GU J Sci 30(4): 584-597 (2017)

Gazi University

Journal of Science

http://dergipark.gov.tr/gujs

The Statistical Analysis of the Earthquake Hazard for Turkey by Generalized Linear Models

Emel KIZILOK KARA[1], Kübra DURUKAN[2,*]

[1]*Department of Actuarial Science, Faculty of Arts and Sciences, Kirikkale University, Yahşihan, Kırıkkale, TURKEY*

[2]*Department of Statistics, Faculty of Arts and Sciences, Kirikkale University, Yahşihan, Kırıkkale, TURKEY*

Article Info

Received: 27/04/2017
Accepted: 30/05/2017

Keywords

Earthquake, magnitude-frequency, Gutenberg-Richter relationship, Generalized Poisson Regression Model, Generalized Negative Binomial Regression Model, dispersion parameter, Turkey

Abstract

In this paper, 4863 earthquake data of magnitude 4.0 and greater from 1900 to 2014 are statistically analyzed for the earthquake hazard in Turkey. The magnitude-frequency relationship in earthquake risk analysis is often performed by Gutenberg-Richter model. With the use of this model, information about earthquake potential of any region can be obtained by previous data and by estimating parameters such as return periods and possibilities of their occurrence. In this study, the relationship between earthquake numbers and magnitudes is modelled with the Generalized Linear Models as an alternative to Gutenberg-Richter model. Generalized Poisson Regression model and Generalized Negative Binomial Regression models as Generalized Linear Models are utilized in the study. Generalized Poisson Regression model is found as the best model when considering the dispersion parameters and model selection criteria. Exceeding probabilities and return periods are calculated for the selected years depending on yearly average occurrence number of earthquakes estimated with the Gutenberg-Richter and Generalized Poisson Regression models. According to the results, Generalized Poisson Regression model can be employed for seismic risk modelling in Turkey.

1. INTRODUCTION

Turkey, located on the Alpine-Himalayan belt, is one of the areas of high seismicity in the world. There are many active faults in Turkey, due to its complex geological structure and geodynamic situation. The active fault map of Turkey by Saroğlu et al. (1992) [1] is given in Figure 1.

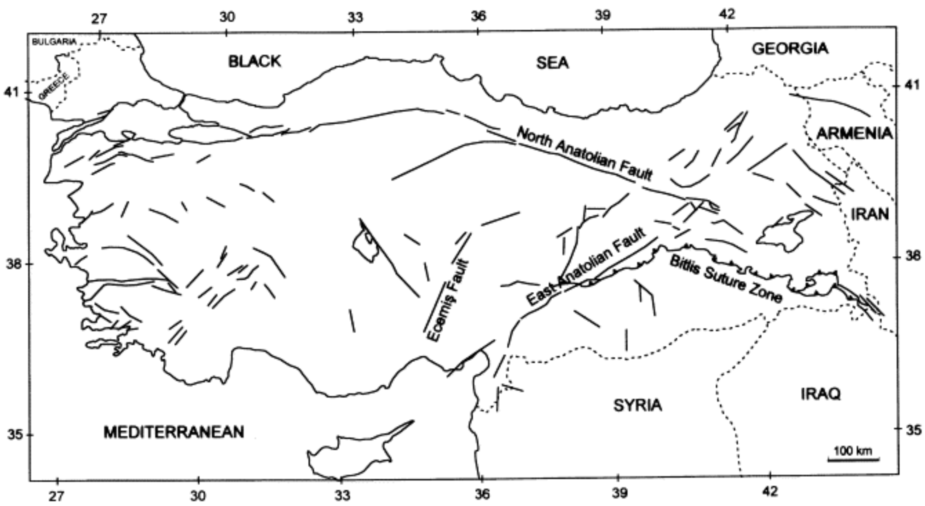

Figure 1. *The active fault map of Turkey*

*Corresponding author, kubraba@gmail.com

According to the active fault map of Turkey, the North Anatolian Fault (NAF) and East Anatolian Fault (EAF), as well the Eastern Anatolia, Marmara and Aegean regions are the highest earthquake risk areas in Turkey. The largest earthquakes (1939 Erzincan $Ms = 8.3$, 1942 Niksar- Erbaa $Ms = 6.9$, 1944 Bolu-Gerede $Ms = 7.5$, 1949 Karlıova $Ms = 7.9$, 1971 Bingöl $Ms = 6.8$, 1999 İzmit $Ms = 7.4$, 1999 Düzce $Ms = 7.2$, 2010 Elazığ $Ms = 6.0$ and 2011 Van $Ms = 7.2$) that have took place in Turkey during the last century occured over the NAF. These earthquakes have caused major damage such as loss of life and property. In this sense, especially in earthquake regions of high risk, the earthquake hazard analysis plays a significant role in minimizing damage.

The aim of the earthquake hazard analysis is to estimate earthquake occurrence probabilities and their recurrence periods by using the relationship between the magnitude and number (frequency) of the earthquakes. Here, one of the main problems is how to find the best suitable model that estimates seismic risk. The basic model is the Gutenberg Richter (GR) model that states the logarithm of the frequency is linearly dependent on the magnitude [2]. However, Firuzan (2008), [3] said that many authors have discussed the usage of GR model especially in the tails of the distribution (Dargahi-Noubary (1986), [4]; Main (1996), [5]). As said in the same paper, Dargahi-Noubary (1986), [4] and Kagan (1993), [6] suggest that more suitable statistical models should be used instead of GR model for the distribution with high magnitudes. For this purpose, Firuzan (2008), [3] give the comparative assessment of four advanced statistical distributions, to provide earthquake probabilites and return periods for all different zone regions in western Anatolia. Here, the statistical models compared are the Exponential (EXP), Extreme value distribution Type1 (Gumbel) (GUM), Log Pearson Type 3 (LP3), and Generalized Pareto (GP) models. The results indicate that the GP and GUM distributions are most appropriate for describing the peaks over the threshold earthquake series and annual maximum earthquake data in western Anatolia, respectively. Çobanoğlu et al. (2006), [7] compared the GR model with the Exponential, Gumbel and Poisson models to determine the magnitude and frequency relationship for the Denizli region. As a result, they found that the return periods obtained with the Poisson model are larger values than the other models.

On the other hand, there are other studies in the literature on earthquake hazard analysis using statistical models due to Turkey's high seismicity: Kalyoncuoğlu (2007) performed an evaluation of seismicity and earthquake hazard parameters using a new approach to the Gutenberg–Richter relation in Turkey and the surrounding area [8]. Bayrak et al. (2005) estimated some fundamental seismic parameters using the Gumbel III asymptotic distribution based on the GR model for different regions in Turkey [9]. Öztürk et al. (2008) calculated the most probable maximum magnitudes, the mean return periods (in years) and the probabilities for different time periods at given magnitudes by using the Gumbel's I asymptotic distribution in order to estimate the seismicity of the 24 seismic regions in Turkey [10]. Some of the approaches in the literature are concerned with the Markov chains [11, 12], stochastic models [13], reliability issues [14, 15], probability density and distribution functions [16], probabilistic assessment of earthquake insurance rates [17] and Poisson approachs [18]. Also, for earthquake hazard in several regions of Turkey and the Earth, the other methods relating to the assessment of the earthquake hazard parameters have been applied for the occurrence probabilities and return periods of earthquakes [19-26].
In classical regression models developed for GR, it is required that the dependent variable be normally distributed with constant variance and be a linear function of independent variables. Independent variables may be constant, categorical or the combination of both [27]. However, it is possible to encounter conditions where these assumptions cannot be provided in practice. Generalized Linear Models (GLMs) do not require continuity and normality assumptions [28]. Therefore, our purpose in this study is to use GLMs in order to statistically make analysis of the earthquake hazard, as an alternative to linear regression GR models.

The concept of GLMs was first developed by Nelder and Wedderburn (1972) [29]. During the years ahead, McCullagh and Nelder (1989), [27] and Dunteman and Ho (2006), [30] worked on GLMs. Additionally, there are GLMs applications in various areas such as actuary, insurance and engineering [28, 31, 32].

GLMs applications are significant in that the dependent variable is rare case. The number of earthquakes with a magnitude greater than a certain value may be considered as rare case. The objective of this study is to present an application of GLMs by using Generalized Poisson Regression (GPR) and Generalized Negative Binomial Regression (GNBR) models for the earthquake hazard analysis of Turkey. The relationship between the number of earthquakes (i.e., frequency) and magnitude are estimated by GPR and GNBR models. Earthquake risk parameters are calculated for GPR and GNBR models and the obtained results for selected model according to dispersion parameters and model selection criteria are compared by the traditional GR model.

As different from other studies, the purpose of present study is to use GLMs, which are more flexible in providing of normality and linerarity assumptions, to calculate earthquake probabilities and return periods. The use of this model also allows to choose the model fit according to the dispersion parameters. Therefore, we have proposed a different statistical approach to model earthquake magnitude-frequency relation with this study. On the other hand, we did not find in our literature reviews of studies conducted with the GLM, considering the dispersion parameter for earthquake data. However, it is said that the GLM gives more meaningful results than other known methods in engineering applications [32, 33]. In addition, in the literature reviews of studies, other statistical methods are generally used to estimate the earthquake probability and return periods in different regions of Turkey. Unlike the methods used in these studies, we used the GLMs to estimate earthquake probabilities and return periods. As a result, we have shown in this study that GPR model, as an alternative to GR model, can be used for Turkey earthquake data. One of the important results of our study is that the predicted return periods by the GPR model are larger than the estimated using the GR model for earthquake magnitudes smaller than 7. The normality and linearity assumptions may not be provided especially for regional earthquake studies since there will be less data. In this case, we think that the use of GLM models in regional earthquake analyzes can make an important contribution to the literature.

The rest of the paper is organized as follows. Section 2 consists methodology chapter that gives required information for the GR, GPR and GNBR models and the estimation of earthquake risk parameters. Also, AIC and BIC criteria used for the choice of the model according to dispersion conditions is described in the same section. Section 3 presents the results and discussion included an application of the earthquake hazard analysis in Turkey using GR, GPR and GNBR models. Finally, some conclusions and suggestions are contained in Section 4.

2. METHODOLOGY AND METHODS

In order to perform the earthquake risk analysis of Turkey (located between 36-42°N latitudes and 26-45°E longitudes) which is among the leading countries subject to earthquake risks, the earthquake data, whose magnitude values (M) are 4.0 or greater between 20.09.1900 and 20.07.2014 were gathered from the Bogazici University Kandilli Observatory and Earthquake Research Institute [34]. In this study, magnitude values of X_m type which have the greatest value in magnitude (M_D: Time dependent, M_L: Local, M_W:Moment, M_S:Surface wave, M_b:Body wave) are used for earthquake data. Turkey earthquake map containing this data are illustrated in Figure 2.

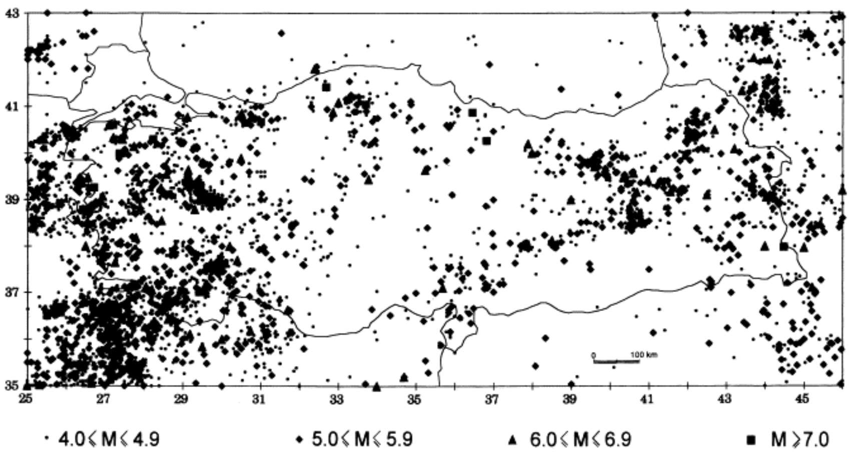

Figure 2. *Earthquake map for Turkey from 1900 to present*

This paper, which helps to perform the earthquake hazard analysis of Turkey, consists of three main stages with the following methodological steps.

(1) The determination of frequency distribution for frequency and magnitudes of earthquakes:

a. *Data:* 4863 earthquake data with magnitudes $4.0 \leq M \leq 7.9$ for 115 years from 1900 to 2014 are utilized.
b. *Frequency table:* The magnitudes with 0.1 class interval, number of earthquakes or frequency (n_i), cumulative frequency (N), cumulative frequency per year (N/t) and log transformation of (*N)* (i.e., $y = LogN$) are occurred.

(2) The performing of dispersion parameters and goodness-of fit test for models.

a. *Descriptive statistics and model selection for frequency:* Frequency (*N*) is taken as dependence variable and descriptive statistics of $LogN$ and LnN variables are calculated. Then Kolomogorov Simirnov goodness-of-fit test is used to show that the normal distribution is suitable to $LogN$ frequency data for GR model and the Poission distribution is suitable to LnN frequency data for GPR model.
b. *Model setup:* The relation between magnitude and frequency are occurred with GR, GPR and GNBR models.
c. *Model selection for the relation between magnitude and frequency:* Model parameters are estimated using SPSS 18.0 and the best model is selected with dispersion parameters and model selection crieteria (Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC))

(3) The analyzing of the earthquake hazard for selected models.

a. *Calculation of earthquake parameters for selected models:* a, b, a', a_1 and a'_1 parameters are calculated for selected models.
b. *Evaluation of the earthquake hazard analysis of Turkey:* Average earthquake number $n(M)$, recurrence probabilities $R(M)$ and return periods $Q(T)$ are estimated from selected models.

In here, the mentioned models (the GR model, the GPR model and GNBR model), the performing of the earthquake hazard analysis and model selection are defined in the following subsections:

2.1. Gutenberg–Richter Model

When any earthquake occurs, earthquake parameters such as focus centre (hypocentre) outer centre (epicentre), strength, magnitude etc. are used for the earthquake to describe and understand. Of these parameters, magnitude is described as a measure of energy emerged during the earthquake. In fact, as the magnitude of earthquake increases, the frequency decreases. On the other hand, it causes significant loss of life and property damages. Therefore, correct modelling of the magnitude-frequency relationship is important to take necessary precautions which eventually enable to reduce risk.

The GR model defined by Gutenberg and Richter in 1954 often used in the literature for the magnitude-frequency relationship in earthquake risk analysis is given by

$$LogN = a + bM \quad (1)$$

where M is the earthquake magnitude, N is the number of cumulative earthquakes whose magnitude is equal to M or greater in a given year, a and b are regression parameters [2].

Here, parameter a changes according to the size of area under examination, duration of the observation and earthquake effectiveness during the observation. Parameter b varies in response to the tectonic characteristics of study area. When parameter b value turns out to be high, it shows energy accumulation and vice versa [2, 35]. Parameters a and b in the model are calculated by least squares.

2.2. Generalized Linear Models

In general regression models, constant variance and normality for dependent variable are important assumptions. Otherwise, a data transformation needs to be performed for the dependent variable. In order to overcome the heteroscedasticity problem, the weighted smallest squares method can be utilized as an effective method. An alternative approach to data transformation is to use GLMs as suggested by Montgomery et al. (2010) [36]. In fact, they are a combination of linear and nonlinear regression models where the dependent variable is may not be normally distributed. It is assumed in GLM that the distribution of the dependent variable is a member of the exponential family including, for example, Normal, Binomial, Poisson, Exponential and Gamma distribution and it described in general form as

$$f(y_i, \theta_i, \emptyset) = exp\left\{\left[\frac{y_i\theta_i - b(\theta_i)}{a(\emptyset)}\right] + c(y_i, \emptyset)\right\} \tag{2}$$

where, $\emptyset\ is$ the dispersion parameter and $\theta_i\ is$ the location parameter.

While the relationship between expected value of y and the independent variable is described with $\eta_i = x_i{}'\beta$ in a normal model, it is described with a g link function in the form of $\eta_i = g[E(y_i)]$ in GLMs. Here, $E(y_i) = g^{-1}(x_i{}'\beta)$ can be obtained by inverse transformation [32].

The likelihood estimators of GLMs with a g link function can be obtained by numerical recursive algorithms such as Newton Raphson method using the following log-likelihood function [32].

$$LL = \sum_{i=1}^{n}\left\{\left[\frac{y_i\theta_i - b(\theta_i)}{a(\emptyset)}\right] + c(y_i, \emptyset)\right\} \tag{3}$$

2.2.1. Generalized Poisson Regression Models and Negative Binomial Regression Models

Normality assumption in modelling dependent variables for traffic accidents, number of people suffering from certain diseases and, number of earthquakes, floods and hurricanes is generally not valid. Therefore, in such cases Generalized Poisson Regression (GPR) models are suggested to be used [28]. The probability distribution of GPR with y_i dependent variable is described as

$$f(y) = \frac{e^{-\mu}\mu^y}{y!},\ y = 0,1, \dots \tag{4}$$

provided that the Poisson Regression model is defined by

$$y_i = E(y_i) + \varepsilon_i\ , i = 0,1, \dots, n \tag{5}$$

The relationship between μ and η_i is given in the form of $\mu_i = g^{-1}(\eta_i)$ with a proper g link function. Here, the mean is equal to expected value of y_i (i.e., $\mu = E(y_i)$), and the link function is a linear dependent function of *x* (i.e., $g(\mu_i) = \eta_i$). The commonly used link functions for the Poisson distribution are the identical link function ($g(\mu_i) = \mu_i$) and the log link function ($g(\mu_i) = Ln(\mu_i)$). Hence, the relationship between μ and η_i is obtained by using the identical and the log link functions in the form of $x_i{}'\beta$ and $e^{x_i{}'\beta}$, respectively [32].

In this model, the mean and the variance of the dependent variable are theoretically assumed to be equal to each other. However, this assumption cannot be true in practice and sometimes overdispersion or underdispersion problems can occur [37, 38]. As known that if the variance is higher than mean overdispersion takes place and vice versa [37]. If dispersion exists it is appropriate to use one of GPR or

GNBR models where a dispersion parameter is added to the model [39, 40]. GPR models are preferred for overdispersion or underdispersion, while GNBR models are employed for overdispersion [41].

Overdispersion for Poisson approach can be determined by various methods. For example, $\emptyset$ dispersion parameter is equal to division of Pearson χ^2 or deviation G^2statistics by the model degree of freedom (i.e., $\frac{\chi^2}{n-p}$ or $\frac{G^2}{n-p}$). χ^2 and G^2 statistics are given for dependent variable y with estimator of mean $\hat{\mu}$ and estimator of variance $\hat{\sigma}^2$, by respectively $\chi^2 = \sum_{i=1}^{n}\left(\frac{y_i-\hat{\mu}_i}{\hat{\sigma}_i}\right)^2$ and $G^2 = \sum_{i=1}^{n} y_i Ln\left(\frac{y_i}{\hat{\mu}_i}\right)$, where, $\hat{\sigma}_i^2$ is equal to $\hat{\mu}_i$ and $\hat{\mu}_i + a\hat{\mu}_i^2$ for Poisson and Negative Binomial models, respectively [32].

If $\emptyset$ calculated from Pearson χ^2 is significantly greater one overdispersion exists. In that case, GNBR is suggested to be used [36]. On the other hand, if $\emptyset$ determined from deviation G^2 is close to zero, GPR is preferred [32]. Another method for determining overdispersion parameter depending on T and T_1 statistics is given respectively, $T = \frac{1}{2}\sum_{i=1}^{n}[(y_i - \hat{y}_i)^2 - y_i]$ and $T_1 = \frac{\sum_{i=1}^{n}[(y_i-\hat{y}_i)^2-y_i]}{\left(2\sum_{i=1}^{n}\hat{y}_i^{\,2}\right)^{1/2}}$.
According to GPR models, overdispersion exists if T statistic is a positive value, otherwise, underdispersion occurs if T statistic is a negative value. T_1 statistic is actually used for large samples [42]. If overdispersion is determined by one of these methods, GNBR can be used as an alternative to GPR for modelling [41].

In this study, the probability distribution of dependent variable y is assumed to be Negative Binomial with mean $E(y) = \mu$ and variance $Var(y) = \mu\,(1 + \kappa\mu)$ as follows

$$f(y) = \frac{\Gamma(y+\frac{1}{\phi})}{y!\,\Gamma(\frac{1}{\phi})}\left(\frac{1}{1+\phi\mu}\right)^{\frac{1}{\phi}}\left(\frac{\phi\mu}{1+\phi\mu}\right)^{y}, y = 0,1,2\ldots \tag{6}$$

In which ϕ is overdispersion or shape parameter. Accordingly, the GNBR model with logarithmic link function is given below [28]

$$Ln\mu = Ln\,n + x'\beta. \tag{7}$$

2.3. The Earthquake Hazard Analysis for Models

In order to identify the earthquake hazard, the seismicity parameters must be calculated in the following way. The GR model identified by Eq. 1 for the magnitude-frequency relationship can be rewritten as

$$N(M) = 10^{a-bM} \tag{8}$$

The annual average number of earthquakes at a given time (i.e., $n(M \geq M_i\,)$) is found by using Eq.8 [43, 44]. The normal frequency value used to determine seismic risk is found by $n(M) = 10^{a_1'-bM}$ [18]. So, $n(M)$ value explains annual average earthquake occurence number which is calculated according to magnitude and seismicity parameters.

In this study, seismicity parameters for GPR and GNBR models were founded smilarly and obtained equations were given in Table 1. In here t shows the investigated time periods (year) and a' shows the relationship between normal and cumulative frequency.

Table 1. *Earthquake parameter equations for GR, GPR and GNBR models*

GR	GPR	GNBR
$a' = a - Log\,(bLn10)$	$a' = a - Ln\,(bLn10)$	$a' = a - Ln\,(bLn10/n)$
$a_1 = a - Log\,(t)$	$a_1 = a - Ln\,(t)$	$a_1 = a - Ln\,(t/n)$
$a_1' = a' - Log\,(t)$	$a_1' = a' - Ln\,(t)$	$a_1' = a' - Ln\,(t/n)$

The earthquake hazard analysis of future earthquakes can be carried out by using observed earthquake data, considering that the occurrence of earthquakes is a random process with respect to time. Assuming seismic activities as independent events, the occurrence of an earthquake can be considered as a Poisson process [12].

According to Poisson model, the seismic risk (i.e: probability of occurrence of earthquakes for certain T years), $R(M)$ and recurrence period , $Q(T)$ can be determined as years [12].The seismic risk for a region with magnitude M within T years for a t-year observation interval is described by Eq. (9) as follows:

$$R(M) = 1 - e^{-n(M)T} \tag{9}$$

The recurrence or return period , which actually gives information about how many years are required for next earthquake with M severity, is defined by Eq. (10)

$$Q(T) = \frac{1}{n(M)} . \tag{10}$$

2.4. Model Selection

The goodness-of-fit tests including the AIC and BIC given in Eqs. (11-12) are utilized to determine the best model [45, 46].

$$AIC = -2LL + 2d \tag{11}$$

$$BIC = -2LL + dLn(n) \tag{12}$$

where LL is log-likelihood, d is the degree of freedom and n is the number of observations. The model with the smallest AIC or BIC criterion values is considered as the best model.

3. RESULTS AND DISCUSSION

This section of the study consists of three main stages: (1) the determination of frequency distribution for frequency and magnitudes of earthquakes, (2) the performing of dispersion tests and goodness-of fit test for models and (3) the analyzing of the earthquake hazard for selected models are performed as application. In the first step, the results of Kolmogorov Simirnov goodness-of-fit tests show that the normal distribution is suitable to $LogN$ frequency data for GR model and the Poission distribution is suitable to LnN frequency data for GPR model. In the second step, according to dispersion parameters and model selection criteria (AIC and BIC), the GPR model as GLMs has been determined as the best-fitting model for the magnitude-frequency relationship. So, GPR can be utilized for analyzing of the earthquake hazard as alternatively to GR model. Finally, in the third step, recurrence probabilities and return periods are calculated for the selected years depending on yearly average occurrence number of earthquakes estimated with the GR and GPR models.

3.1. The earthquake hazard analysis in Turkey using GR, GPR and GNBR models

In this study, 4863 earthquake data with magnitudes (M) between 4.0 and 7.9 for 115 years from 1900 to 2014 were utilized. Here, M shows the magnitude values of the X_m.such as explained in Section 2. The magnitudes with 0.1 class interval, number of earthquakes or frequency (n), cumulative frequency (N), cumulative frequency per year (N/t) and log transformation of N (i.e., $y = LogN$) are given in Table 2. Here, the values of $y = LogN$ were used as the dependent variable and magnitudes were taken as the independent variable in order to examine the magnitude-frequency relationship. In addition, cumulative frequency values (N) were divided by time period examined t (=115 years) for the calculation of dispersion parameters and model selection criteria (AIC and BIC).

Table 2. *The magnitude and frequency of earthquakes in Turkey between 1900-2014 years*

M	n	N	$LogN$	$Log(N/t)$	M	n	N	$LogN$	$Log(N/t)$
4.0	651	4863	3.687	1.626	6.0	23	97	2.107	-0.074
4.1	478	4212	3.624	1.564	6.1	13	74	1.987	-0.191
4.2	412	3734	3.572	1.511	6.2	6	61	1.869	-0.275
4.3	456	3322	3.521	1.461	6.3	9	55	1.785	-0.320
4.4	366	2866	3.457	1.397	6.4	4	46	1.740	-0.398
4.5	380	2500	3.398	1.337	6.5	4	42	1.663	-0.437
4.6	264	2120	3.326	1.266	6.6	6	38	1.623	-0.481
4.7	328	1856	3.269	1.208	6.7	3	32	1.580	-0.556
4.8	282	1528	3.184	1.123	6.8	9	29	1.505	-0.598
4.9	327	1246	3.095	1.035	6.9	2	20	1.462	-0.760
5.0	141	919	2.963	0.903	7.0	4	18	1.301	-0.805
5.1	67	778	2.891	0.830	7.1	3	14	1.255	-0.915
5.2	121	711	2.852	0.791	7.2	5	11	1.146	-1.019
5.3	175	590	2.771	0.710	7.3	1	6	1.041	-1.283
5.4	82	415	2.618	0.557	7.4	1	5	0.778	-1.362
5.5	88	333	2.522	0.462	7.5	1	4	0.699	-1.459
5.6	49	245	2.389	0.328	7.6	1	3	0.602	-1.584
5.7	33	196	2.292	0.232	7.7	1	2	0.477	-1.760
5.8	35	163	2.212	0.151	7.8	0	1	0.301	-2.061
5.9	31	128	2.107	0.047	7.9	1	1	0	-2.061

Table 3. *Descriptive statistics, model parameter estimations and goodness of fit tests for dependent variables*

Dependent Variables	Descriptive Statistics		Model	Parameter Estimations		Goodness of Fit Test	
	Mean	Std. Deviation		Mean	Std. Deviation	K-S	p value
$LogN$	2.154	1.113	Normal	2.154	1.113	1.166	0.132
LnN	4.897	2.414	Poisson	4.897	4.897	0.689	0.729

The descriptive statistics, model parameter estimations and goodness of fit test results for dependent variables are tabulated in Table 3. In here, the values of LogN and LnN respectively are taken as the dependent variable in GR and GPR models. Kolmogorov-Simirnov (K-S) test is used for goodness of fit tests. According to the results, it is shown that LogN and LnN are statistically fit to the Normal distribution ($\text{p value} = 0.132 > 0.05$) and Poisson distribution ($\text{p value} = 0.729 > 0.05$), respectively. Therefore, in this study, Normal and Poisson distributions are used for dependent variables of GR and GPR models, respectively.

The regression analysis for GR, GPR and GNBR models were performed by SPSS 18.0 package software. The parameter estimations, model significance tests, dispersion parameters and model selection criteria results for each model are given in Table 4. Here, the GR model obtained as $LogN = 7.579 - 0.927M$ with coefficient of determination $R^2 = 0.9910$ was found statistically significant ($p < 0.05$) indicating that earthquake data can be described with this model.

Although both GPR and GNBR models were also found statistically significant ($p < 0.05$), dispersion tests need to be carried out before the model selection. For this purpose, T and T_1 statistics were first calculated as -93.0926 and -1.47965, respectively. This shows the existence of underdispersion since T was found negative and $T_1 > Z_{table}(= -1.645)$ at 5% significance level. The dispersion parameters for both models were found less than one ($\phi = 0.22$ for GPR and $\phi = 0.083$ for GNBR) according to Pearson χ^2 indicating underdispersion. Therefore, GNBR cannot be employed if underdispersion is encountered as suggested by Famoye and Singh (2006), [41].

In addition, the results of AIC and BIC for the choice of the best model are given in Table 4. According to these results, it is seen that the data fit better to GPR model which has relatively small AIC and BIC values. As an alternative to GR model, GPR model is obtained as $LnN = 15.905 - 1.821M$ by taking the underdispersion condition into account. The graphs of magnitude-frequency relations estimated by GR and GPR models are shown in Figure 3.

In order to perform the earthquake hazard analysis, we determine the magnitude-frequency relationship with GR and GPR models using the seismicity parameters given in Table 5.

Table 4. *Significance tests, dispersion tests and model selection criteria for GR, GPR and GNBR models*

		GR		GPR		GNBR	
	Parameter	**Estimation (Std.error)**	**p-value**	**Estimation (Std.error)**	**p-value**	**Estimation (Std.error)**	**p-value**
Significance of the coefficients	a	7.579 (0.0914)	0.000	15.905 (0.0461)	0.000	17.278 (0.9040)	0.000
	b	-0.927 (0.0151)	0.000	-1.821 (0.0102)	0.000	-2.098 (0.1511)	0.000
Dispersion Test Statistics		**Test value**	**ϕ**	**Test value**	**ϕ**	**Test value**	**ϕ**
	G^2	32.734	0.861	12.010	0.316	4.745	0.125
	χ^2	32.734	0.861	8.368	0.220	3.149	0.083
Model Selection Criteria	LL	-52.748		-48.978		-70.959	
	AIC	111.497		101.956		145.918	
	BIC	116.569		105.334		149.295	

Table 5. *Earthquake parameters for GR and GPR models*

Models	a	b	a'	a_1	a_1'
GR	7.579	0.927	7.250	5.518	5.189
GPR	15.905	1.821	14.472	11.160	9.727

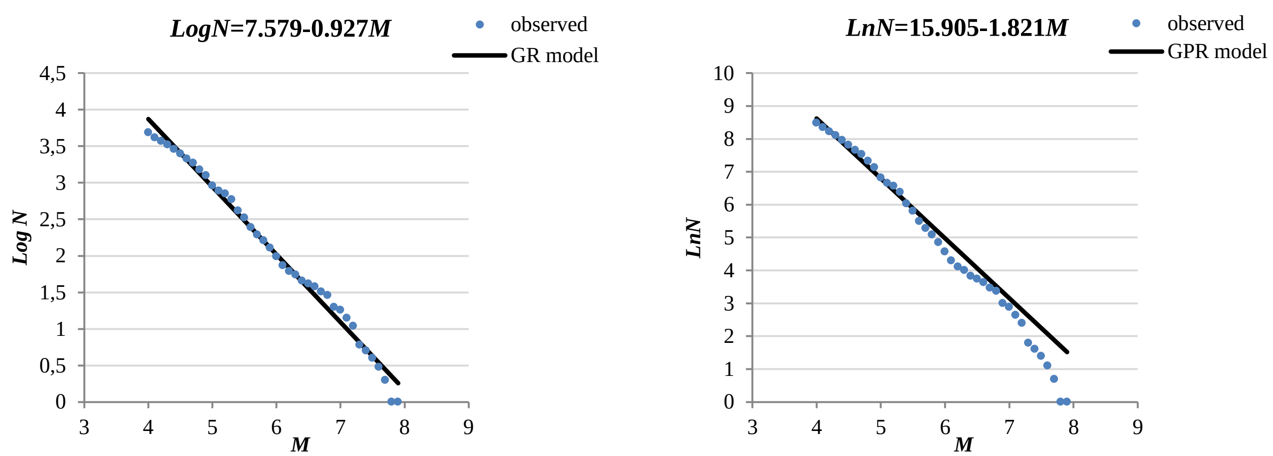

Figure 3. *Magnitude-freqency relations estimated by GR and GPR models for Turkey.*

The results of seismic risk values and recurrence periods are shown in Table 6. According to the results obtained, it is found that the occurrence possibility of an earthquake with $M \geq 6$ within 20 years is 99.98%, return period is 2.3604 years according to GR model; 99.76% and return period is 3.3180 years according to GPR model. Seismic risk values found by GPR is higher than the results obtained with GR model for $M > 7.0$. Similarly, when return periods are investigated according to earthquake magnitude, GR and GPR models give close values but GPR model produces smaller values of return periods compared to the GR model for 7.1 and bigger magnitude earthquakes. For example, return periods and occurrence probabilities in 20 years for 7.4 magnitude earthquake was found to be 46.8590 years and 34.67% for GR model and 42.4680 years and 37.56% for GPR model, respectively (Table 7).

Table 6. *Earthquake risk analysis results obtained by using GR and GPR models*

Model	**GR**	$\boldsymbol{LogN = 7.579 - 0.927M}$								
M	$n(M)$	1 year	5 year	10 year	20 year	30 year	50 year	75 year	100 year	$Q(T)$
4	30.2696	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0330
4.5	10.4113	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0960
5	3.5810	0.9722	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.2793
5.5	1.2317	0.7082	0.9979	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8119
6	0.4236	0.3453	0.8798	0.9855	0.9998	1.0000	1.0000	1.0000	1.0000	2.3604
6.5	0.1457	0.1356	0.5174	0.7671	0.9454	0.9874	0.9993	1.0000	1.0000	6.8627
7	0.0501	0.0489	0.2217	0.3942	0.6321	0.7777	0.9184	0.9767	0.9933	19.9523
7.5	0.0172	0.0171	0.0826	0.1583	0.2910	0.4038	0.5777	0.7255	0.8216	58.0087

Model	**GPR**	$\boldsymbol{LnN = 15.905 - 1.821M}$								
M	$n(M)$	1 year	5 year	10 year	20 year	30 year	50 year	75 year	100 year	$Q(T)$
4	11.5035	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0869
4.5	4.6281	0.9902	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.2161
5	1.8620	0.8446	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.5371
5.5	0.7491	0.5272	0.9764	0.9994	1.0000	1.0000	1.0000	1.0000	1.0000	1.3349
6	0.3014	0.2602	0.7784	0.9509	0.9976	0.9999	1.0000	1.0000	1.0000	3.3180
6.5	0.1213	0.1142	0.4546	0.7026	0.9115	0.9737	0.9977	0.9999	1.0000	8.2470
7	0.0488	0.0476	0.2165	0.3860	0.6231	0.7686	0.9128	0.9742	0.9924	20.4985
7.5	0.0196	0.0194	0.0935	0.1782	0.3247	0.4450	0.6252	0.7705	0.8595	50.9503

Table 7. *Seismic risk parameters for GR and GPR models.*

	$\boldsymbol{n(M)}$		$\boldsymbol{Q(T)}$		$\boldsymbol{R(M)}$	
$\boldsymbol{M}$	**GR**	**GPR**	**GR**	**GPR**	**GR**	**GPR**
4	30.2696	11.5035	0.0330	0.0869	1.0000	1.0000
4.5	10.4113	4.6281	0.0960	0.2161	1.0000	1.0000
5	3.5810	1.8620	0.2793	0.5371	1.0000	1.0000
5.5	1.2317	0.7491	0.8119	1.3349	1.0000	1.0000
6	0.4236	0.3014	2.3604	3.3180	0.9998	0.9976
6.5	0.1457	0.1213	6.8627	8.2470	0.9454	0.9115
7	0.0501	0.0488	19.9523	20.4985	0.6321	0.6231
7.1	0.0405	0.0407	24.6998	24.5927	0.5541	0.5566
7.2	0.0327	0.0339	30.5769	29.5047	0.4792	0.4923
7.3	0.0264	0.0283	37.8524	35.3978	0.4097	0.4316
7.4	0.0213	0.0235	46.8590	42.4680	0.3467	0.3756
7.5	0.0172	0.0196	58.0087	50.9502	0.2910	0.3247

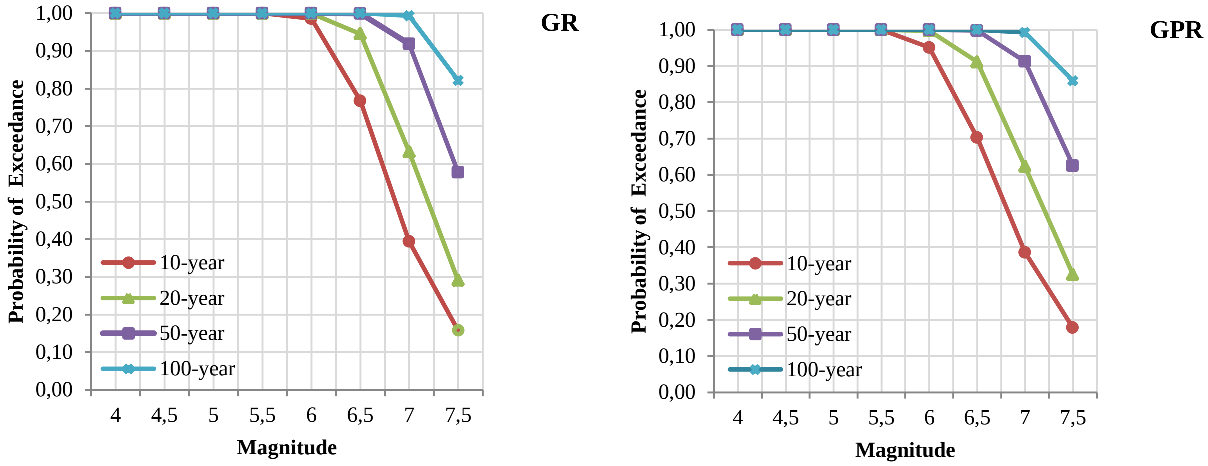

Figure 4. *The probability of exceedence graphs for GR and GPR models*

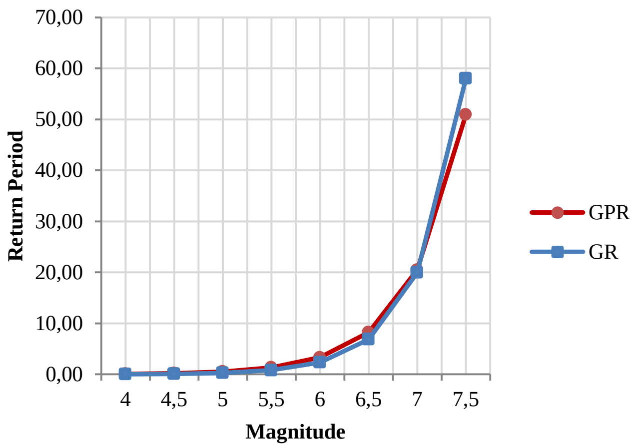

Figure 5. *The return period graphs for GR and GPR models*

Possibilities of exceeding the earthquake magnitudes in the given periods (10-year, 20-year, 50-year and 100-year) for GR and GPR models are shown in Figure 4. The probability of exceedence graphs for GR and GPR model. Here, while magnitude values for each model are increasing, corresponding earthquake possibilities are decreasing. In addition, while the given periods increases, the probability of exceedance increases for each model at a given magnitude value (for example: $M = 7.0$). In Figure 5, the return periods of earthquakes are shown for GR and GPR models. Here, while magnitude values increase, return period also increase for each model. The results show that GPR model coherent with the known GR model.

4. CONCLUSIONS

In this study, the magnitude-frequency relationship for seismicity of Turkey is modelled with GLM (GPR and GNBR models) as an alternative to Gutenberg Richter (GR) model. Modelling of the magnitude-frequency relationship with GLM is important because it takes dispersion into account. The other main advantage of using GLM for the earthquake hazard analysis is that it is also applicable for not normally distributed data.

In this paper, it is shown that the magnitude-frequency relationship in Turkey can be better explained with GPR model which produced relatively small AIC and BIC values compared to both GR and GNBR models. Hence, GPR modeling as a statistical approach is utilized for estimating parameters including earthquake return periods and occurrence probabilities. Actually, it is shown that GR and GPR models generated relatively close values, but the GPR model produced smaller values of return periods compared to the GR model for 7.1 and higher magnitude earthquakes. It can be concluded from the results that in contrast to the GR model, GPR model gives an estimated before occurrence time for the high magnitude earthquake. For example, using the GR and GPR model, the recurrence probability is around 35% and 38% for an earthquake with magnitude greater than 7.4 in a 20-year period in Turkey and such an earthquake to occur in a 43-year and 47-year periods, respectively.

It is known that high-magnitude earthquake caused the loss of life and property. Hence this study, the using of proposed the GPR model for earthquake risk prediction will be an essential factor in reducing the loss of property since it is estimated the occurrence time before with GPR model.
In conclusion, the study results (i.e., return period and occurrence probabilities) based on GR and GPR models revealed that Turkey is located an earthquake region with relatively high risk. In further studies, additional variables such as earthquake risk regions, amount of live and material loss can be added to the earthquake hazard analysis with GPR models.

ACKNOWLEDGEMENTS

We would like to thank the referees for their careful reading of the manuscript and their valuable suggestions. Also, we thank Prof. Dr. Osman Yıldız for his valuable contribution in edited the English of the final text.

CONFLICTS OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

[1] Saroğlu, F., Emre, O., Kuşcu, I., "The East Anatolian fault zone of Turkey", Ann. Tectonicae, 6: 99–125, (1992).

[2] Gutenberg, B., Richter, C.F., "Seismicity of the earth and related phenomena", Second Printed, Princeton University Press, Princeton, (1954).

[3] Firuzan, E., "Statistical earthquake frequency analysis for Western Anatolia", Turkish Journal of Earth Sciences, 17: 741-62, (2008).

[4] Dargahi-Noubary, G.R., "A method for predicting future large earthquakes using extreme order statistics", Physics of the Earth and Planetary International, 42: 241–45, (1986).

[5] Main, I., "Statistical physics seismogenesis and seismic hazard", Review of Geophysics, 34: 433–62, (1996).

[6] Kagan, Y.Y., "Statistics of characteristic earthquakes", Bulletin of Seismological Society of America, 83: 7–24, (1993).

[7] Cobanoğlu, I., Bozdağ, Ş., Dinçer, I., Erol, H., "Statistical approaches to estimating the recurrence of earthquakes in the Eastern Mediterranean Region", Istanbul University Earth Sciences Review, 19: 91-100, (2006).

[8] Kalyoncuoğlu, Y., "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: 131-48, (2007).

[9] Bayrak, Y., Yılmaztürk, A., Öztürk, S., "Relationships between fundamental seismic hazard parameters for the different source regions in Turkey", Natural Hazards, 36: 445–62, (2005).

[10] Öztürk, S., Bayrak, Y., Cınar, H., Koravos, G.C., Tsapanos, T.M., "A quantitative appraisal of earthquake hazard parameters computed from Gumbel I method for different regions in and around Turkey", Natural hazards, 47(3): 471-95, (2008).

[11] Ünal, S., Çelebioğlu, S., "A Markov chain modelling of the earthquakes occurring in Turkey", GU J Sci, 24: 263–74, (2011).

[12] Ünal, S., Çelebioğlu, S., Özmen, B., "Seismic hazard assessment of Turkey by statistical approaches", Turkish J Earth Sci, 23: 350-60, (2014).

[13] Yücemen, M.S., Akkaya, A., "Stochastic models for the estimation of seismic hazard and their comparison", in In:Proceedings of the 3rd Earthquake Engineering Conference. İstanbul, Turkey, pp. 466–77, İstanbul, Turkey, (1995).

[14] Kestel, S.S., Yücemen, M.S., "Earthquake reliability of life line networks", Turk J Engin Environ Sci, 24: 143-60, (2000).

[15] Selçuk, A.S., Yücemen, M.S., "Reliability lifeline networks with multiple sources under seismic hazard", Natural Hazards, 21: 1-18, (2000).

[16] Kasap, R., Gürlen, Ü., "Obtaining the return period of earthquake magnitudes: as an example Marmara Region", Doğus University Journal, 4: 157–66 (article in Turkish with English abstract), (2003).

[17] Yücemen, M.S., "Probabilistic assessment of earthquake insurance rates for Turkey", Natural Hazards Springer, 35: 291-313, (2005).

[18] Özmen, B., "Probability of earthquake occurrences to Ankara", Bulletin of the Earth Sciences Application and Research Centre of Hacettepe University, vol. 34, no. 1, pp. 23-26 (article in Turkish with English abstract), 2013.

[19] Kayabali, K., "Modeling of seismic hazard for Turkey using the recent neotectonic data", Eng Geol, 63: 221–32, (2002).

[20] Koravos, G.C., Main, I.G., Tsapanos, T.M., Musson, R.M.W., "Perceptible earthquakes in the broad Aegean area", Tectonophysics, 371(1): 175-86, (2003).

[21] Mäntyniemi, P., Tsapanos, T.M., Kijko, A., "An estimate of probabilistic seismic hazard for five cities in Greece by using the parametric-historic procedure", Engineering Geology, 72(3): 217-31, (2004).

[22] Bayrak, Y., Öztürk, S., Cınar, H., Kalafat, D., Tsapanos, T.M., Koravos, C.G., Leventakis, G.A., "Estimating earthquake hazard parameters from instrumental data for different regions in and around Turkey", Eng Geol, 105: 200–10, (2009).

[23] Xie, F., Wang, Z., Liu, J., "Seismic hazard and risk assessments for Beijing–Tianjin–Tangshan, China, Area", Pure and applied geophysics, 168(3-4): 731-38, (2011).

[24] Wang, J.P., Lin, C.W., Taheri, H., Chan, W.S., "Impact of fault parameter uncertainties on earthquake recurrence probability examined by Monte Carlo simulation—an example in Central Taiwan", Engineering geology, 126: 67-74, (2012).

[25] Akol, B., Bekler, T., "Assessment of the statistical eartquake hazard parameters for NW Turkey", Natural hazards, 68(2): 837-53, (2013).

[26] Liu, J., Xie, F., Lv, Y., "Seismic hazard assessments for the Ordos Block and its periphery in China", Soil Dynamics and Earthquake Engineering, 84: 70-82, (2016).

[27] McCullagh, P., Nelder, J.A., "Generalized linear models", 2nd Ed., London: Chapman and Hall, (1989).

[28] Jong, P., Heller, G.Z., "Generalized linear models for insurance data", New York: Cambridge University Press, (2008).

[29] Nelder, J.A., Wedderburn, R.W.M., "Generalized linear models", J. Roy. Statist. Soc. A, 135: 370-84, (1972).

[30] Dunteman, G.H., Ho, M.R., "An introduction to generalized linear models", Thousand Oaks, Calif.: Sage Publications, (2006).

[31] Tüzel, S., Sucu, M., "Zero-inflated discrete models for claim frequencies", Journal of Statisticians, 5: 23-31 (article in Turkish with English abstract), (2012).

[32] Myers, R.H., Montgomery, D.C., Vining, G.G., "Generalized linear models", New York: John Wiley and Sons, (2002).

[33] George, J., Letha, J., Jairaj, P.G., “Daily rainfall prediction using generalized linear bivariate model–a case study”, Procedia Technology, 24: 31-38, (2016).

[34] Bogazici University Kandilli Observatory And Earthquake Research Institute (KOERI), Istanbul, Turkey, http://www.koeri.boun.edu.tr, 2.8.2014 [Online].

[35] Taban A., Gençoğlu, S., "Earthquakes and parameters", Earthquake Research Institute, 11: 7-83 (article in Turkish), (1975).

[36] Montgomery, D.C., Peck, E.A., Vining, G.G., "Introduction to linear regression analysis", 5th Edition, John Wiley and Sons, (2010).

[37] Greene, W.H., "Accounting for excess zeros and sample selection in Poisson and negative binomial regression models", Tech. Rep., NYU Working Paper No. EC-94-10, March (1994).

[38] Giles, D., "Notes on the zero inflated Poisson model", Department of Economics, University of Victoria, (2010).

[39] Lawless, J.F., "Negative binomial and mixed Poisson regression", The Canadian Journal of Statistics, 15(3): 209-25, (1987).

[40] Famoye, F., "Restricted generalized Poisson regression model", Communications in Statistics-Theory and Methods, 22(5): 1335-54, (1993).

[41] Famoye, F., Singh, K.P., "Zero-Inflated generalized Poisson regression model with an application to domestic violence data", Journal of Data Science, 4: 117-30, (2006).

[42] Dean, C., Lawless, J.F., "Test for defecting overdispersion in Poisson regression models", Journal of American Statistical Association, 84: 467-72, (1989).

[43] Tuksal, I., "Seismicity of the North Anatolia Fault System in the domain of space, time and magnitude", M.S. Thesis. Saint-Louis Universty, Saint-Louis, Missouri (article in Turkish with English abstract), (1976).

[44] Alptekin, O., "Magnitude–frequency relationships and deformation release for the earthquakes in and around Turkey", Thesis for Promoting to Associate Professor Level, Karadeniz Technical University, p. 107 (in Turkish), (1978).

[45] Akaike, H., "A new look at the statistical model identification", Automatic Control, IEEE Transactions 19: 716–23, (1974).

[46] Schwarz, G., "Estimating the dimension of a model", Annals of Statistics, 6: 461–64, (1978).