and slow response to a quake. The California region is studied (Ebel et al. 2007) with earthquake data which have magnitude values $M \geq 4$. Inter-event times between earthquakes were assumed to be exponential and the short-term future earthquakes (in one day) were forecasted in a dynamic way using HMM. Observations were the inter-event times and one of four spatial quadrants. Poisson HMM was used to model the seismic variations and earthquake frequencies (Orfanogiannaki et al. 2010). The earthquakes greater than the magnitude $M \geq 3.2$ around the Killini, Ionian Sea were used between 1990 and 2006. Observations were assumed the daily number of earthquakes and inter-event times. Poisson HMM was fitted for 2,3,4 and 5 states and models with four and five hidden states were selected. HMM is applied to earthquake data ($M \geq 4$) from 1932 to 1964 in the California region (Chambers et al. 2012). They used two different models with two and four states, respectively. Inter-event time distribution was assumed to be exponential. Observations were assumed inter-event times and one of two regions (east or west) that an earthquake may occur. HMM is used for earthquake data from 1845 to 2013 in the Greece region (Votsi et al. 2013). Hidden states were assumed the stress level of the ground. Different HMMs with a different number of states were examined and a two-state model was selected. Observations were the classes of earthquakes magnitudes and their aim was the estimation of the expected time until the next earthquake. When the stress level of the ground exceeds a threshold, an earthquake occurs (Yip et al. 2018). They assumed this threshold divides the underground stress level into two states: low and high level of earthquake frequency. Their model includes earthquake magnitudes as observations and a non-stationary transition probability matrix that varies by time. A comprehensive implementation of hidden semi-Markov models for the seismic hazard assessment was presented (Pertsinidou et al. 2017). The earthquakes which have a magnitude $M \geq 2.7$ around the North and South Aegean Sea were taken into consideration. Models with different dimensions were developed and compared with different sojourn time distributions.

In this study, an HMM is proposed for modeling the earthquake data of the Aegean Sea region. The complete data contains the depth, location, magnitude, and inter-event time of an earthquake. HMMs with different numbers of hidden states are constructed and the most likely model is selected according to likelihood values. By using the most suitable model, short-term probabilities of earthquake magnitudes and earthquake locations are forecasted. The conditional probability distribution of earthquake magnitudes and earthquake locations are obtained. Then, the joint probability distribution of random variables magnitude and location is obtained. The resulting model is capable of forecasting the probabilities of earthquake magnitude classes and the probabilities of earthquake locations.

The common point of all existing studies is that they assumed only one variable as the observation for the estimation of model parameters. Other variables are included in the model after parameter estimation with some independency assumptions. However, three types of observations which are depth, location, and magnitude are used for the estimation of model parameters in our study. We aim to obtain the estimates with more earthquake information. These three types of observations are redefined as a single variable and then, marginal and joint probability distributions of earthquake magnitude and location are obtained.

The main contributions of our paper can be summarized as follows: (1) A modeling approach is proposed in which multiple observations are used for the parameter estimation. (2) Marginal and joint probability distributions of all observation random variables are obtained by this approach. (3) The proposed HMM gives probabilistic forecasts of the magnitude and location of a future earthquake in the Aegean Sea, Turkey. Earthquake forecasting for this region is provided for the first time by this modeling approach.

In section two, a brief introduction to HMMs and its complete parameter set is given. Then, the method of modeling the earthquake data and selecting the best representative model between different HMMs are presented. In section three, results and the related probability distributions of earthquake magnitude and location is presented. The probabilities of occurring earthquakes in a region are forecasted in one day and one week. And finally, discussions, concluding remarks, and some possible further studies are presented in the section four.

2. Material and Method

HMMs are discrete-time stochastic processes which consist of two sets of random variables. One is a Markov chain with some finite states and the other is a set of observations occurring immediately after each transition of the underlying Markov chain. The transitions of the Markov chain are invisible (hidden) to a viewer outside of the system and the current state is always dependent on the previous state because of the Markovian property. The observations are emitted by a probability distribution corresponding to the current state of the Markov chain. They are independent of each other and only depend on the current hidden state. It is not possible to observe the states of the system directly and someone can record only the observations. HMMs are different from traditional Markov models due to these assumptions. The observed data is not the actual state of the system for HMMs, these observations are emitted by the underlying hidden states. On the other hand, the observations correspond directly to the states for traditional Markov models. So, the term "hidden" comes from the first-order Markov chain behind the observations.

2.1. General definition of HMM

Let v_k be the k th observation and there are M possible emitted observations per state. The system makes a transition and emits one of the observations $\{v_1, v_2, ..., v_M\}$. There are N hidden states, hence the transition probability matrix is N -dimensional square matrix. The hidden state sequence and the observation sequence are finite sets and T is the length of these sets. Let X_t be the random variable that represents the state of the hidden Markov chain at time t ,where $X_t \in \{1, 2, ..., N\}, t = 0, 1, ..., T$. Furthermore let Y_t be the random variable that represents the observation in a state at time t, where $Y_t \in \{v_1, v_2, ..., v_M\}$ and $t = 0,1,...,T$. So, the sequence of hidden states and observations are denoted by $\{X_1, X_2, ..., X_T\}$ and $\{Y_1, Y_2, ..., Y_T\}$, respectively.

Transition probability matrix (TPM) is denoted by $P = \{p_{ij}\}$. Each element of P is a transition probability from hidden state *i* to state *j*. That is,

$$p_{ij} = P(X_{t+1} = j \mid X_t = i\},\ i,j = 1,2,...,N \text{ and } t = 1,2,...T. \quad (1)$$

The model emits an observation after each transition and emission information is given by *observation probability distribution*. The emission probability of an observation at time t may change from state to state. Since there are N hidden states and there are M possible observations in each state, the observation probability matrix is denoted by $B = \{b_j(k)\}$ and consists of $N \times M$ entry. This matrix gives the probability of emitting an observation given that the system is in any hidden state. That is,

$$b_j(v_k) = P(Y_t = v_k \mid X_t = j),\ j = 1,2,...,N,\ k = 1,2,...,M \text{ and } t = 1,2,...T. \quad (2)$$

The initial state distribution gives the starting probabilities and usually given by $N \times 1$ vector. Let π_i be the probability of starting at state *i*. Then,

$$\pi_i = P(X_o = i),\ \text{where } i = 1,...N. \quad (3)$$

These three parameters construct the complete parameter set $\lambda = (P, B, \pi)$ of a HMM.

2.2. Modeling of earthquake data by HMM

HMM has been one of the ways of modeling the earthquake data that include magnitudes, inter-event times, earthquake frequencies, depth of earthquakes, or locations. Earthquakes are caused by a sudden release of stress along faults. When the stress on the fault exceeds a threshold, an earthquake occurs, and energy releases. The stress level of the fault changes before each earthquake and these changes are unobservable. It is possible to observe and measure the magnitude or the occurrence time of an earthquake but, the same does not apply to the stress level that causes the earthquake. Hence, the states of the hidden Markov chain is assumed the "stress level" of the fault. Hidden states are denoted by X_t which is defined on the state space $S = \{1,2,...,N\}$, and also when a hidden state number decreases, the stress level of fault decreases as well: state 1 corresponds to the minimum stress level and state N corresponds to the maximum stress level. After each transition, four different observations are emitted: location, depth, inter-event time between previous and last earthquake, and the magnitude of an earthquake.

Location (L). The Aegean Sea region is selected for modeling because of the active nature of earthquakes. The region is divided into two locations. Region I(L_I) is the Mytilene Island area and region II(L_{II}) is the Aegean Sea area. These two regions continuously experience successive earthquakes. An earthquake can occur either in region I or region II. Selected regions are shown in Figure 1.

Depth (D). The depth of earthquakes is divided into two classes according to the median value. The median value is found as *Med* = 10 and 10*km* is accepted as the threshold value between two classes. That is, an earthquake can occur either more than or equal to 10*km* deep ($D \geq 10$) or less than 10*km* deep ($D < 10$).

Magnitude (M). For the corresponding regions, the earthquakes with a magnitude greater than 3 are included in the data set. The magnitude values of M are classified into four groups with the help of the *k*-means clustering method. An earthquake magnitude between [3.0, 3.3) is included in class 1, between [3.0, 3.6) is included in class 2, between [3.6, 3.9) is included in class 3, and if $M \geq 3.9$, it is included in class 4.

Inter-event times (T). Inter-event times between successive earthquakes are assumed to be independent of other variables. It is assumed that it depends only on the hidden states and inter-event times between two successive earthquakes follow an exponential distribution. So, the parameter of the related inter-event time distribution depends on the hidden state it is emitted. In other words, each state must have different exponential rates. The Viterbi (Viterbi 1967) algorithm, which is an algorithm to find the most likely hidden state

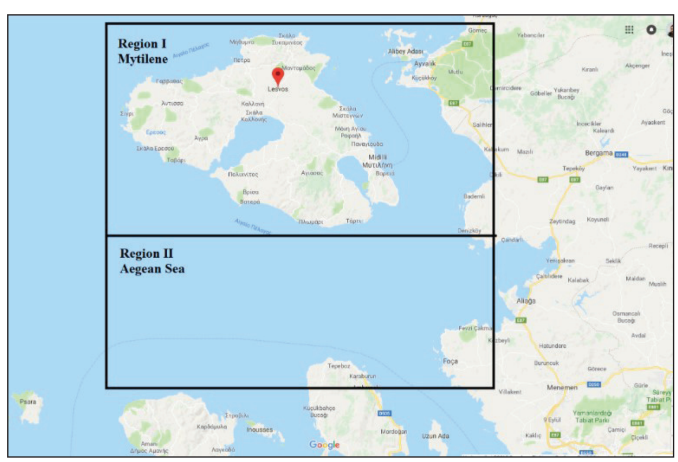

Figure 1. Possible earthquake locations.

sequence, is needed to estimate model parameters of inter-event times, in days, for each hidden state. Since there is no information about hidden states of past earthquakes, these 1107 hidden states are estimated with the Viterbi algorithm.

The main contribution of our study is to allow multiple observations included in the parameter estimation part of the model. The traditional Baum-Welch algorithm (Baum et al. 1970) allows using only one variable as an observation. Existing studies with HMM for earthquake forecasting generally include only one observation for learning problem and only one variable is used while estimating the model parameters with the Baum-Welch algorithm. Some of the existing studies assume the observation as the number of earthquakes occurred in independent time intervals, whereas some of them assume the magnitude of an earthquake as an observation. The inter-event times and locations of earthquakes are generally assumed as independent factors and the joint distribution of these random variables with the inter-event time is obtained by multiplying their marginal distributions. In our study, inter-event times between successive earthquakes are also assumed to be exponential. Different from the previous studies, three variables which are *location*, *depth,* and *magnitude* are included as the observations to the Baum-Welch algorithm. These three variables are expressed as a new single random variable for the Baum-Welch algorithm iterations. That is, more information is used to estimate the complete parameter set. Obtaining the marginal distributions of magnitude values and locations, and also the joint distribution of these two variables allows us to draw some clear and accurate conclusions.

2.3. Model selection

Four different models are constructed with different numbers of hidden states $(N = 2,...,5)$ and the best model that fits the data is to be selected. There exist 2 locations, 2 depth classes, and 4 magnitude classes. It is possible to show these multivariate data sets as a one-dimensional new variable by taking the Cartesian product of these three random variables. So, the new type of observation includes sixteen types of possible observation. For instance, $\{L_{II}, D < 10\,km, M \in [3.0, 3.3]\}$ is the joint event of an occurrence of an earthquake between magnitude 3 and 3.3 at the Aegean Sea region with a depth of less than 10*km*. This joint event is just one of the sixteen events. Each of these joint events is one of the possible observations to be emitted. Now, a hidden state sequence can be obtained. The complete earthquake data is fitted for the process of the Baum-Welch algorithm. Note that this algorithm does not give us the marginal distributions of each variable. The observation probability distribution of the new one-dimensional variable is obtained.

Model parameters are estimated by the Baum-Welch algorithm which is an iterative algorithm based on expectation-maximization. The starting values for initial state distribution, observation emission distribution, and transition probabilitiy distribution are selected from Uniform distribution. Stopping criteria for estimates is $\varepsilon = 10^{-4}$. That is, when the difference between the last estimates and the previous estimates gets smaller than the identified value of ε, the algorithm is stopped. Parameters are estimated for four different models. The likelihood values of the four models are used to select the most likely model. Akaike Information Criterion (AIC) (Akaike 1974) and Bayesian Information Criterion (BIC) (Schwarz 1978) values are obtained to decide the most suitable model. The model with the lowest AIC and BIC value is selected as the best representative model that fits data. Likelihood, AIC, and BIC values are shown in Table 1. *P* is the degrees of freedom of each model. From Table 1, AIC and BIC take the smallest value for the two-state model. According to Table 1, the HMM with two hidden states fits better the data.

Table 1. Likelihood and information criterion values

State	P	*lnL*	AIC	BIC
2	4	-918.15	1844.3*	1864.3*
3	9	-1447.92	2913.84	2958.92
4	16	-917.54	1867.08	1947.23
5	25	-1124.76	2249.52	2424.75

For two-state HMM, state 1 corresponds to low stress level whereas state 2 corresponds to the high-stress level. The estimated values of hidden state transition probabilities are given below.

$$\hat{P} = \begin{pmatrix} 0.5075 & 0.4925 \\ 0.9999 & 0.0001 \end{pmatrix} \quad (4)$$

According to the transition probability distribution, the probability of a transition from “low-stress level” to “high-stress level” is 0.4925 and the probability of a transition from “high-stress level” to “low-stress level” is 0.9999. It is known that one of the main reasons for severe earthquakes is the high-stress level of the ground and these earthquakes generally occur very rarely. The high-stress level makes a transition to the low-stress level just after a severe

Table 2. Observation probability distribution given the hidden states

Obs.	Event	State I	State II	Obs.	Event	State I	State II
1	L_I, $D \le 10$, $M \in [3.0,3.3)$	0.1240	0.1479	9	L_{II}, $D \le 10$, $M \in [3.0,3.3)$	0.1213	0.1199
2	L_I, $D \le 10$, $M \in [3.3,3.6)$	0.0317	0.0846	10	L_{II}, $D \le 10$, $M \in [3.3,3.6)$	0.1259	0.0182
3	L_I, $D \le 10$, $M \in [3.6,3.9)$	0.0095	0.0136	11	L_{II}, $D \le 10$, $M \in [3.6,3.9)$	0.0172	0.0564
4	L_I, $D \le 10$, $M \ge 3.9$	0	0.0055	12	L_{II}, $D \le 10$, $M \ge 3.9$	0	0.0547
5	L_I, $D \ge 10$, $M \in [3.0,3.3)$	0.1525	0	13	L_{II}, $D \ge 10$, $M \in [3.0,3.3)$	0.1672	0.1069
6	L_I, $D \ge 10$, $M \in [3.3,3.6)$	0.0451	0.0317	14	L_{II}, $D \ge 10$, $M \in [3.3,3.6)$	0.0771	0.1544
7	L_I, $D \ge 10$, $M \in [3.6,3.9)$	0	0.0328	15	L_{II}, $D \ge 10$, $M \in [3.6,3.9)$	0.0546	0.0596
8	L_I, $D \ge 10$, $M \ge 3.9$	0.0038	0.0224	16	L_{II}, $D \ge 10$, $M \ge 3.9$	0.0701	0.0914

earthquake. There are very frequent non-violent earthquakes in the investigated area, Turkey side of the Aegean Sea. In other words, the stress level of the region is mostly low. This is why the system tries to make a transition to the low-stress level, especially from the high-stress level. When the region's tendency to non-violent earthquakes and the selected model is considered, the transition probability matrix gives the expected probabilities of hidden transitions. The process tends to make transitions from the low-stress level to both states with almost the same probabilities. The estimated values of observation emission probabilities are given in Table 2.

Recall that hidden states depend only on the previous state and the transition probability matrix is independent of observations. Table 2 gives the probabilities of emissions given the hidden states. Emission probabilities for sixteen observations are obtained. It is clear from Table 2 that when the magnitude value gets higher, the probability of observation emission gets lower. That is, there is a small tendency for an earthquake with high magnitudes.

3. Results

It is possible to obtain the conditional probability distributions of magnitudes and locations using Table 2 by considering the marginalization according to other observations. If the marginalization is taken by the depth and location, the distribution of magnitude classes is obtained given the hidden states. If the marginalization is taken by the depth and magnitude, the distribution of locations is obtained given hidden states. These probabilities are presented in Table 3.

This conditional distribution of magnitude classes presented in Table 3 gives the conditional probability,

Table 3. Marginal distribution of earthquake class given the hidden states

	State I	State II
1: $M \in [3.0, 3.3)$	0.5650	0.3747
2: $M \in [3.3, 3.6)$	0.2798	0.2889
3: $M \in [3.6, 3.9)$	0.0813	0.1624
4: $M \ge 3.9$	0.0739	0.1740

$$P(M = m \mid X_t = j), \quad j = 1,2 \text{ and } m = 1,2,3,4. \tag{5}$$

For instance, given that stress level is low, the probability of observing an earthquake with a magnitude greater than or equal to 3.9 is 0.0739. It is seen from the magnitude distribution that when the stress level is low, there is a tendency of occurring non-violent earthquakes. The probabilities of class 1 and class 2 are much greater than the probabilities of class 3 and class 4. However, the same conclusion cannot be made for the high-stress level. The probabilities of class 3 and class 4 for the high-stress level are approximately two times greater than that of the low-stress level. And also the probabilities of class 1 and class 2 for the high-stress level are much lower than that of the low-stress level. The tendency of occurring earthquakes with higher magnitude increases when the stress level is high.

Table 4. Marginal distribution of location given the hidden states

	State I	State II
Mytilene Island (L_I)	0.3666	0.3385
Aegean Sea (L_{II})	0.6334	0.6615

The conditional distributions related to locations presented in Table 4, gives the probability,

$$P(L = l \mid X_t = j), \; j = 1,2 \text{ and } l = L_I, L_{II}. \tag{6}$$

For instance, the probability of observing an earthquake in Mytilene Island is 0.3666 given that the stress level is low. The location distribution shows that the number of earthquakes that the Aegean Sea is exposed, is more than that of the Mytilene Island for both low and high-stress levels.

The joint observation probability distribution of the magnitude and location is obtained by obtaining the marginalization of depth.

Table 5. Joint distribution of location and magnitude given the hidden states

	State I	State II
$L_I, M \in [3.0,3.3)$	0.2765	0.1479
$L_I, M \in [3.3,3.6)$	0.0768	0.1163
$L_I, M \in [3.6,3.9)$	0.0095	0.0464
$L_I, M \geq 3.9$	0.0038	0.0279
$L_{II}, M \in [3.0,3.3)$	0.2885	0.2268
$L_{II}, M \in [3.3,3.6)$	0.2030	0.1726
$L_{II}, M \in [3.6,3.9)$	0.0718	0.1160
$L_{II}, M \geq 3.9$	0.0701	0.1461

The conditional joint distribution of magnitude and location presented in Table 5 gives the probability,

$$P(M=m, L=l \mid X_t=j),\ j=1,2 \text{ and } m=1,2,3,4 \text{ and } l=L_I, L_{II}. \quad (7)$$

From Table 5, the probability of observing an earthquake in Mytilene Island with a magnitude between 3.3 and 3.6 is 0.0768. Note that the sum of observation probabilities for each hidden state is equal to one.

Apart from the marginal and joint probabilities for magnitude and location with the condition of hidden states, it is possible to obtain unconditional probability distribution for each observation by using the limiting probabilities of hidden states. Using the estimated transition probability matrix $(\hat{\boldsymbol{P}})$, it is possible to compute the limiting distribution of hidden states. Let $\boldsymbol{w}$ be the vector of steady-state probabilities, then

$$\boldsymbol{w} = [w_1, w_2] = [0.6699, 0.3301]. \quad (8)$$

The limiting probabilities can be considered as the weights and the weighted means of hidden states are calculated for the conditional probability distributions. Then, the marginal probability distribution of magnitude variable given in Table 6 is obtained as,

$$P(M=m) = \sum_{j=1}^{2} w_j b_{j_M}(k),\ m=1,2,3,4. \quad (9)$$

Table 6. Marginal distribution of magnitude class

	1:	2:$M \in [3.3,3.6)$	3:	4: M≥3.9
P (M=m)	0.5022	0.2828	0.1080	0.1070

The marginal probability distribution of location variable given in Table 7 is obtained as,

$$P(L=l) = \sum_{j=1}^{2} w_j b_{j_L}(k),\ \text{where } l=L_I, L_{II}. \quad (10)$$

Table 7. Marginal Distribution of Location

	Mytilene Island (L_I)	Aegean Sea (L_{II})
P (L=l)	0.3574	0.6426

The joint probability distribution of magnitude and location variables given in Table 8 is obtained as,

$$P(M=m, L=l) = \sum_{j=1}^{2} w_j b_{j_{ML}}(k),\ \text{where } m=1,2,3,4 \ and\ l=L_I, L_{II}. \quad (11)$$

Table 8. Joint distribution of location and magnitude

	Mytilene Island (L_I)	Aegean Sea (L_{II})
1: $M \in [3.0,3.3)$	0.2341	0.2681
2: $M \in [3.3,3.6)$	0.0898	0.1930
3: $M \in [3.6,3.9)$	0.0216	0.0864
4: $M \geq 3.9$	0.0119	0.0951

To check the independency of the random variables magnitude, location, and depth, we check the following equality.

$$P(M=k)P(L=m) = P(M=k, L=m)$$

For instance, $P(M=1)=0.5022$ and $P(L=L_I)=0.6426$ and their product is 0.3227. On the other hand, their joint probability is $P(M=1, L=L_I)=0.2681$. Hence, it can be concluded that these two random variables are not independent. So, including magnitude, location and depth variables together give more accurate results.

Having estimated the hidden state sequence, now it is possible to estimate the exponential distribution parameters for each hidden state. Inter-event times for state 1 (low-stress level) follows an exponential distribution with $\hat{\lambda}_1 = 0.1723$ and inter-event times for state 2 (high-stress level) follows an exponential distribution with $\hat{\lambda}_2 = 0.2025$

. So, the mean times until the next earthquake can be calculated using the parameters, that is, $\mu_1 = \frac{1}{\hat{\lambda}_1} = 5.8048$ days and $\mu_2 = \frac{1}{\hat{\lambda}_2} = 4.9371$ days, for each state respectively. The mean waiting time until the next earthquake is larger for the low-stress level than that of the high-stress level. That is, when the stress level is high, it is more likely to occur an earthquake more quickly. The mean waiting time ($E(t)$) is calculated with a mixture of two non-identical and independent exponential distributions ($\hat{\lambda}_1$ and $\hat{\lambda}_2$). Let p_j is the probability of observing an inter-event time from the jth hidden state and note that $\sum_{j=1}^{2} p_1 = 1$. Two proportions p_1 and p_2 are estimated using the 1106 inter-event time according to which hidden state it was generated. The estimated values of proportions based on hidden states are, $\hat{p}_1 = 0.5844$ and $\hat{p}_2 = 0.4156$. It means that 58.44% of the inter-event times are emitted at the low-stress level and 41.56% of the inter-event times are estimated at the high-stress level. So, the mean waiting time between two successive earthquakes is, $E(t) = \sum_{j=1}^{2} \frac{p_i}{\lambda_j} = 5.4441$ days. Besides, the mean rate of the distribution ($\hat{\lambda}$) can be obtained as $\hat{\lambda} = \frac{1}{5.4441} = 0.1836$.

Short-term probability forecasting. It is possible to give forecasts of short-term probabilities for earthquake occurrences. The aim is to find the probability of earthquake occurrence in a region, in t days. For this purpose, we must use the joint probability distribution of magnitude (M), location (L), and inter-event time (T). It has been assumed before that inter-event time is independent of other variables. So, the joint probability distribution of M, L, and T can be obtained by the production of the joint distribution of M and L, and distribution of T. So, the probability of observing an earthquake at a magnitude class k at one of two locations in t days is,

$$P(M = k, L = l, T \leq t) = P(M = k, L = l)\,(1 - e^{-\hat{\lambda} t}) \qquad (12)$$

where $M = 1,2,3,4,\ l = L_I, L_{II},$ and $t \geq 0$.

If the time index t goes to infinity, the probabilities approach to some limiting values given in section 2.4. and become independent of t. We already know that the probability distributions defined in section 2.4 hold the conditions to be a probability function. So, the joint probability distributions of the variables M, L, and T also hold the conditions to be a probability function under the limiting case.

We can make short-term forecasting for all the elements of the variable vector. Note that observations only depend on the related hidden states according to the HMM assumptions. Table 9 shows the short-term probabilities of observing earthquakes in one week.

According to Table 9, the probabilities of occurring an earthquake in any region increases as t goes to infinity. When $t \to \infty$, the probabilities converge to the joint probabilities given in Table 8. Suppose that today an earthquake occurred. Then, the probability of occurring an earthquake with the magnitude between [3.0, 3.3) in the Aegean Sea deficits in 5 days is,

$$P(M = 1, L = L_{II}, T \leq 5) = 0.1610. \qquad (13)$$

4. Discussion and Conclusion

HMM is used to model the earthquake occurrences in the Aegean Sea region. Among the possible models, the most likely model is selected by AIC and BIC after the learning problem. Model parameters are estimated by the Baum-Welch algorithm. The traditional Baum-Welch algorithm allows only one observation to perform the iterations. Different from the existing studies, three types of variables which are magnitude, location, and depth are represented as a one-dimensional random variable to perform the algorithm. This procedure allowed us to estimate the model

Table 9. One-week earthquake occurrence probabilities

Event	t (in days)							
	1	2	3	4	5	6	7	∞
$M = 1, L = L_I$	0.0393	0.0719	0.0991	0.1218	0.1406	0.1563	0.1693	0.2341
$M = 2, L = L_I$	0.0150	0.0276	0.0380	0.0467	0.0539	0.0599	0.0649	0.0898
$M = 3, L = L_I$	0.0036	0.0066	0.0091	0.0112	0.0130	0.0144	0.0156	0.0216
$M = 4, L = L_I$	0.0020	0.0036	0.0050	0.0062	0.0071	0.0079	0.0086	0.0119
$M = 1, L = L_{II}$	0.0449	0.0824	0.1135	0.1394	0.1610	0.1789	0.1939	0.2681
$M = 2, L = L_{II}$	0.0323	0.0593	0.0817	0.1004	0.1159	0.1288	0.1396	0.1930
$M = 3, L = L_{II}$	0.0145	0.0265	0.0366	0.0449	0.0519	0.0577	0.0625	0.0864
$M = 4, L = L_{II}$	0.0159	0.0292	0.0403	0.0495	0.0571	0.0635	0.0688	0.0951

parameters with more information. Studies on this subject generally consider only one variable as an observation. In this case, the effects of other variables on the estimation of earthquake occurrence probability are ignored, resulting in loss of information and inconsistent forecasts. After the parameter estimation, marginal and joint probability distributions of magnitude and location are obtained using the observation probability distribution. Since there is usually only one observation in other studies, the joint probability function can not be obtained. However, with the approach in our study, the joint and marginal probability distributions of depth, location, and magnitude variables can be obtained. Inter-event times are assumed to be exponential and the mean rate of the times between earthquakes is estimated by the mean waiting time of two non-identical and independent exponential distributions. The mean time until the next earthquake is obtained. It is concluded that the mean time until the next earthquake is larger for the low-stress level than that of for high-stress level. When the stress level is high, it is more likely to occur an earthquake more quickly so, this is an expected result. It is obvious that the inter-event times between successive earthquakes will affect the magnitude of the next earthquake. In this context, there is a lack of information in studies that do not include inter-event times to the model and that fit discrete-time modeling.

The proposed model is capable of forecasting the probabilities of short-term earthquake occurrences using the magnitude and the location information. The joint probability distribution of magnitude, location, and inter-event time is obtained. The presented joint distribution has the properties of a probability distribution function. Probability distribution approaches its limiting value when the time index t goes to infinity. An observer is capable of learning the probability of an earthquake occurring in a region within a magnitude class in the short-term.

Further Study. The dependency structure between the observations and hidden states can be studied in more detail. Different dependency structures may be considered. Besides, a possible interest may be the frequencies of earthquake occurrences or the estimation of severe earthquake arrival times. As in this study and other studies, obtaining a joint distribution function of inter-event time and other variables, instead of assuming that the inter-event time is independent becomes a challenging topic for future studies.

Acknowledgement

The authors would like to thank the anonymous referees for their invaluable and helpful comments for the improvement of the study.

5. References

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Automat.* Control, 19:716-723. https://doi.org/10.1007/978-1-4612-1694-0_16

Alvarez, EE. 2005. Estimation in stationary Markov renewal processes, with application to earthquake forecasting in Turkey. *Methodol. Comput. Appl.*, 7:119-130. https://doi.org/10.1007/s11009-005-6658-2

Anagnos, T., Kiremidjian, AS. 1988. A review of earthquake occurrence models for seismic hazard analysis. *Probab. Eng. Mech.*, 3:3-11. https://doi.org/10.1016/0266-8920(88)90002-1

Baum, LE., Petrie, T. 1966. Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Stat.*, 37:1554-1563. http://dx.doi.org/10.1214/aoms/1177699147

Baum, LE., Petrie, T., Soules, G., Weiss, N. 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Stat.*, 41:164-171. http://dx.doi.org/10.1214/aoms/1177697196

Chambers, DW., Baglivo, JA., Ebel, JE., Kafka, AL. 2012. Earthquake forecasting using hidden Markov models. *Pure. Appl. Geophys.*,169:625-639. https://doi.org/10.1007/s00024-011-0315-1

Chambers, DW., Ebel, JE., Kafka, AL., Baglivo, JA. 2003. A hidden Markov approach to modeling interevent earthquake times. *Eos. Trans. AGU Fall. Meet. Suppl.*, 84(46), Abstract S52F-0179

Ebel, JE., Chambers, DW., Kafka, AL., Baglivo, JA. 2007. Non-Poissonian earthquake clustering and the hidden Markov model as bases for earthquake forecasting in California. *Seismol. Res. Lett.*, 78:57-65. http://dx.doi.org/10.1785/gssrl.78.1.57

Granat, R., Donnellan, A. 2002. A hidden Markov model based tool for geophysical data exploration.. *Pure. Appl. Geophys.*, 159:2271-2283. https://doi.org/10.1007/s00024-002-8735-6

Masala, G. 2012. Earthquakes occurrences estimation through a parametric semi-Markov approach. *J. Appl. Stat.*, 39:81-96. https://doi.org/10.1080/02664763.2011.578617

Orfanogiannaki, K., Karlis, D., Papadopoulos, GA. 2010. Identifying seismicity levels via Poisson hidden Markov models. *Pure. Appl. Geophys.*, 167:919-931. https://doi.org/10.1007/s00024-010-0088-y

Pertsinidou, CE., Tsaklidis, G., Papadimitriou, E., Limnios, N. 2017. Application of hidden semi-Markov models for the seismic hazard assessment of the North and South Aegean Sea, Greece. *J. Appl. Stat.*, 44:1064-1085. http://dx.doi.org/10.1080/02664763.2016.1193724

Reid, H.F. 1910. The mechanics of the earthquake, the California earthquake of April 18, 1906. Report of the state investigation comission, vol 2. Carnegie Institution of Washington, Washington D.C. (reprinted in 1969).

Sadeghian, R. 2012. Forecasting time and place of earthquakes using a semi-Markov model (with case study in Tehran province). *J. Ind. Eng. Int.*, 8:20. https://doi.org/10.1186/2251-712X-8-20

Schwarz, GE. 1978. Estimating the dimension of a model. *Ann. Statist.*, 6:461-464. https://www.jstor.org/stable/2958889

Tiampo, KF., Shcherbakov, R. 2012. Seismicity-based earthquake forecasting techniques: Ten years of progress. *Tectonophysics*, 522-523:89-121.

Vere-Jones, D. 1978. Earthquake prediction: a statistician's view. *J. Phys. Earth.*, 26:129-146. https://doi.org/10.4294/jpe1952.26.129

Viterbi, AJ. 1967. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans. Inf. Theory.*, 13:260-269. DOI:10.1109/TIT.1967.1054010

Votsi, I., Limnios, N., Tsaklidis, G., Papadimitriou, E. 2012. Estimation of the expected number of earthquake occurrences based on semi Markov models. *Methodol. Comput. Appl. Probab.*, 14:685-703. https://doi.org/10.1007/s11009-011-9257-4

Votsi, I., Limnios, N., Tsaklidis, G., Papadimitriou, E. 2013. Hidden Markov models revealing the stress field underlying the earthquake generation.. *Physica A*, 392:2868-2885. https://doi.org/10.1016/j.physa.2012.12.043

Votsi, I., Limnios, N., Papadimitriou, E., Tsaklidis, G. 2019. Earthquake statistical analysis through multi-state modeling. Wiley, New Jersey.

Yip, CF., Ng, WL., Yau, CY. 2018. A hidden Markov model for earthquake prediction. *Stoch. Environ. Res. Risk. Assess.*, 32:1415-1434. https://doi.org/10.1007/s00477-017-1457-1