A Multivariate Non-Parametric Hazard Model for Earthquake Occurrences in Turkey

Nihal Ata* and Gamze Özel
Department of Statistics, Hacettepe University

Abstract: This paper provides an introduction to multivariate non-parametric hazard model for the occurrence of earthquakes since the hazard function defines the statistical distribution of inter-event times. The method is applied to the Turkish seismicity since a significant portion of Turkey is subject to frequent earthquakes and presents several advantages compared to other more traditional approaches. Destructive earthquakes from 1903 to 2009 between the longitudes of (39-42)N° and the latitudes of (26-45)E° are used. The paper demonstrates how seismicity and tectonics/physics parameters that can potentially influence the spatio-temporal variability of earthquakes and presents several advantages compared to more traditional approaches.

Key words: Earthquake catalogue, proportional hazard model, survival analysis, Turkish seismicity.

1. Introduction

The statistical modelling of the spatio-temporal distribution of earthquakes is an indispensable tool for extracting information from the available data on the physics of the earthquake occurrence process, and for making reliable earthquake forecasts. There are several statistical methods that can reveal the precursory seismic activity based on earthquake catalogues (Vere-Jones, 1970; Shimazaki and Nakata, 1980; Nishenko, 1985; Boschi Gasperini and Mulargia, 1995; Ellsworth *et al.*, 1998; Ogata, 1998; Kagan and Jackson, 2000; Stock and Smith, 2002; Posadas *et al.*, 2002), but the results obtained seem, in some cases, contradictory. Apparently, the factors that contribute mainly to such differences are the magnitude threshold and the spatial scale investigated. A spatio-temporal clustering for mainshock-aftershock sequences is widely accepted, but it has been suggested that the statistical distribution of large earthquakes might be different (Faenza *et al.*, 2003). The distribution of large earthquakes is studied only in

---

*Corresponding author.

the temporal domain, by selecting small seismic areas where the hypothesis of spatially homogeneous sampling probably holds. Unfortunately, these small areas do not usually contain a sufficient number of earthquakes to test adequately any hypothesis and model (Jackson and Kagan, 1993). For this reason, quite different distributions, such as Poisson (Kagan and Jackson, 1994), Poisson generalized (Kagan, 1991), Brownian passage time (Ellsworth *et al.*, 1998), Weibull (Nishenko, 1985), lognormal (Nishenko and Buland, 1987; Michael and Jones, 1998) distributions, or competitive models, such as general seismic gap hypothesis (McCann *et al.*, 1979), time-predictable model (Shimazaki and Nakata, 1980; Papazachos, 1992), clustering of earthquakes (Kagan and Jackson, 2000), are still commonly used.

Several studies have modelled earthquakes in Turkey as a Poisson process (Özmen and Kocaefe, 1999; Kasap and Gürlen, 2003; Kalyoncuoğlu, 2007). In those, the time between earthquakes is assumed to be exponentially distributed. This assumes that the probability of observing an earthquake at any given time is independent of both the elapsed time since the last earthquake and its severity.

The technique described here based on survival analysis and may account for any tectonics/physics factor that can potentially influence the distribution of the events, and test statistically their importance. This is accomplished by attaching at each inter-event time between earthquakes a vector of covariates, which describes the seismic events. The technique is robust because it considers all of the spatially inhomogeneous data simultaneously to build the statistical model; moreover, it allows use of the censoring times that carry out a certain amount of information on the process of earthquake occurrence and that are not usually considered in traditional approaches. At the same time, the technique is very flexible, being based on a small number of mild assumptions and does not require the knowledge of distribution (Faenza *et al.*, 2003).

The aim of this paper is to give some insight on the spatio-temporal distribution of large earthquakes in Turkey, including 111 destructive earthquakes having magnitude M ≥ 5.0 between years 1903-2009. A multivariate non-parametric hazard model is used to present several advantages compared to traditional approaches and this model has not been used before for the earthquake analysis in Turkey. This paper is organized as follows. In Section 2, survival analysis and its key properties are introduced. In Section 3, sample data of the Turkish seismicity, determination of tectonics/physics parameters are included and the results obtained applying the technique to the M ≥ 5.0 earthquakes that occurred in past century are reported. Finally, in Section 4, some future research perspectives are discussed.

2. Survival Analysis

2.1 Introduction

Survival analysis is a class of statistical methods for studying the occurrence and timing of events in both social and natural sciences. Survival time is defined as the time to the occurrence of a given event. This event can be the death, development of a disease, response to a treatment, equipment failures, births, marriages, divorces, promotions, retirements, earthquakes and so forth. Let T represents the survival time being in a given state or the time between two events. When subjects have not experienced the event of interest at the end of the study, the exact survival times of these subjects are unknown and these are called censored observations or censored times. In this study, the occurrence of an earthquake is defined as an event. The non-occurrence of an earthquake by the end of the study is referred to a censored observation, i.e. the most recent earthquake in each neotectonic zone represents a censored observation. Such data observations can be modeled as censored data and analyzed by survival models.

The distribution of survival time is characterized by probability density function $f(t)$, hazard function $h(t)$ or survival function $S(t)$. $S(t)$ gives the probability that failure will occur after time t and is given by

$$S(t) = P(T > t) = \int_t^\infty f(x)dx, \qquad 0 < t < \infty.$$

Many parametric (exponential, Weibull, log-logistic, log-normal, etc.) and non-parametric (Kaplan and Meier, 1958) approximations are used to estimate $S(t)$. Kaplan-Meier estimator is more often used and given by

$$\hat{S}(t) = \prod_{j=1} \left(\frac{n_j - d_j}{n_j} \right),$$

where $t_1 < \cdots < t_k$, $j = 1, \cdots, k$ is the ordered failure time; n_j is the number of individuals still at risk at time t_j, $n_j - d_j$ is the number of individuals surviving longer than t_j.

Hazard function is defined by

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t / T \geq t)}{\Delta t},$$

for $t > 0$ and represents the probability that an individual alive at t experiences the event in the next period Δt.

2.2 Proportional Hazard Model

The well-known model within survival models is proportional hazard model (PHM) which used to investigate the relation between the survival time and some covariates. Although it is based on proportional hazard assumption, no particular form of the probability distribution is assumed for the survival times. Therefore it is called as a semi-parametric model.

The data based on a sample of size n, consists of (t_i, d_i, x_i), $i = 1, \cdots, n$, where t_i is the time on study for the ith individual, d_i is the event indicator ($d_i = 1$ if the event has occurred and $d_i = 0$ if the lifetime is censored) and x_i is the vector of covariates for the ith individual. Hazard function for PHM is given by

$$h_i(t) = h_0(t) \exp(\boldsymbol{\beta}' \mathbf{x}_i),$$

where $h_0(t)$ is the baseline hazard function and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters.

The ordered earthquake occurrence times are denoted by and the set of places which are at risk at time t_j are denoted by $R(t_j)$, so that $R(t_j)$ is the set of places which are uncensored at a time just prior to t_j. Then, the likelihood for PHM is given by

$$L(\beta) = \prod_{j=1}^{k} \frac{\exp(\boldsymbol{\beta}' \mathbf{x}_j)}{\sum_{\ell \in R(t_j)} \exp(\boldsymbol{\beta}' \mathbf{x}_\ell)},$$

where $\mathbf{x}_j$ is the vector of covariates for the place where the earthquake occurs at the jth ordered earthquake occurrence time, t_j.

Survival times are usually recorded to the nearest day, month or year and so tied survival times can arise. In order to accommodate the tied observations, Breslow (1974) proposed an approximation to the likelihood function,

$$L(\beta) = \prod_{j=1}^{k} \frac{\exp(\boldsymbol{\beta}' \mathbf{s}_j)}{\left[\sum_{\ell \in R(t_j)} \exp(\boldsymbol{\beta}' \mathbf{x}_\ell) \right]^{d_j}}. \tag{2.1}$$

In (2.1), the $\mathbf{s}_j$ is the vector of sums of each of the p covariates for those individuals who die at the jth earthquake occurrence, $t_{(j)}$, $j = 1, \cdots, r$. If there are d_j earthquake occurrence at $t_{(j)}$, the hth element of s_j is $s_{hj} = \sum_{r=1}^{d_j} x_{hjr}$, where x_{hjr} is the value of the hth variable, $h = 1, \cdots, p$, for the rth of d_j individuals, $r = 1, \cdots, d_j$ who die at the jth earthquake occurrence time, $j = 1, \cdots, k$ (Collett, 2003).

Parameter estimation is obtained by the Newton-Raphson procedure in PHM. Let $\mathbf{u}(\beta)$ be the $p \times 1$ vector of first derivatives of the log-likelihood function with respect to the β parameters. Let $\mathbf{I}(\beta)$ be the $p \times p$ matrix of negative second

derivatives of the log-likelihood, so that the $(j, r)^{th}$ element of $\mathbf{I}(\beta)$ is

$$-\frac{\partial^2 \log L(\beta)}{\partial \beta_j \partial \beta_r}. \tag{2.2}$$

An estimate of the vector of β parameters at the $(s + 1)^{th}$ cycle of the iterative procedure, $\hat{\beta}_{s+1}$, is

$$\hat{\beta}_{s+1} = \hat{\beta}_s + \mathbf{I}^{-1}(\hat{\beta}_s)\mathbf{u}(\hat{\beta}_s), \tag{2.3}$$

for $s = 0, 1, 2, \cdots$, where $\mathbf{u}(\hat{\beta}_s)$ is the vector of efficient scores and $\mathbf{I}^{-1}(\hat{\beta}_s)$ is the inverse of the information matrix, both evaluated at $\hat{\beta}_s$. The process is started by taking $\hat{\boldsymbol{\beta}}_0 = \mathbf{0}$ and terminated when the change in the log-likelihood function is sufficiently small, or when the largest of the relative changes in the values of the parameter estimates is sufficiently small. When the iterative procedure is converged, the variance-covariance matrix of the parameter estimates is approximated by the inverse of the information matrix, evaluated at $\hat{\beta}$, that is, $\mathbf{I}^{-1}(\hat{\beta})$.

If the number of covariates is relatively large, the number of possible models that need to be fitted is computationally expensive. In this situation, automatic routines based on forward selection, backward elimination or stepwise procedures for variable selection are used.

In survival analysis, comparisons between a number of possible model is made on AIC (Akaike's information criteria), $\text{AIC} = -2 \log \hat{L} + \alpha q$ where q is the unknown β parameters in the model and α is a predetermined constant. The values of α between 2 and 6 can be used in computing the value of statistic. The choice for $\alpha = 3$ is equivalent to using a 5% significance level in judging the difference between the values of $-2 \log L$ for the two nested models which differ by between one and three parameters.

3. Empirical Study

3.1 Data Set and Tectonics/Physics Parameters

Turkey and its surrounding area are known as an excellent natural laboratory to study postcollisional intracontinental convergence and tectonic escape-related deformation, and the consequent structures that include fold and thrust belts, suture zones, active strike-slip faulting, and active normal faulting and the associated basin formation (Kalyoncuoğlu, 2007). It is located on a highly active Eurasian Plate which has caused numerous large scale earthquakes throughout history. A significant portion of Turkey is subject to frequent earthquakes, most significantly from the North Anatolian Fault Zone (NAFZ), which stretches across the country and is responsible for many of Turkey's largest historical earthquakes (Şengör et al., 2004). Other main sources of seismic activity in Turkey

are, East Anatolian Fault Zone (EAFZ) and West Anatolian Graben Complex (Erdik, 2001). To better understand the neotectonic features and active tectonics of Turkey, the simplified tectonic map of Turkey is given in Figure 1.
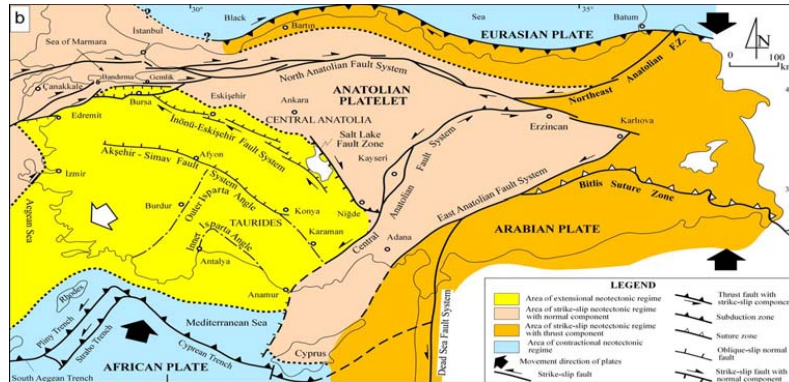

Figure 1: Simplified map showing the neotectonic subdivision of Turkey and adjacent areas (Koçyiğit and Özacar, 2003)

Turkish earthquake catalogue is examined in this study and the investigated area is presented in Figure 2. This data are obtained from an earthquake monitor which is based on real-time earthquake list of Kandilli Observatory for Turkey. In this area, 111 earthquakes with magnitude (M) 5 or larger have occurred in the past 106 years. The examined area is located between between the longitudes of 36°42N and the latitudes of 26°45E.

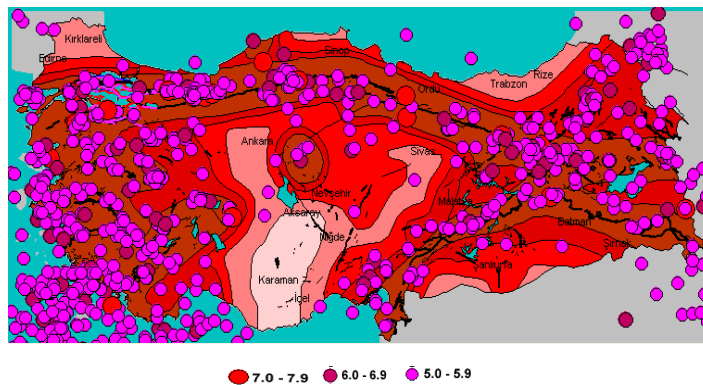Figure 2: The area of investigation (Source: http://www.koeri.boun.edu.tr)

The Ministry of Reconstruction and Settlement Institute published seismic zoning map of Turkey in 1996 based on maximum acceleration. This map is given in Figure 3.
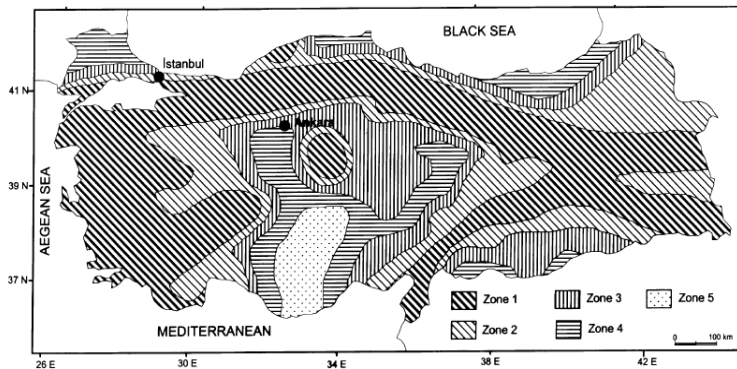
Figure 3: Turkey earthquake hazard map (Ministry of Reconstruction and Settlement, 1996)

As seen in Figure 3, the whole country is divided into the five zones. Zone I has the highest level of seismic hazard and the majority of Turkey is in Zone I. Turkey's economy is becoming more dependent on industry in major cities and these major cities are in Zone II.

Survival model for Turkish earthquake catalogue would be useful to predict earthquake probabilities. In this study, PHM is applied to destructive earthquakes in Turkey. Seismic hazard analysis requires assessment of earthquake source parameters in order to estimate the seismic source potential. Our model deals with two kinds of random variable, namely the inter-event time (IET) between two consecutive earthquakes, and the censoring time (CT), i.e. the time interval between the present time and the last time earthquake occurred. At each one of these random variables a vector of covariates is attached that carries out any kind of information relative to the IET or CT considered.

In this study, at each inter-event time (IET) and censoring time (CT), a covariate vector is attached with components: neotectonic subdivision of Turkey (N), transformed energy of earthquake (TE), depth of earthquake (D), number of foreshocks (FS), number of aftershocks (AS), magnitude of the earthquake (M), Mercalli scale of intensity (I), calculation of acceleration values that will effect the construction (AV) and fault system (FS). It is now worthwhile to overview these covariates that are supposed to be statistically significant.

- **_Neotectonic Subdivision (N)_**: As seen in Figure 1, Turkey and adjacent areas are divided into four main neotectonic domains: area of extensional neotectonic regime, area of strike-slip neotectonic regime with normal component, area of strike-slip neotectonic regime with thrust component and area of contractional neotectonic regime.

- **_Transformed Energy (TE)_**: The energy of an event is chosen as measure of its strength. The formula of Gutenberg (1956) is used to calculate the energy of an earthquake (Jarkov, 1983)

$$\log E_s = 11.8 + 1.5M.$$

  To reduce, the very broad variability range of the energy released in an earthquake, this quantity is transformed through the formula

$$E_i^{\mathrm{tr}} = \frac{(E_i - E_{\min})}{(E_{\max} - E_{\min})}, \quad E \in [0.1].$$

  where $E_{\max}$ and $E_{\min}$ are respectively the upper and lower bound of the observed values, and E is the energy of the ith event (Gospodinov and Rotondi, 2001)

- **_Depth (D)_**: Earthquakes occurring at a depth of less than 70 km are classified as 'shallow-focus' earthquakes, while those with a focal-depth between 70 and 300 km are commonly termed 'mid-focus' or 'intermediate-depth' earthquakes. In fact, the great majority of earthquake in Turkey are shallow. Shallow earthquakes often tend to cause serious damage. That is, the deeper an earthquake is, the less effect it has in the surface of the Earth.

- **_Foreshock (FS) and Aftershock (AS)_**: To detect the significant seismic quiescence, it is necessary to decluster the earthquake catalogue. For this process, it has been made use of the algorithm introduced by Reasenberg (1985). It removes all the dependent events from each cluster, and substitutes them with a unique event. The foreshocks are searched with $M \geq 3.0$ within a distance of 50 km and five days from each earthquake with $M \geq 5.0$. The aftershocks of each earthquake with magnitude $M \geq 4.0$ within one month and shallower than 40 km are used.

- **_Magnitude (M)_**: The great majority of earthquake in Turkey are shallow and surface wave magnitude is generally used for shallow earthquakes. Surface wave magnitude is estimated using the formula $M = \log(A/t) + 1.66 \log \Delta + 3.3$, where A is the maximum amplitude (in micrometers) of the Rayleigh waves, T is the period (usually about 20 seconds) and Δ is the distance (in degrees).

- **_Mercalli Scale of Intensity (I)_**: Intensity measures the strength of shaking produced by the earthquake at a certain location. The Mercalli intensity scale is a scale used for measuring the intensity of an earthquake. The scale quantifies the effects of an earthquake on the Earth's surface, humans, objects of nature, and man-made structures on a scale of I through XII, with I denoting not felt, and XII total destruction.

- **Acceleration Value (AV)**: Earthquake zones of Turkey classified in Figure 3 as 1st degree earthquake zone (more than 0.4g), 2nd degree earthquake zone (between 0.3g - 0.4g), 3rd degree earthquake zone (between 0.2g - 0.3g), 4th degree earthquake zone (between 0.2g - 0.1g) and 5th degree earthquake zone (less than 0.1g), where g is gravity (981 cm/s*s).

- **Fault System (FS)**: Fault is a planar or gently curved fracture in the Earth's crust across which there has been relative displacement. There are six kinds of fault systems in Turkey that are strike-slip fault, thrust fault with strike-slip component, subduction zone, suture zone, oblique-slip normal fault, strike-slip fault with normal component as seen in Figure 1. Since subduction zone did not produce earthquakes with $M \geq 5.0$ from 1903 to 2009, five fault zones are used.

3.2 The Results of Survival Analysis

A probabilistic seismic hazard and methodology is applied to Turkish seismicity in order to determine earthquake occurrences. The collective median survival time for all earthquakes is estimated at 1.40 years. Considering each neotectonic subgroup, median survival time and the number of earthquakes are presented in Table 1.

Table 1: The median survival time for each neotectonic zone

	Zone I	Zone II	Zone III	Zone IV
Median Survival Time (year)	1.40	0.8	3.08	24.80
Number of Events	39	44	23	1

Table 1 shows that most earthquakes have been occurred in Zone I and II. It is also found that Zone IV had the longest median survival (24.80 years), while Zone II had the shortest survival. In the non-parametric approach to seismicity, the estimates of Kaplan-Meier survival probability are presented in Table 2.

As shown in Table 2, the risk of the earthquake occurrence after the preceding one is 0.29 in six months, 0.42 in one year, 0.83 in five years and 0.99 in 25 years, respectively. Kaplan-Meier method provides useful graphical presentation of survival distribution as well as estimates of survival probabilities. Kaplan-Meier estimate of the survival function is plotted in Figure 4.

Table 2: Kaplan-Meier estimations of survival probability

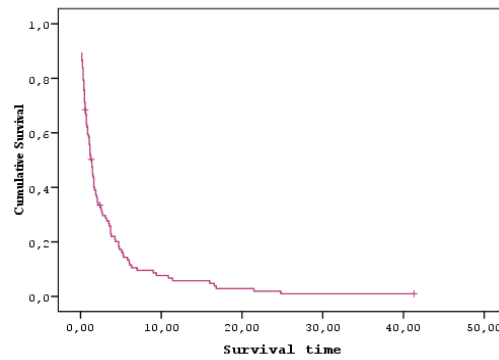| Time Interval (year)	Survival Probability ± Std.Error	Time Interval (year)	Survival Probability ±Std.Error
0.0-0.5	0.7117 ± 0.0430	5.0-6.0	0.1339 ± 0.0331
0.5-1.0	0.5843 ± 0.0469	6.0-7.0	0.1052 ± 0.0299
1.0-1.5	0.4463 ± 0.0475	7.0-9.0	0.0861 ± 0.0259
1.5-2.0	0.3626 ± 0.0460	9.0-12.0	0.0574 ± 0.0227
2.0-3.0	0.2965 ± 0.0439	12.0-17.0	0.0287 ± 0.0163
3.0-4.0	0.2200 ± 0.0401	17.0-25.0	0.0096 ± 0.0095
4.0-5.0	0.1721 ± 0.0366		

Figure 4: Kaplan-Meier curve

Considering the four neotectonic zones, the estimates of Kaplan-Meier survival probability are summarized in Table 3. As expected, Zone II shows the highest recurrence probability for an earthquake within six months. Since the major industry centers are located in Zone II, this result is important for Turkey. The recurrence probability of an earthquake within 1.5 years is 0.5593 in Zone I; 0.6681 in Zone II; 0.375 in Zone III, respectively. On the other hand, the survival probability of Zone IV differs from other zones because of different seismic characteristics. An earthquake will be occurred in twenty five years with the probability 0.5.

Kaplan-Meier curves for each zone in Figure 5 also indicate that Zone IV has a better survival curve than other zones. Moreover, as the number of years increases, the curves appear to get farther apart. However, the weakness of this approach is that it does not provide a comparison of the total survival experience of the four groups, but rather gives a comparison at some arbitrary time point (s). For this reason, the survival functions of the neotectonic zones are also

compared statistically using the log-rank test. It has the considerable advantage that it does not require us to know anything about the shape of the survival curve or the distribution of survival times. The result of log-rank test is found as $\chi^2 = 13.497$ (p-value $= 0.004$) and so that the equality of the survival functions for neotectonic zones rejected at the 95% confidence level.

Table 3: Kaplan-Meier estimations of survival probability for neotectonic zones

Neotectonic Zone I		Neotectonic Zone II	
Time interval (year)	Survival Probability $\pm$ Std.Error	Time interval (year)	Survival Probability $\pm$ Std.Error
0.0-0.5	0.7250 ± 0.0706	0.0-0.5	0.6222 ± 0.0723
0.5-1.0	0.6481 ± 0.0758	0.5-1.0	0.4667 ± 0.0744
1.0-1.5	0.4407 ± 0.0795	1.0-1.5	0.3319 ± 0.0704
1.5-2.0	0.3111 ± 0.0743	1.5-2.0	0.2607 ± 0.0662
2.0-3.0	0.2333 ± 0.0679	2.0-3.0	0.1896 ± 0.0595
3.0-4.0	0.1556 ± 0.0583	3.0-4.0	0.1422 ± 0.0533
4.0-5.0	0.1296 ± 0.0540	4.0-7.0	0.0711 ± 0.0394
5.0-6.0	0.0519 ± 0.0357	7.0-9.0	0.0474 ± 0.0326
6.0-10.0	0.0259 ± 0.0256	9.0-16.0	0.0237 ± 0.0234
10.0-22.0	0.0000 ± 0.0000	16.0-17.0	0.0000 ± 0.0000
Neotectonic Zone III		Neotectonic Zone IV	
Time interval (year)	Survival Probability $\pm$ Std.Error	Time interval (year)	Survival Probability $\pm$ Std.Error
0.0-0.5	0.8333 ± 0.0761	0.0-25.0	0.5 ± 0.3536
0.5-1.0	0.6667 ± 0.0962	25.0-42.0	0.0000 ± 0.0000
1.0-1.5	0.6250 ± 0.0988		
1.5-2.0	0.5833 ± 0.1006		
2.0-3.0	0.5417 ± 0.1017		
3.0-4.0	0.4062 ± 0.1020		
4.0-5.0	0.2257 ± 0.0880		
5.0-6.0	0.1806 ± 0.0811		
6.0-7.0	0.1354 ± 0.0723		
7.0-10.0	0.0903 ± 0.0607		
10.0-12.0	0.0451 ± 0.0440		
12.0-17.0	0.0000 ± 0.0000		

PHM with stepwise variable selection is used to determine the predictive use of seismological and geological variables for the earthquake occurrences in Turkey. This study defines duration variable as, namely inter-event time (IET), between two consecutive earthquakes. Neotectonic subdivision (N), transformed energy (TE), depth (D), number of foreshocks (FS) and aftershocks (AS), magnitude
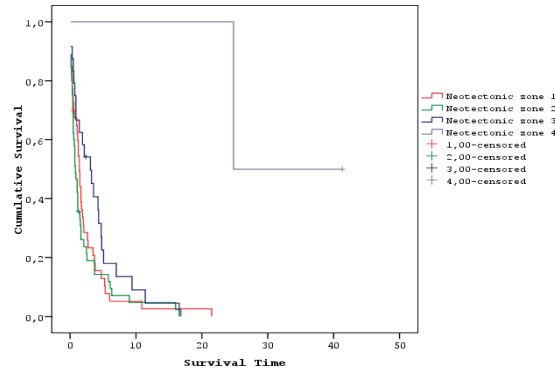
Figure 5: The graph of cumulative survival probabilities for neotectonic zones

(M), Mercalli scale of intensity (I), acceleration values (AV) and fault system (FS) are potential covariates. Neotectonic subdivision is coded with area of extensional neotectonic regime as the reference category. Mercalli scale of intensity (I) is entered as a discrete variable with reference category intensity I. A reference category is choosen as 1st earthquake zone for acceleration values (AV). Similarly, fault system is coded with strike-slip fault as the reference category. Transformed energy (TE), depth (D), number of foreshocks (FS) and aftershocks (AS), magnitude (M) are entered as continuous variables. The results from fitting PHM on Turkish seismicity are given in Table 4.

Table 4: Partial likelihood inference on Turkish seismicity data from PHM

Variable	Parameter Estimate	Std. Error	Chi-Sq.	Pr > Chi-Sq	Hazard Ratio	95% Hazard Ratio Confidence Interval Lower−Upper
Final Model for Turkish Seismicity Data						
Magnitude	-0.68817	0.17958	14.6855	0.0001	0.502	$0.353 - 0.714$
Transformed Energy	3.19736	0.99713	10.2821	0.0013	24.468	$3.466 - 172.723$
Model Fit Statistics						
	-2logL	671.192	AIC	675.192		
Testing Global Null Hypothesis: Beta = 0						
	Likelihood Ratio		Wald Statistics			
	Chi-Square	Pr > Chi-Sq.	Chi-Square	Pr > Chi-Sq.		
	14.8810	0.0006	16.2952	0.0003		

As seen in Table 4, PHM is statistically significant at 95% confidence level. By using equations (2.1)-(2.3), we obtain $\beta_{M} = -0.69 \pm 0.17$ and $\beta_{TE} = 3.2 \pm 0.1$. The results show that the models containing magnitude (M) and transformed energy (TE) are found as important covariates which effect the occurrence of earthquakes whereas others are not significant at a 95% confidence level.

The estimated hazard of magnitude is $\exp(-0.68) \doteq 0.50$; that is, for a 1-unit decrease in magnitude, the risk for occurrence of earthquake increases a 0.5-unit. For transformed energy, the value of $\exp(3.20)$ corresponds to the change in hazard for an increase of a unit in the log scale. Thus, the estimated hazard increases 24.5 times if the transformed energy is higher by a factor of 10.

4. Discussion

United Nations Development Programme (UNDP) announces Turkey as the third country after Iran and Yemen according to the number of deaths as a result of earthquakes. Turkey is located on a highly active Eurasian Geological Plate and there have been 111 earthquakes with magnitudes 5.0 or greater. Sixteen earthquakes with casualties more than 1.000 have occurred since 1903. For this reason, a research on earthquake occurrences in Turkey can be informative for understanding the seismicity of the investigated area.

In this paper, a statistical analysis of the spatio-temporal distribution of Turkish seismicity is performed. A nonparametric multivariate statistical model is used for the first time that is able to account for seismological/geological parameters simultaneously. The method has been applied to Turkey by using a regionalization based on tectonic parameters for the period 1903-2009. Kaplan-Meier estimations provide that neotectonic Zones I and II are more risky relative to neotectonic Zones III and IV. The majority of Turkey is located in Zone I and Zone II is one of Turkey's most industrialized regions, home to much of Turkey's heavy industry, including petrochemical plants, car manufacturers, tire companies, paper mills, steel fabrication plants, cement plants and pharmaceutical firms. The results of PHM suggest that magnitude and transformed energy of earthquake are the important factors that influence the occurrence of earthquake.

Acknowledgements

References

Boschi, E., Gasperini, P. and Mulargia, F. (1995). Forecasting where larger crustal earthquakes are likely to occur in Italy in the near future. *Bulletin Seismological Society of America* **85**, 1475-1482.

Boğaziçi University Kandilli Observatory and Earthquake Research Institute, http://www.koeri.boun.edu.tr (accessed on 15 March, 2010).

Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics* **30**, 89-99.

Collett, D. (2003). *Modelling Survival Data in Medical Research*. Chapman and Hall, New York.

Earthquake Research Institute, http://www.deprem.gov.tr (accessed on 25 March, 2010).

Ellsworth, W. L., Matthews, M. V., Nadeau, R. M., Nishenko, S. P., Reasenberg, P. A. and Simpson, R.W. (1998). A physically-based earthquake recurrence model for estimation of long-term earthquake probabilities, in *Proceedings of the Second Joint Meeting of the UJNR panel on Earthquake Research*, 135-149.

Erdik, M. (2001). Report on 1999 Kocaeli and Düzce (Turkey) earthquakes, Kandilli Obsarvotary and Earthquake Research Institute, Turkey. http://www.koeri.boun.edu.tr/depremmuh/eski/eqspecials/kocaeli/kocaelireport.pdf

Faenza, L., Marzocchi, W. and Boschi, E. (2003). A non-parametric hazard model to characterize the spatio-temporal occurrence of large earthquakes: An application to the Italian catalogue. *Geophysical Journal International* **155**, 521-531.

Faenza, L., Marozcchi, W., Lombardi, A. M. and Console, R. (2004). Some insights into the time clustering of large earthquakes in Italy. *Annals of Geophysics* **47**, 1635-1640.

Ghasemi, H., Cooper, J. D., Imbsen, R., Piskin, H., Inal, F. and Tiras, A. (2000). *The November 1999 Duzce Earthquake*: *Post-Earthquake Investigation of the Structures on the TEM*. Report No. HWA-RD-00-146, McLean, VA.

Gospodinov D. and Rotondi R. (2001). Exploratory analysis of marked poisson processes applied to balkan eartquake sequences. *Journal of the Balkan Geophysical Society* **4**, 61-68.

Jackson, D. D. and Kagan, Y. Y. (1993). Reply to comment: Seismic gap hypothesis: Ten years after. *Journal of Geophysical Research* **98**, 9917-9920.

Kagan, Y. Y. (1991). Likelihood analysis of earthquake catalogues. *Geophysical Journal International* **106**, 135-148.

Kagan, Y. Y. and Jackson, D. D. (1994). Long-term probabilistic forecasting of earthquakes. *Journal of Geophysical Research* **99**, 13685-13700.

Kagan, Y. Y. and Jackson, D. D. (2000). Probabilistic forecasting of earthquakes. *Geophysical Journal International* **143**, 438-453.

Kalyoncuoglu, Y. (2007). Evaluation of seismicity and seismic hazard parameters in Turkey and surrounding area using a new approach to the Gutenberg-Richter relation. *Journal of Seismology* **11**, 131-148.

Kandilli Observatory and Earthquake Research Institute of University of Bogazici, Istanbul, Turkey, http://www.koeri.boun.edu.tr (accessed on 10 January, 2010).

McCann, W. R., Nishenko, S. P., Sykes, L. R. and Krause, J. (1979). Seismic gaps and plate tectonics: seismic potential for major boundaries. *Pure and Applied Geophysics* **117**, 1082-1147.

Michael, A. J. and Jones, L. M. (1998). Seismicity alert probabilities at Parkfield, California. *Bulletin of the Seismological Society of America* **88**, 117-130.

Ministry of Reconstruction and Settlement (1996). *Seismic Hazard Map of Turkey*.

Nishenko, S. P. (1985). Seismic potential for large and great interplate earthquakes along the Chilean and Southern Peruvian margins of South America: A quantitative reappraisal. *Journal of Geophysical Research* **90**, 3589-3615.

Nishenko, S. P. and Buland, R. A. (1987). A generic recurrence interval distribution for earthquake forecasting. *Bulletin of Seismological Society of America* **77**, 1382-1399.

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

Ozbey, C., Sari, A., Manuel, L., Erdik, M. and Fahjan, Y. (2004). An empirical attenuation relationship for northwestern Turkey ground motion using a random effects approach. *Soil Dynamics and Earthquake Engineering* **24**, 115-125.

Özmen, B. and Kocaefe, S. (1999). The seismicity and earthquake hazard of Ankara. *International Workshop on Recent Earthquakes and Disaster Prevention Management*, ODTU, Ankara.

Papazachos, B. C. (1992). A time and magnitude predictable model for generation of shallow earthquakes in the aegean area. *Pure and Applied Geophysics* **138**, 287-308.

Shimazaki, K. and Nakata, T. (1980). Time-predictable recurrence model for large earthquakes. *Geophysical Research Letters* **7**, 279-282.

Şengör, A. M. C., Tüysüz, O., İmren, C., Sakınç, M., Eyidoğan, H., Görür, N., Le Pichon, X. and Rangin, C. (2004). The North Anatolian fault: A new look. *Annual Review Earth and Planetary Sciences* **33**, 1-75.

Stock, C. and Smith, E. G. C. (2002). Adaptative kernel estimation and continuous probability representation of historical earthquake catalogs. *Bulletin of Seismological Society of America* **92**, 904-912.

U. S. Geological Survey (USGS), http://www.usgs.gov (accessed on 2 June, 2010).

Vere-Jones, D. (1970). Stochastic models for earthquake occurrence. *Journal of Royal Statistical Society, Series B* **32**, 1-62.

Nihal Ata
Department of Statistics
Faculty of Science Hacettepe University
Ankara, Turkey
nihalata@hacettepe.edu.tr

Gamze Özel
Department of Statistics
Faculty of Science Hacettepe University
Ankara, Turkey
gamzeozl@hacettepe.edu.tr