Turkish Journal of Earth Sciences (Turkish J. Earth Sci.), Vol. 17, 2008, pp. 741–762. *Copyright ©TÜBİTAK*
First published online 13 March 2008

Statistical Earthquake Frequency Analysis for Western Anatolia

ESİN FİRUZAN

Dokuz Eylül University, Department of Statistics, Tınaztepe Campus, Buca, TR–35160 İzmir, Turkey
(e-mail: esin.firuzan@deu.edu.tr)

Abstract: In order to apply a statistical earthquake frequency analysis to western Anatolia, the area has been restricted within coordinates 36.50º–40º North and 26.00º–30.00º East. Data used in the analysis belong to the instrumental period from 1.1.1900 to 10.11.2006 (Ms ≥ 4.0). The earthquake record was assumed to be representative of longer periods and the possibility of clusters of seismicity or of extreme magnitude events were ignored. The study area has been divided into eight zones with different seismotectonic properties. All computations have been performed for these zones separately. The earthquake magnitude series is termed as the instantaneous peak over the threshold (POT) and the series of the largest earthquake magnitude per year in a zone as the annual maximum (AM). POT and AM series of the surface wave magnitude (M_s) are obtained from the catalog of the Kandilli Observatory and Earthquake Research Institute (KOERI) of Bosphorus University. Exceedance probabilities and return periods of earthquakes of various magnitudes have been calculated for each zone.

This paper discusses the comparative assessment of four advanced distributions to provide accurate and reliable earthquake estimates for all sub-regions in western Anatolia. The models compared are the Exponential (EXP), Extreme value distribution Type 1 (Gumbel) (GUM), Log Pearson Type 3 (LP3), and Generalised Pareto (GP). Model parameters are estimated using the method of moments and the quantitative assessment of each model is based on the Anderson Darling Test results and quantile estimation of each model. On the basis of these comparisons, it is concluded 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.

The results show that *b* values in the computed magnitude-frequency relations range from 0.40–0.76. These results are supported by previous studies of the Aegean and surrounding area (Papazachos 1990; Öncel & Wilson 2004; Sayıl & Osmanşahin 2007). The highest *b* value has been obtained for Zone 7 (Gökova Gulf). As a result of the GUM, EXP and GP models, Zone 8 has the lowest risk for magnitude 7.0 and longest return periods; however, LP3 model results show that Zone 7 has the lowest risk for magnitude 7.0 and the longest return period.

Key Words: extreme value distribution, Western Anatolia, zonation, seismic assessment, earthquake

Batı Anadolu İçin İstatistiksel Deprem Frekans Analizi

Özet: Çalışma, Batı Anadolu'ya istatistiksel frekans analizini uygulamak için (36.50º–40º) Kuzey ve (26.00º–30.00º) Doğu koordinatları ile sınırlandırılmıştır. Analizde kullanılan veriler aletsel dönemi 1.1.1900–10.11.2006 (Ms ≥ 4.0) tarihlerini içermektedir. İncelenen deprem verilerinin daha uzun dönem verilerini temsil ettiği ve bu dönem içerisinde sismisitenin kümelenme olasılığının veya çok şiddetli depremler olma olasılığının olmadığı varsayıldı. Çalışma alanı sismotektonik özellikler nedeniyle sekiz bölgeye bölünmüştür. Bütün hesaplamalar bu bölgeler için ayrı ayrı yapılmıştır. Deprem verileri, belirli bir deprem şiddetini veya büyüklüğünü aşan ani çıkışlar (POT) ve belirli bir bölgede meydana gelen yıllık maksimum deprem şiddetleri ya da büyüklükleri (AM) olmak üzere iki kısımda ifade edilmiştir.

Yüzey dalga şiddeti (M_s) POT ve AM serileri, Boğaziçi Üniversitesi Kandilli Rasathanesi kataloğundan elde edilmiştir. Her bölge için çeşitli deprem şiddetlerinin aşılma olasılıkları ve geri dönüş periyodları hesaplanmıştır.

Bu çalışma, Batı Anadolu'daki bütün bölgeler için daha güvenilir ve daha doğru tahminler sağlamak amacıyla dört istatistiksel dağılım kullanılarak, karşılaştırmalı bir değerlendirilme yapılmasını amaçlamaktadır. Karşılaştırılmak üzere seçilen modeller Üstel, Gumbel, Log Pearson Tip 3 ve Genelleştirilmiş Pareto dağılımlarıdır.

Model parametreleri moment yöntemi ile tahminlenmiştir. Her bir model için niceliksel değerlendirme Anderson Darling Test sonuçlarına ve kantil tahminlerine dayandırılmıştır. Tüm karşılaştırmalara dayanarak, GP dağılımının belirli bir deprem şiddetini veya büyüklüğünü aşan ani çıkışlar için, Gumbel dağılımının da yıllık maksimum deprem verileri için en uygun dağılım olduğu sonucuna varılmıştır.

Elde edilen sonuçlara göre, şiddet-frekans ilişkisinde hesaplanan *b* değerleri 0.40–0.76 aralığındadır. Bu sonuçlar, daha önce Ege Bölgesi ve çevresi için yapılan çalışmalarla desteklenmektedir (Papazachos 1990; Öncel & Wilson 2004; Sayıl & Osmanşahin 2007). En yüksek *b* değeri Bölge 7 (Gökova Körfezi) için belirlenmiştir. Gumbel, Üstel ve Genelleştirilmiş Pareto modelleri sonuçlarına göre, Bölge 7 ve 8, 7 şiddeti için en düşük risk ve en uzun geri dönüş periyoduna sahiptir. Bununla birlikte, LP3 dağılım sonuçları, Bölge 7'nin 7 şiddeti için en düşük risk ve en uzun geri dönüş periyoduna sahip olduğunu göstermiştir.

Anahtar Sözcükler: uçdeğer dağılımı, Batı Anadolu, bölgelendirme, sismik değerlendirme, deprem

Introduction

Western Anatolia is a part of the Aegean Province of tectonic extension and is one of the most seismically active continental regions of the world (Akyol *et al*. 2006). It comprises part of a region of distributed extension that also includes parts of Greece, Macedonia, Bulgaria, and Albania as shown in Figure 1. As a result of this extension western Anatolia is characterized by East–West-trending horsts and deep sediment filled grabens. Approximately E–W-trending grabens (e.g., Edremit, Bakırçay, Kütahya, Simav, Gediz, Küçük Menderes, Büyük Menderes and Gökova grabens) and their basin-bounding active normal faults are the most dominant neotectonic features of western Turkey as shown in Figure 2. Other, less dominant structural elements are NNE–SSW-trending basins and horsts (e.g., Gördes, Demirci, Selendi and Uşak-Güre basins). In terms of total displacement and strike, the Gediz and Büyük Menderes fault zones are the most important (Sarı & Şalk 2006).

Because earthquakes are extreme and complex phenomena, many scientific studies are interested in predicting their occurrence time and assessing hazards and a statistical approach is one of these research areas. To apply statistical models to determine the seismicity in western Anatolia, this region must be seismotectonically homogenous (constant value of *b* parameter of the frequency-magnitude relation, where every point is assumed to have the same probability of a future earthquake). Thus, it is necessary to separate the whole study region into seismic zones that should be as seismotectonically homogenous as possible. Since the study area is rather broad and the seismicity is not homogenous, the detailed zonation of Papazachos & Kiratzi (1996), which separates the studied area into 19 seismogenic sources, has been adopted. Based on this article, the study area has been separated to eight sub-regions called zones. The details of seismic zonation of earthquakes in western Anatolia are given in the 2nd section.

Statistical models of earthquake occurrence have gained more importance as earthquakes occur very often. Such statistical models can be used to forecast earthquake occurrence, recurrence intervals, maximum ground motions, and earthquake hazard at a given region. Recurrence intervals are means of expressing the likelihood that a given magnitude earthquake will be exceeded in a specified number of years and is an important factor in earthquake control, emergency planning, and insurance considerations.

The probabilistic approach is convenient for comparing risks in various parts of a country and comparing the earthquake risk with other natural disasters. In this paper, the approach used for earthquake-frequency estimation is to consider a sequence of maximum annual earthquakes and peaks over the threshold earthquake series that will provide suitable advanced statistical distributions.

All estimates of probabilistic seismic hazards are based on the frequency-magnitude relationship, most commonly the Gutenberg-Richter (GR) law. However, the validity of GR has been questioned by many investigators especially in the tails of the distribution (Dargahi-Noubary 1986; Main 1996). An important drawback of the GR model is that an upper bound for the largest earthquake magnitude does not exist. Dargahi-Noubary (1986), Kagan (1993) and Main (1996) suggest that more flexible statistical models should be used instead of the GR model to limit the distribution at high magnitudes. The application of the GR model to the series of earthquake magnitudes is equivalent to the application of an EXP probability density function (pdf) (Thompson *et al*. 2007) and therefore does not have an upper bound. Also, the GUM distribution is the annual maximum earthquake equivalent of the EXP and also does not have an upper bound. Unlike GUM, GR and the EXP models, GP includes an upper bound and an additional complication required of truncated distributions is not required. Moreover, the models introduced in this paper fit the observed large earthquakes far better than the GR model, especially in the tails of the distributions. In addition, three parameter distributions will always fit the observations better than fewer parameter distributions. They are more flexible although they may be less stable for predictions.

The purpose of the present paper is to estimate the probability of earthquake occurrence and earthquake magnitude corresponding to any return period using statistical models from instrumental data for each zone in western Anatolia. This includes the choice of frequency distributions, distributions fitting to the data, and the estimation of earthquakes based on probabilistic approach.

Methods, Material Studied and Descriptions

Definition of Seismogenic Zones

Seismic zonation is important not only for theoretical reasons (improvement of understanding of the geodynamics of a region) but also for practical reasons.

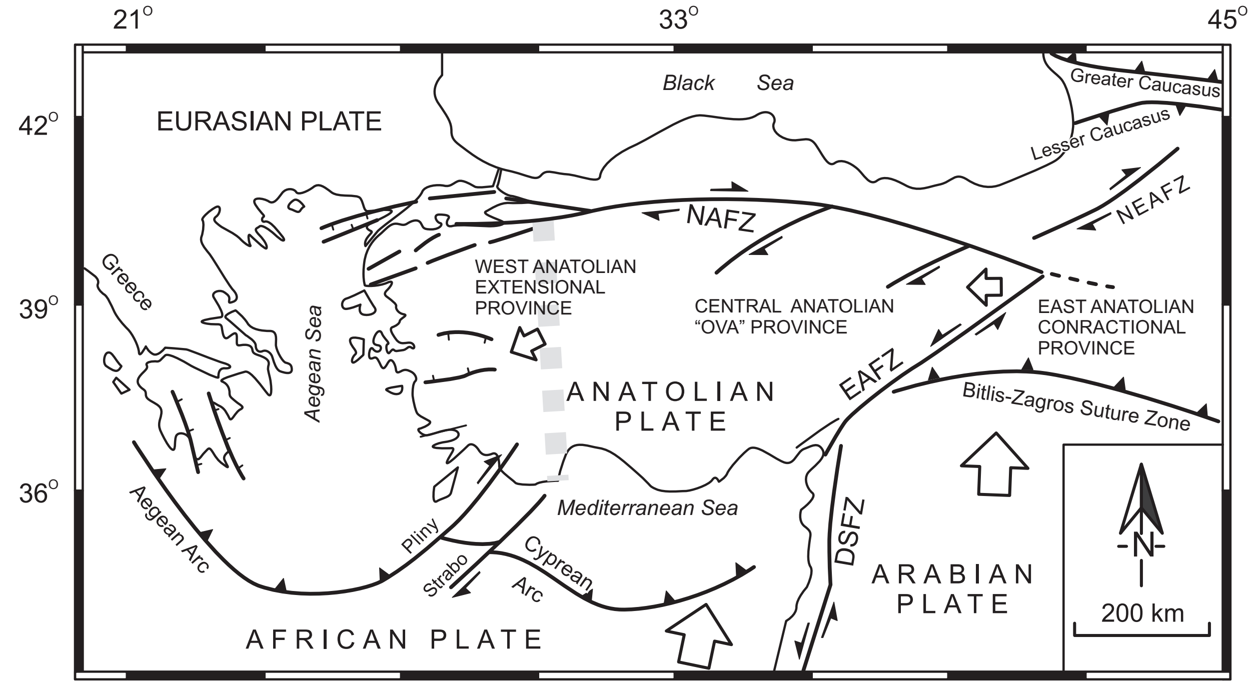

Figure 1. Simplified tectonic map of Turkey showing major neotectonic structures and neotectonic provinces (from Şengör *et al.* 1985; modified from Barka & Reilinger 1997; Kiratzi & Louvari 2001; Bozkurt & Sözbilir 2004). K– Karlıova, KM– Kahramanmaraş, DSFZ– Dead Sea Fault Zone, EAFZ– East Anatolian Fault Zone, NAFZ– North Anatolian Fault Zone, NEAFZ– Northeast Anatolian Fault Zone. Heavy lines with half arrows are strike-slip faults with the arrows showing relative movement sense. Heavy lines with filled triangles show major folds and thrust belts with the triangles indicating the direction of convergence. Heavy lines with open triangles indicate an active subduction zone. Bold filled arrows indicate the movement directions of the African and Arabian plates relative to Eurasia (Akyol *et al.* 2006). The hatched area shows the transition zone between the western Anatolian extensional province and the central Anatolian 'ova' province from Bozkurt (2001).

Seismic hazard evaluation and earthquake prediction are the most important practical problems of seismology. The solutions to both problems depend much on seismic zonation.

Several seismologists (Papazachos 1990; Öncel & Wilson 2004; Sayıl & Osmanşahin 2007) have studied seismic zonation in the Aegean and surrounding area. A seismogenic zone must include seismically homogenous fault segments where every point is assumed to have the same probability of a future earthquake. Zones are mainly defined by two fundamental characteristics. These are seismic profile and the tectonic regime of the region (Sayıl & Osmanşahin 2007). The map in Figure 3 showing the main faults demonstrates the quite high seismic activity in the region. Based on these observations eight seismic zones are defined in this paper. A detailed description of these zones is made with respect to their seismotectonic properties. In addition, parameters, which are currently used as measures of seismicity, have been determined for each seismic zone and are given in Table 1. The name of the sub-regions, latitude and longitude intervals, completeness date, number of earthquakes, depth intervals, parameter *b* values GR law, maximum magnitude, and recording periods for the western Anatolian dataset are shown in Table 1. Seismic risks have been calculated separately.

Data

Instrumental period (Ms ≥ 4.0) data obtained from the KOERI (Kandilli Observatory and Earthquake Research Institute) catalog from 1 January 1900 to 31 December 2006 are used in this study. The earthquake record is assumed to be representative of longer periods and the possibilities of clusters of seismicity or of extreme magnitude events are ignored. Homogeneity of the data is very important in the analysis. In order to ensure homogeneity, all of the magnitudes have been taken as

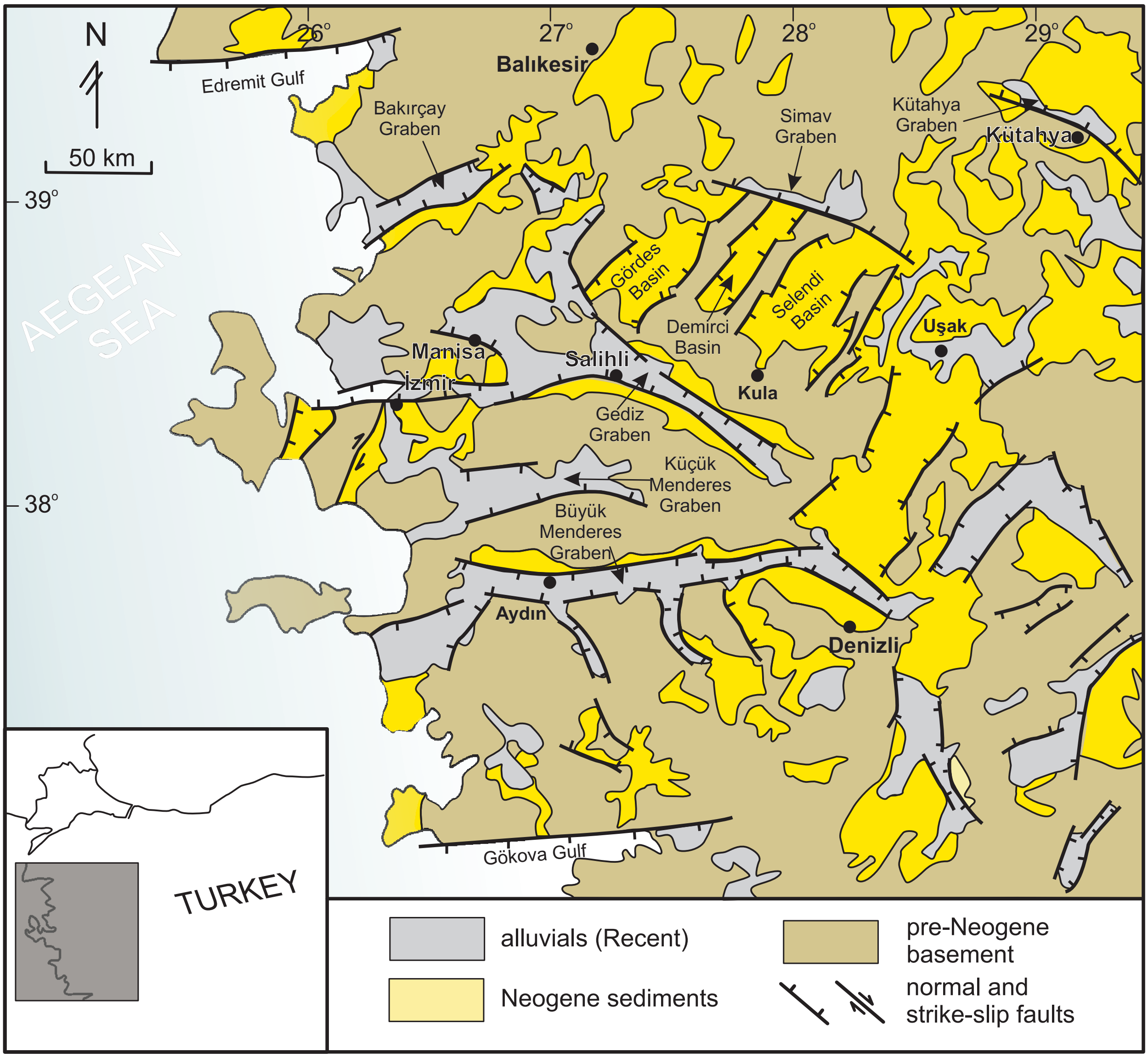

Figure 2. Simplified map showing major structural elements of western Anatolia. Heavy lines with hachures show normal fault: hachures indicate down-thrown side, other heavy lines show strike-slips faults and the arrows along them show the relative direction of movement(from Bozkurt 2001; Bozkurt & Sözbilir 2004).

surface wave magnitude. These magnitudes have been determined by seismologists who compiled the catalogues either from recordings of long-period seismometers or through the use of experimental scaling relations (Sayıl & Osmanşahin 2007). We thus study the earthquake size distribution for a revised and extended earthquake catalog.

An important criterion for the analysis is the completeness of the data. Namely, the data must include all of the earthquakes that occurred in a certain seismogenic region during a specific time period with magnitudes larger than a specific minimum (cut-off) magnitude (Sayıl & Osmanşahin 2007). The smallest magnitude from which earthquakes were reliably reported in the catalogues has been chosen as a minimum magnitude (M_{min}= 4.0 in our case for all zones) in each sub-region.

The probability densities and distribution functions of magnitude random variables have been found and then the return periods of earthquakes of various magnitudes have

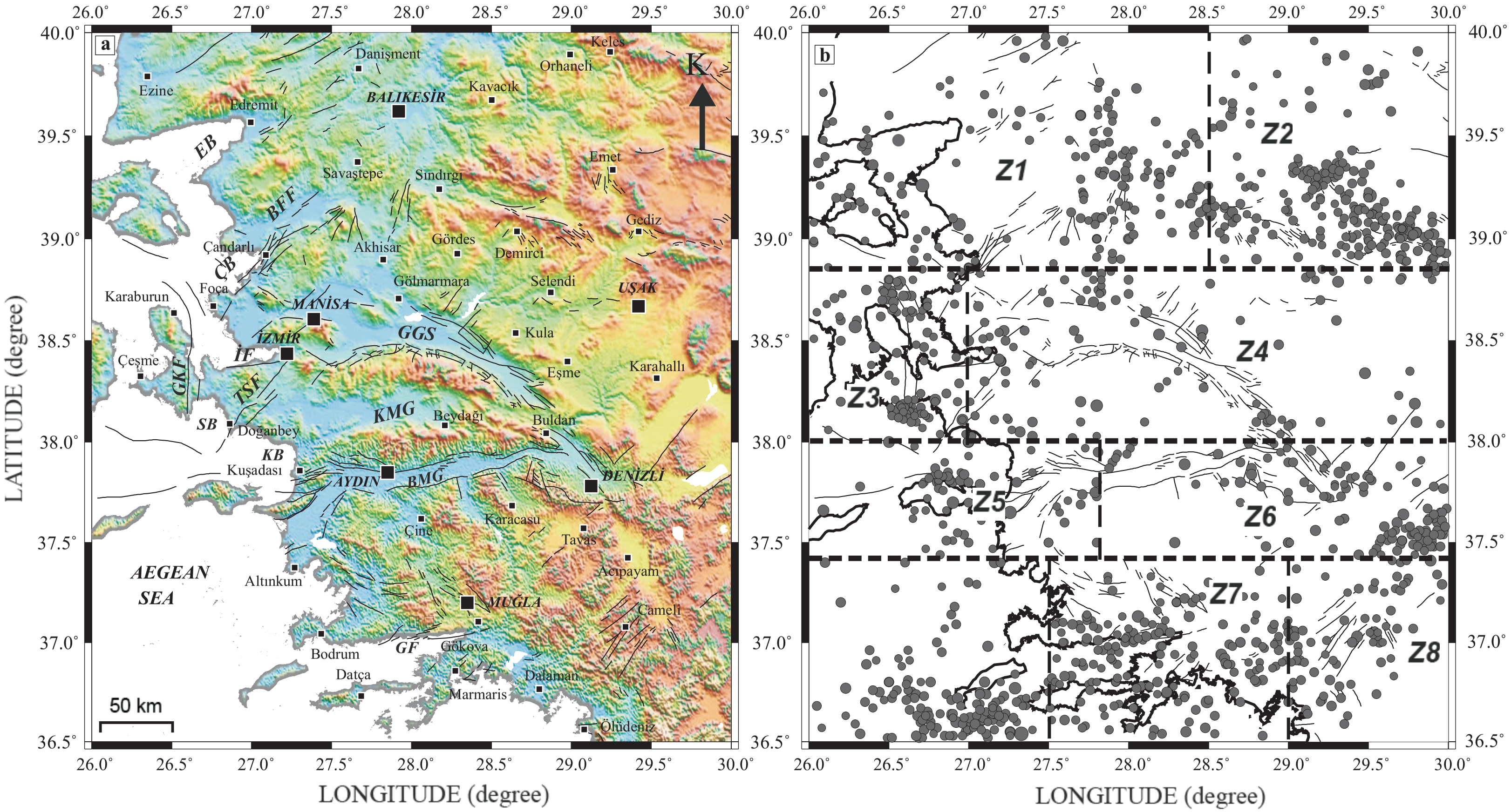

Figure 3. (a) Main tectonic elements together with topography, and (b) seismotectonic properties of the study area. Bathymetric image was drawn by using Earth Topography 2 Min (Etopo2), and the topographic image was produced from the Global Topography 30s Digital Elevation Model (Gtopo20-Dem) Data. Figures are plotted by using Generic Mapping Tool (GMT) Software (Wessel & Smith 1995). Tectonic features are compiled from Bozkurt (2001), Emre *et al.* (2005), Ocakoğlu *et al.* 2004, 2005; Şaroğlu *et al.* 1992. GGS– Gediz Graben system, GCFZ– Gölhisar-Çameli Fault Zone, TSF– Tuzla-Seferihisar Fault, ZBF– Zeytindağ-Bergama Fault, GKF– Gülbahçe Karaburun Fault, MF– Manisa Fault, İF– İzmir Fault.

Table 1. Parameters of the seismic zones of the earthquakes in western Anatolia.

Sub-Regions	Sources D.M.Yr	Completeness Date	Earthquake Number	Depth Model	Gutenberg-Richter	Mmax	Ti
Zone 1	26.0 ≤ Lon. ≤ 28.5 38.8 ≤ Lat. ≤ 40.0	4.4.1903	194	2 ≤ Depth ≤ 80	logN(M)=3.67-0.584M *(0.338) (0.06)*	6.7	103
Zone 2	28.5 ≤ Lon. ≤ 30.0 38.8 ≤ Lat. ≤ 40.0	2.5.1928	248	1 ≤ Depth ≤ 67	logN(M)=4.44-0.722M *(0.409) (0.08)*	6.5	78
Zone 3	26.0 ≤ Lon. ≤ 27.0 38.0 ≤ Lat. ≤ 38.8	17.8.1916	110	2 ≤ Depth ≤ 40	logN(M)=3.57-0.615M *(0.527) (0.10)*	6.2	90
Zone 4	27.0 ≤ Lon. ≤ 30.0 38.0 ≤ Lat. ≤ 38.8	19.5.1904	104	1 ≤ Depth ≤ 88	logN(M)=3.05-0.496M *(0.385) (0.08)*	6.1	102
Zone 5	26.0 ≤ Lon. ≤ 27.7 37.4 ≤ Lat. ≤ 38.0	11.8.1904	64	1 ≤ Depth ≤ 60	logN(M)=2.37-0.404M *(0.467) (0.09)*	6.4	102
Zone 6	27.7 ≤ Lon. ≤ 30.0 37.4 ≤ Lat. ≤ 38.8	23.2.1904	157	2 ≤ Depth ≤ 130	logN(M)=3.47-0.564M *(0.87) (0.17)*	5.7	105
Zone 7	27.4 ≤ Lon. ≤ 29.0 36.5 ≤ Lat. ≤ 37.4	1.5.1920	184	1 ≤ Depth ≤ 157	logN(M)=4.47-0.757M *(0.38) (0.07)*	6.1	86
Zone 8	29.0 ≤ Lon. ≤ 30.0 36.5 ≤ Lat. ≤ 37.4	1.3.1926	80	1 ≤ Depth ≤ 94	logN(M)=3.04-0.498M *(0.52) (0.11)*	5.8	80

been calculated based on these probabilities. The method of moments is used to characterize probability distributions. Moments are useful for providing physically descriptive information about distributions. For example, the zeroth, first, second, and third moments are related to the distribution total probability, mean, variance, and skewness. In addition to their utility as general descriptive measures, moments are also commonly employed for estimating the parameters of a probability distribution. Model parameters are obtained by equating measured moments, determined through numerical integration of earthquake data.

Four advanced distributions, which involve the best approximation to earthquake data, Exponential, Extreme Value Distribution Type 1 (Gumbel), Log-pearson Type 3, Generalised Pareto, are commonly applied to floods, wind speeds, and wave heights. There are many recent developments in the field of regional flood frequency analysis (Stedinger *et al*. 1993; Hosking & Wallis 1997) which may offer attractive alternatives to the traditional models for earthquake frequency analysis. The Anderson-Darling Test is performed to determine the best-fitting distributions at high magnitudes and classify the resulting frequency-magnitude distributions.

The results indicate that more flexible models can replace GR, EXP and other models which are more parsimonious. GP Distribution gives better results for peaks over the threshold earthquake series than EXP Distribution and GUM Distribution gives also better results for annual maximum earthquake data than LP3 Distribution.

The Probability Distribution of Peaks over Threshold Series

The distributions of earthquake magnitudes *m* have long been assumed to follow the power law model given by Gutenberg & Richter (1954)

$$\log_{10}(N) = a - bm \qquad (1)$$

where *a* and *b* are model parameters and *N* is the frequency of occurrence of earthquakes which have exceeded magnitude m.

In any region, occurrence risk in T years of an earthquake with any magnitude M for observation interval of T year is calculated from equation 2 and recurrence period of an earthquake is estimated by equation 3

$$R(M) = 1 - e^{-n(M)T} \qquad (2)$$

$$Q(M) = 1/n(M) \tag{3}$$

In computations, magnitudes of Ms ≥ 5.0 and increment interval of 1 were chosen, and equations 1, 2, and 3 were used. Observational time interval (T_i) has been determined by the completeness condition (Table 1, third column) of each zone. Computations have been performed for decades in the next 100 years. The results are discussed in Section 3.

The Exponential probability density function (pdf) is given by the following equation:

$$f_{Exp}(m) = \lambda e^{[-\lambda(m-\theta)]} \qquad \theta \le m < \infty \ \lambda > 0 \tag{4}$$

where $\lambda = (\overline{m} - \theta)^{-1}$ and *m* is the observed earthquake magnitude in excess of some minimum threshold level θ (Utsu 1999). Mean, variance, skewness, and kurtosis of earthquake magnitudes are given by:

$$\mu_{Exp} = \theta + \left(\frac{1}{\lambda}\right) \quad \text{(5a)} \qquad \sigma^2_{Exp} = \left(\frac{1}{\lambda^2}\right) \quad \text{(5b)}$$

$$\gamma_{Exp} = 2 \quad \text{(5c)} \qquad \kappa_{Exp} = 9 \quad \text{(5d)}$$

Having fixed skewness and kurtosis regarded as the EXP model, this model is rigidly constrained. That partially explains why the EXP model is unable to show the tail behaviour of very large earthquakes in most regions. The EXP pdf is needed to construct probability plots and estimate earthquake magnitudes corresponding to a recurrence interval given by:

$$m_{Exp}(p) = \frac{\lambda\theta - \ln(1-p)}{\lambda} \tag{6}$$

where *p* is the non-exceedance probability given by the cumulative density function (cdf): $p = F_{Exp}(m) = P(M \le m)$.

The Generalized Pareto model (GP) is much more flexible than the EXP model for earthquakes and is given by:

$$f_{GP}(m) = \beta[1 - \kappa\beta(m - m_0)^{\left(\frac{1}{\kappa}\right)-1}$$

for (7)

$$m_0 < m < \infty \quad \beta > 0 \quad \kappa \ne 0$$

where β and κ are scale and shape parameters respectively, and β is the same as λ in the EXP pdf. The GP model has a great advantage over the EXP model for modeling the distribution of earthquake magnitudes because the upper bound is equal to $m_0 + [1/(\beta\kappa)]$. When the shape parameter is positive, its skew is not constant as it is in the case for the GR and EXP models. In general, for the GP model in (7) the bounds on *m* are

$$m_0 \le m \le m_0 + [1/(\beta\kappa)] \text{ for } \kappa > 0 \tag{8a}$$

$$m_0 + [1/(\beta\kappa)] \le m \le \infty \text{ for } \kappa < 0 \tag{8b}$$

Thus an upper bound on the distribution of *m* implies that $\kappa > 0$. For the GP model in (7), Hosking & Wallis (1997) report the mean, variance, skew, and kurtosis as:

$$\mu_{GP} = m_0 + \frac{1}{\beta(1+\kappa)} \tag{9a}$$

$$\sigma^2_{GP} = \frac{1}{\beta^2[(1+\kappa)^2(1+2\kappa)]} \tag{9b}$$

$$\gamma_{GP} = \frac{2(1-\kappa)(1+2\kappa)^{1/2}}{(1+3\kappa)]} \tag{9c}$$

$$\kappa_{GP} = \frac{3(1+2\kappa)(3-\kappa+2\kappa^2)}{[(1+3\kappa)(1+4\kappa)]} - 3 \tag{9d}$$

The quantile function for a GP distribution is given by

$$m_{GP}(p) = \frac{1 + \kappa\beta m_0 - (1-p)^{\kappa}}{\kappa\beta} \tag{10}$$

where $p = F_{GP}(m) = P(M \le m)$. More detailed parameter estimation methods for the GP distribution are provided by Hosking & Wallis (1997) and Rosbjerg *et al.* (1992).

The Probability Distribution of Annual Maximum Series

There is a family of Generalised Extreme Value distributions, each member of which is characterized by the value of a parameter denoted by κ. The family can be divided into three classes corresponding to different ranges of κ values. Negative κ corresponds to Type 2 (Fisher-Tippett); zero κ corresponds to Type 1 (Gumbel), and positive κ corresponds to Type 3 (Weibull).

According to the relationship κ and skewness of the annual maximum magnitude on 106 years, the extreme value Type 1 (GUM) model has been chosen. Its cdf is given by:

$$F_{Gum}(m) = e^{[-e^{(-\lambda(m-\theta))}]} \quad \theta \le m < \infty \tag{11}$$

Here the scale parameter λ has been estimated from the observed earthquake data.

The GUM distribution has mean, variance, skewness, and kurtosis defined as:

$$\mu_{Gum} = \xi + \left(\frac{0.5772}{\beta}\right) \quad (12a)$$

$$\sigma^2_{Gum} = \frac{\pi^2}{6\beta^2} \quad (12b)$$

$$\gamma_{Gum} = 1.1396 \quad (12c)$$

$$\kappa_{Gum} = 5.4 \quad (12d)$$

The most important drawback of using the GUM distribution for modeling annual maximum magnitude series is due to the fixed skew and kurtosis and lack of an upper bound. The quantile function of a GUM pdf is given by:

$$m_{Gum}(p) = \xi - \frac{\ln(-\ln(p))}{\beta} \quad (13)$$

This distribution is widely used for frequency analysis. Due to the popularity of this distribution for floods, a great deal of research into its better utilization has been undertaken including its application to earthquakes (Thompson *et al.* 2007).

The LP3 distribution is widely used for frequency analysis. Unlike the GUM distribution, the cdf of the LP3 distribution cannot be stated in a closed form. The cdf must be stated as the integral of the pdf from negative infinity to the cutoff point. Using the variables, the pdf as:

$$f(m) = [(\ln(m) - \varepsilon)/\alpha]\beta - 1 \quad (14)$$

m is the cutoff value, α is a scale, β is a shape, and ε is a location parameter, $\Gamma(\beta)$ is the Gamma Function evaluated at β.

The LP3 distribution tells the likely values of magnitudes of earthquakes to expect at various recurrence intervals based on the available historical data. It is important to design structures to protect the distribution against the largest expected event and for this reason it is customary to perform the earthquake frequency analysis by using the annual maximum earthquake magnitude data. However, the LP3 distribution can be constructed by using instantaneous peak earthquake data.

The pdf for the LP3 distribution can be used to find equations giving the first three moments as functions of α, β, and ε. By rearranging these equations, the parameters μ_m, σ_m and C_m of the logarithms of *m* can be found as:

$$\alpha = \frac{\sqrt{\beta}}{\sigma_m} \qquad \beta = \left(\frac{2}{C_m}\right)^2 \qquad \varepsilon = \mu_m - \frac{\beta}{\lambda}\alpha \quad (15)$$

where all of the variables are as previously defined.

The quantile function of LP3 distribution is

$$m_{LP3}(p) = 1 - (1 - p)^n \quad (16)$$

where *n* is the number of earthquakes.

Parameter Estimation for Distributions

Many methods are used to estimate the parameters of the various models introduced above. Unfortunately, individual earthquake datasets are often too short to provide reliable estimates of extreme events without additional information. The method of moments (MM) is used to estimate parameters of the four pdfs introduced above. MM estimators are used instead of the maximum likelihood (ML) because a number of studies have shown that MM estimators are more efficient and less biased than ML (Hosking 1990) estimators for the small samples, which are normally encountered in annual maximum magnitude series of earthquakes.

Peaks over Threshold Series Estimators. Assuming that a peaks over threshold series of earthquake magnitudes m_i for $i=1,\dots,n$ is available with $m_i \geq m_0$ for all $i=1,\dots,n$ where *n* is the sample size of the peaks over threshold series. Sample estimates of the mean $\overline{m}$ and standard deviation s_m are given by:

$$\overline{m} = \frac{1}{n}\sum_{i=1}^{n} m_i \quad \text{and} \quad s_m = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(m_i - \overline{m})^2} \quad (17)$$

Solving the expression for the mean of an EXP variable in (3) for λ leads to:

$$\hat{\lambda} = (\overline{m} - \theta)^{-1} \quad \text{and} \quad \hat{b} = \frac{1}{(\overline{m} - \theta)\ln(10)} \quad (18)$$

Combining the moment equation for the mean and variance of a GP variable in equations (7a) and (7b) leads to the following MM sample estimators for κ and β:

$$\hat{\beta} = \frac{1}{(\overline{m} - m_0)(1+\kappa)} \quad \text{and} \quad \hat{\kappa} = \frac{1}{2}\left[\frac{\overline{m} - m_0}{s_m} - 1\right] \quad (19)$$

Seismic risk and return period values for GR model have been estimated by using *a* and *b* parameters given in Table 2 and graphic results are shown in Figure 4. Model

Table 2. Seismic risk and return period values estimated by Gutenberg-Richter models of each zone.

Sub-Regions	Magnitudes (M_s)	Seismic Risk Period (Year) R(M)%										Return Period Q(M)
		10	20	30	40	50	60	70	80	90	100	
Zone 1	5.0	33.4	55.6	70.4	80.3	86.9	91.3	94.2	96.1	97.4	98.3	25
(T_1=103 Years)	6.0	10.0	19.1	27.2	34.5	41.1	47.0	52.3	57.1	61.4	65.3	94
	7.0	2.7	5.4	7.9	10.4	12.8	15.2	17.5	19.7	21.9	24.0	364
Zone 2	5.0	40.6	64.7	78.9	87.5	92.6	95.6	97.4	98.4	99.1	99.5	19
(T_2=78 Years)	6.0	9.4	17.7	25.5	32.4	38.7	44.4	49.6	54.3	58.6	62.5	101
	7.0	1.9	3.6	5.4	7.2	8.9	10.6	12.3	13.9	15.5	17.1	534
Zone 3	5.0	21.7	38.7	52.0	62.5	70.6	77.0	82.0	85.9	88.9	91.3	40
(T_3=90 Years)	6.0	5.7	11.2	16.3	21.1	25.7	30.0	34.1	37.9	41.5	44.8	168
	7.0	1.4	2.8	4.1	5.4	6.8	8.0	9.3	10.6	11.8	13.0	714
Zone 4	5.0	27.1	46.8	61.2	71.7	79.4	84.9	89.1	92.1	94.2	95.8	31
(T_4=102 Years)	6.0	9.6	18.1	25.9	32.9	39.3	45.1	50.3	55.0	59.3	63.2	100
	7.0	3.1	6.2	9.2	12.1	14.9	17.6	20.2	22.7	25.2	27.5	310
Zone 5	5.0	20.9	36.9	49.8	60.1	68.3	74.8	80.0	84.1	87.3	89.9	43
(T_5= 102 Years)	6.0	8.8	16.8	24.1	30.8	36.9	42.4	47.5	52.1	56.3	60.1	108
	7.0	3.5	6.9	10.2	13.4	16.4	19.4	22.2	25.0	27.6	30.2	277
Zone 6	5.0	28.1	48.3	62.8	73.2	80.8	86.2	90.1	92.9	94.9	96.3	30
(T_6= 105 Years)	6.0	8.5	16.3	23.4	29.9	35.9	41.4	46.4	50.9	55.1	58.9	112
	7.0	2.4	4.7	7.1	9.2	11.4	13.6	15.7	17.7	19.7	21.6	409
Zone 7	5.0	27.6	47.3	61.7	72.2	79.8	85.3	89.4	92.2	94.4	95.9	29
(T_7= 86 Years)	6.0	5.4	10.6	15.5	20.0	24.4	28.5	32.4	36.1	39.6	42.9	178
	7.0	0.9	1.9	2.9	3.8	4.8	5.7	6.6	7.5	8.4	8.3	1016
Zone 8	5.0	31.9	53.2	68.0	78.1	85.0	89.8	93.0	95.2	96.7	97.8	26
(T_8= 80 Years)	6.0	11.5	21.6	30.6	38.6	45.7	51.9	57.4	62.3	66.6	70.5	81
	7.0	3.8	7.5	11.0	14.4	17.7	20.8	23.8	26.7	29.5	32.2	257

parameters estimation results and exceedance probabilities are tabulated in Table 3 for EXP distribution and in Table 4 for GP distribution.

Annual Maximum Series Estimators. Consider that an annual maximum series of earthquake magnitudes m_i i=1,…,n is available where n is the number of years in the annual maximum series. m_i is equal to the annual earthquake magnitude in the year, and the annual maximum magnitude series have a lower bound ξ which differs from θ.

For the GUM model, MM estimates of the scale and lower bound parameters are obtained from (12) which leads to:

$$\hat{\beta} = \frac{\pi}{s_m \sqrt{6}} \qquad \hat{\xi} = \overline{m} - \frac{0.5772}{\hat{\beta}} \tag{20}$$

The MM estimate of the shape parameter κ is obtained by the substitution of sample skewness with $\hat{\gamma}_m$ computed from:

$$\hat{\gamma}_m = \frac{\frac{1}{n}\sum_{i=1}^{n}(m_i - \overline{m})^3}{s_m^3} \tag{21}$$

The advantage of the LP3 distribution is that extrapolation can be made of the values for extreme events with return periods well beyond the observed earthquake events. The general equation for LP3 is:

$$\log m = \overline{\log m} + \kappa\sigma_{\log m} \tag{22}$$

where *m* is the observed annual maximum earthquake magnitude, is the average of the *log m*, κ is a frequency factor, and σ is the standard deviation of the *log m* values. The frequency factor κ is a function of the skewness

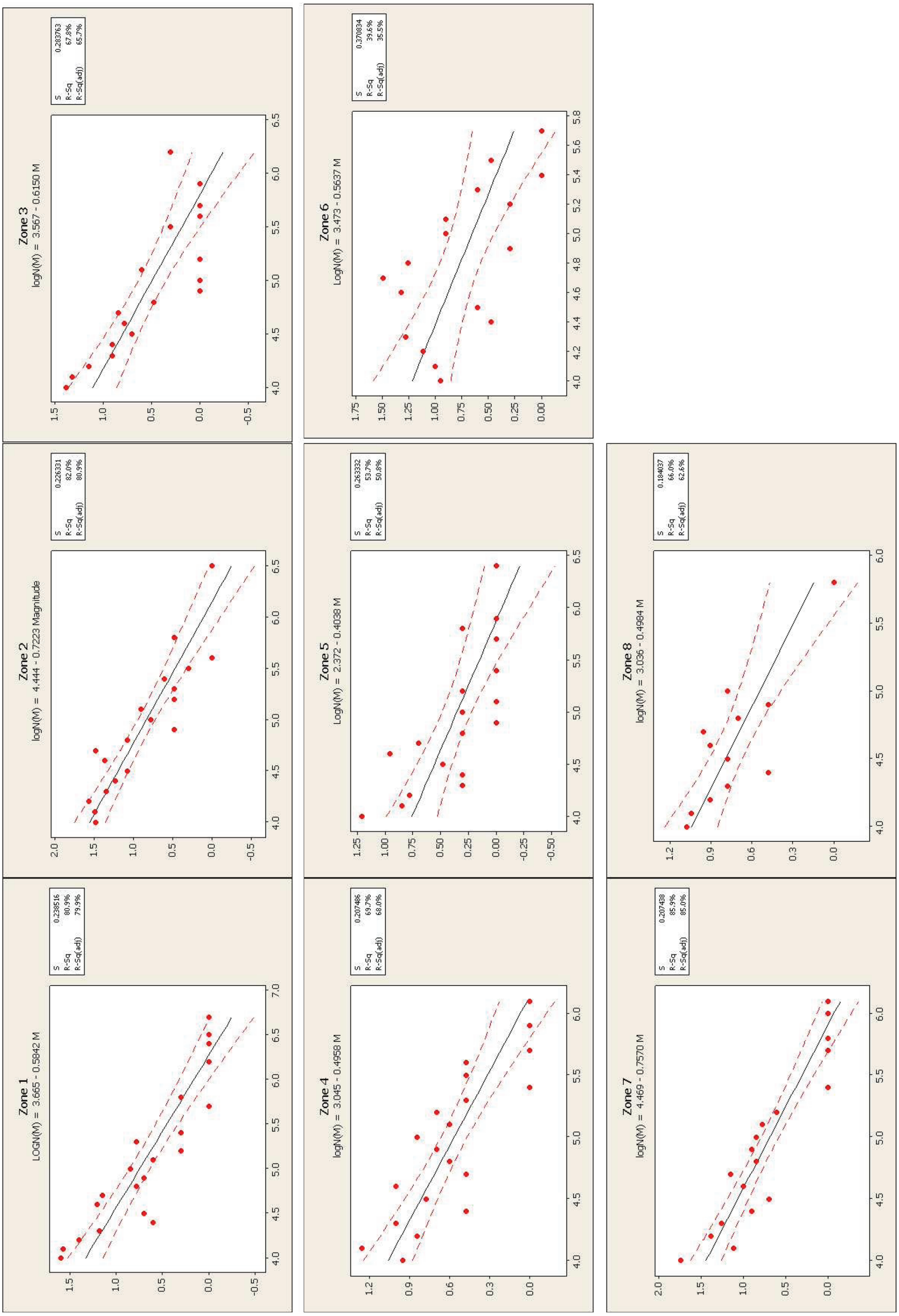

Figure 4. Magnitude-frequency relations computed by GR model for each zone.

Table 3. Exceedance probabilities and return period values estimated by Exponential distribution of each zone.

Exponential Distribution Results													
Zone 1		5.0	5.5	6.0	6.5	7.0	Zone 2		5.0	5.5	6.0	6.5	7.0
$\mu = 5.2$	Exceed. prob.	0.43	0.28	0.18	0.11	0.07	$\mu = 4.9$	Exceed. prob.	0.43	0.29	0.20	0.16	0.13
$\sigma^2 = 1.45$							$\sigma^2 = 0.81$						
$\beta = 0.83$	Return period	5.04	7.64	11.57	17.52	26.53	$\beta = 1.11$	Return period	5.27	9.8	16.00	27.88	48.56
Zone 3		5.0	5.5	6.0	6.5	7.0	Zone 4		5.0	5.5	6.0	6.5	7.0
$\mu = 5.0$	Exceed. prob.	0.36	0.22	0.13	0.08	0.04	$\mu = 5.1$	Exceed. prob.	0.40	0.25	0.16	0.10	0.06
$\sigma^2 = 1$							$\sigma^2 = 1.21$						
$\beta = 1$	Return period	5.17	8.52	14.05	23.16	38.19	$\beta = 0.91$	Return period	5.11	8.05	12.70	20.02	31.56
Zone 5		5.0	5.5	6.0	6.5	7.0	Zone 6		5.0	5.5	6.0	6.5	7.0
$\mu = 4.9$	Exceed. prob.	0.33	0.19	0.11	0.06	0.04	$\mu = 4.9$	Exceed. prob.	0.32	0.17	0.10	0.05	0.03
$\sigma^2 = 0.81$							$\sigma^2 = 0.75$						
$\beta = 1.11$	Return period	5.27	9.18	16.00	27.88	48.56	$\beta = 1.15$	Return period	5.32	9.45	16.81	29.86	53.07
Zone 7		5.0	5.5	6.0	6.5	7.0	Zone 8		5.0	5.5	6.0	6.5	7.0
$\mu = 4.8$	Exceed. prob.	0.27	0.14	0.07	0.04	0.02	$\mu = 4.7$	Exceed. prob.	0.22	0.10	0.05	0.03	0.02
$\sigma^2 = 0.61$							$\sigma^2 = 0.42$						
$\beta = 1.28$	Return period	5.51	10.45	19.81	37.58	71.28	$\beta = 1.54$	Return period	6.08	13.14	28.38	61.31	132.41

Table 4. Exceedance probabilities and return period values estimated by Generalised Pareto distribution of each zone.

Generalised Pareto Distribution Results													
Zone 1		5.0	5.5	6.0	6.5	7.0	Zone 2		5.0	5.5	6.0	6.5	7.0
$\mu = 5.2$	Exceed. prob.	0.51	0.33	0.19	0.10	0.03	$\mu = 4.9$	Exceed. prob.	0.39	0.19	0.06	0.002	0.001
$\sigma^2 = 0.77$							$\sigma^2 = 0.39$						
$\beta = 0.58$	Return period	4.70	6.02	8.06	11.58	19.42	$\beta = 0.73$	Return period	4.22	5.83	9.71	33.62	75.58
$\kappa = 0.44$							$\kappa = 0.52$						
Zone 3		5.0	5.5	6.0	6.5	7.0	Zone 4		5.0	5.5	6.0	6.5	7.0
$\mu = 5.0$	Exceed. prob.	0.38	0.22	0.11	0.04	0.01	$\mu = 5.1$	Exceed. prob.	0.50	0.30	0.15	0.04	0.01
$\sigma^2 = 0.61$							$\sigma^2 = 0.55$						
$\beta = 0.75$	Return period	4.91	7.16	11.29	20.20	46.51	$\beta = 0.57$	Return period	4.48	5.45	7.18	11.52	34.94
$\kappa = 0.32$							$\kappa = 0.59$						
Zone 5		5.0	5.5	6.0	6.5	7.0	Zone 6		5.0	5.5	6.0	6.5	7.0
$\mu = 4.9$	Exceed. prob.	0.34	0.19	0.10	0.05	0.02	$\mu = 4.9$	Exceed. prob.	0.40	0.17	0.09	0.04	0.01
$\sigma^2 = 0.67$							$\sigma^2 = 0.31$						
$\beta = 1.00$	Return period	4.99	8.36	14.46	25.97	48.68	$\beta = 0.66$	Return period	3.88	4.83	9.33	14.78	57.92
$\kappa = 0.10$							$\kappa = 0.74$						
Zone 7		5.0	5.5	6.0	6.5	7.0	Zone 8		5.0	5.5	6.0	6.5	7.0
$\mu = 4.8$	Exceed. prob.	0.31	0.13	0.04	0.005	0.002	$\mu = 4.9$	Exceed. prob.	0.23	0.07	0.01	0.002	0.001
$\sigma^2 = 0.35$							$\sigma^2 = 0.35$						
$\beta = 0.94$	Return period	4.43	7.41	15.24	52.05	77.08	$\beta = 0.66$	Return period	4.58	9.37	29.92	65.32	113.91
$\kappa = 0.36$							$\kappa = 0.74$						

coefficient, and the return period and can be found using the frequency factor table. The annual maximum earthquake magnitudes for the various return periods are found by solving the general equation.

The mean, variance, and standard deviation of the data can be calculated by using the two formulas below:

$$\overline{\log m} = \frac{\sum_{i=1}^{n} (\log m_i)}{n} \tag{23a}$$

$$\sigma^2_{\log m} = \frac{\sum_{i=1}^{n} (\log m_i - \overline{\log m_i})^2}{n-1} \quad (23b)$$

$$\sigma_{\log m} = \sqrt{\frac{\sum_{i=1}^{n} (\log m_i - \overline{\log m_i})^2}{n-1}} \quad (23c)$$

The skewness coefficient C_m can be calculated as follows:

$$C_m = \frac{n \sum_{i=1}^{n} (\log m_i - \overline{\log m_i})^3}{(n-1)(n-2)(\sigma_{\log m})^3} \quad (24)$$

where $\sigma_{\log m}$ is the standard deviation.

Model parameters estimation results and exceedance probabilities are tabulated in Table 5 for GUM distribution and in Table 6 for LP3 distribution.

Plotting Position of Distributions

Plotting position-based methods are probably the most straightforward of the various methods. For each observed magnitude a plotting position formula is used to estimate an exceedance probability. A curve is then fitted through the data points.

A probability plot is built by plotting the ordered observations $m_{(i)}$, $i=1,\ldots,n$ versus an estimate of the ordered observation based on the assumed distribution that is obtained from the quantile function for that distribution $m(p_i)$, where p_i is a suitable plotting position. The plotting position for EXP, GP, and LP3 is obtained from the formulation given by:

$$p_i = \frac{i-0.4}{n+0.2} \quad (25)$$

For the GUM distribution, the plotting position introduced by Gringortens (1963) is given by:

$$p_i = \frac{i-0.44}{n+0.12} \quad (26)$$

The quantile functions of distributions are shown in Figure 5.

Probability Plot Goodness of Fit for Regional Earthquake Distributions. Probability plots are a useful graphical tool to evaluate the goodness-of-fit of alternative pdfs to earthquake data. Probability plots for western Anatolia earthquake data are shown in Figure 6 (see Figure 3 for location map). Figure 6 compares the EXP and GP pdfs to the observed peaks over the threshold series and also compares the GUM and LP3 pdfs to the observed annual maximum series for each zone.

Anderson Darling (AD) Evaluation. The Anderson-Darling (AD) Test is devised to give heavier weightings to the tails of the distribution where unexpectedly high or low values, called outliers, are located (Kottegoda & Rosso 1997). Of the many quantitative goodness-of-fit techniques (e.g., Kolmogorov-Smirnov(K-S), Shapiro-Wilk, von Mises, etc.), the AD Test (Stephens 1974) is preferred because it is a modification of the K-S test and gives more weight to the tails of the distribution than the older K-S test. The AD test is used to test if a sample of data came from a population with a specific distribution.

Table 5. Exceedance probabilities and return period values estimated by Gumbel distribution of each zone.

Extreme Value Distribution Type 1 (Gumbel) Results													
Zone 1		5.0	5.5	6.0	6.5	7.0	**Zone 2**		5.0	5.5	6.0	6.5	7.0
μ = 5.10	Exceed. Prob.	0.48	0.29	0.17	0.06	0.03	μ = 4.9	Exceed. Prob.	0.36	0.16	0.09	0.03	0.01
σ^2 = 0.69	Return Period	2.88	6.63	15.27	35.15	80.90	σ^2 = 0.46	Return Period	3.43	7.84	17.9	41.1	93.9
Zone 3		5.0	5.5	6.0	6.5	7.0	**Zone 4**		5.0	5.5	6.0	6.5	7.0
μ = 4.8	Exceed. Prob.	0.29	0.16	0.11	0.04	0.01	μ = 5.0	Exceed. Prob.	0.43	0.18	0.06	0.02	0.009
σ^2 = 0.50	Return Period	3.27	6.83	14.25	29.75	62.09	σ^2 = 0.39	Return Period	2.82	7.38	19.32	50.6	132.4
Zone 5		5.0	5.5	6.0	6.5	7.0	**Zone 6**		5.0	5.5	6.0	6.5	7.0
μ = 4.9	Exceed. Prob.	0.42	0.23	0.10	0.04	0.01	μ = 4.8	Exceed. Prob.	0.29	0.10	0.03	0.01	0.003
σ^2 = 0.52	Return Period	3.84	9.1	21.28	50.1	117.8	σ^2 = 0.54	Return Period	3.98	15.48	60.25	234	912
Zone 7		5.0	5.5	6.0	6.5	7.0	**Zone 8**		5.0	5.5	6.0	6.5	7.0
μ = 4.9	Exceed. Prob.	0.36	0.18	0.06	0.02	0.009	μ = 4.6	Exceed. Prob.	0.18	0.06	0.02	0.004	0.001
σ^2 = 0.43	Return Period	4.57	15.67	55.55	184.1	666	σ^2 = 0.26	Return Period	6.61	24.83	93.3	350.75	1318.2

Table 6. Exceedance probabilities and return period values estimated by Log Pearson Type 3 distribution of each zone.

Log Pearson Type 3 Distribution Results													
Zone 1		**5.0**	**5.5**	**6.0**	**6.5**	**7.0**	**Zone 2**		**5.0**	**5.5**	**6.0**	**6.5**	**7.0**
$\mu = 5.1$	Exceed. Prob.	0.53	0.25	0.15	0.06	0.03	$\mu = 5.0$	Exceed. Prob.	0.42	0.14	0.06	0.04	0.02
$\sigma^2 = 1.12$							$\sigma^2 = 1.07$						
$\kappa = 0.75$	Return period	1.88	4.00	6.4	16.00	32.00	$\kappa = 1.03$	Return period	2.38	7.14	16.67	25.00	50.00
Zone 3		**5.0**	**5.5**	**6.0**	**6.5**	**7.0**	**Zone 4**		**5.0**	**5.5**	**6.0**	**6.5**	**7.0**
$\mu = 5.0$	Exceed. Prob.	0.46	0.28	0.17	0.07	0.03	$\mu = 5.1$	Exceed. Prob.	0.46	0.25	0.12	0.06	0.03
$\sigma^2 = 1.14$							$\sigma^2 = 1.12$						
$\kappa = 0.81$	Return period	2.15	3.5	5.6	14.00	28.01	$\kappa = 0.62$	Return period	2.13	4.00	8.00	16.00	32.00
Zone 5		**5.0**	**5.5**	**6.0**	**6.5**	**7.0**	**Zone 6**		**5.0**	**5.5**	**6.0**	**6.5**	**7.0**
$\mu = 5.1$	Exceed. Prob.	0.45	0.36	0.18	0.09	0.04	$\mu = 4.9$	Exceed. Prob.	0.36	0.14	0.06	0.04	0.02
$\sigma^2 = 1.18$							$\sigma^2 = 1.10$						
$\kappa = 0.33$	Return period	2.2	2.75	5.5	11.1	22.00	$\kappa = 1.30$	Return period	2.72	7.00	16.33	24.5	49.00
Zone 7		**5.0**	**5.5**	**6.0**	**6.5**	**7.0**	**Zone 8**		**5.0**	**5.5**	**6.0**	**6.5**	**7.0**
$\mu = 4.8$	Exceed. Prob.	0.27	0.12	0.07	0.03	0.01	$\mu = 4.9$	Exceed. Prob.	0.28	0.14	0.10	0.07	0.03
$\sigma^2 = 1.11$							$\sigma^2 = 0.35$						
$\kappa = 1.86$	Return period	3.66	7.85	13.75	27.5	55.1	$\beta = 0.66$	Return period	3.5	7.00	9.33	14.00	28.1
							$\kappa = 0.74$						

AD can also be applied to any distribution. The AD test is an alternative to the chi-square and K-S goodness-of-fit tests. The AD test makes use of the specific distribution in calculating critical values and this has the advantage of allowing a more sensitive test. The disadvantage is that critical values must be calculated for each distribution. The critical values for the AD test depend upon the specific distribution being tested. Tabulated values and formulas have been published by D'Agostino & Stephens (1986) for a few specific distributions such as normal, lognormal, exponential, weibull, logistic, gumbel, double exponential, uniform, gamma, and cauchy.

In this study, the critical values of EXP and GUM distributions have been obtained from D'Agostino & Stephens (1986). Choulakian & Stephens (2001) applied goodness-of-fit tests for the GP distribution based on Cramer-von Mises statistic and Anderson-Darling statistic (A^2). In the absence of such tests, the authors used to test exponentiality. Upper-tail asymptotic percentage points of the A^2 for different values of shape parameter (for values of κ between –0.9 and 0.5) were given in the aforementioned article. For further information Choulakian & Stephens (2001) can be checked.

In order to apply the AD test for LP3, a transformation that produces a new test statistic independent of distribution has been used. This transformation was proposed by Laio (2004).

The null hypothesis is H_o: The random sample $X_1, \ldots, X_n$ comes from the specific distribution

In order to implement the AD test, we follow a well-defined series of steps:

(1) Estimate the parameters from the sample data.

(2) Sort the data in ascending order for any distribution and calculate A^2 from equation 27. These steps are the same for all distributions. The AD test statistic is defined as:

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\ln F(x_i) + \ln\left(1 - F(x_{n+1-i})\right)\right] \tag{27}$$

where $F(x_i)$ is the cumulative distribution function of the specific distribution.

(3) For a few specific distributions that are mentioned above, compare AD test statistics with critical values for each distribution.

(4) The null hypothesis that the distribution is of specific form is rejected if the test statistic A^2 is greater than the critical value.

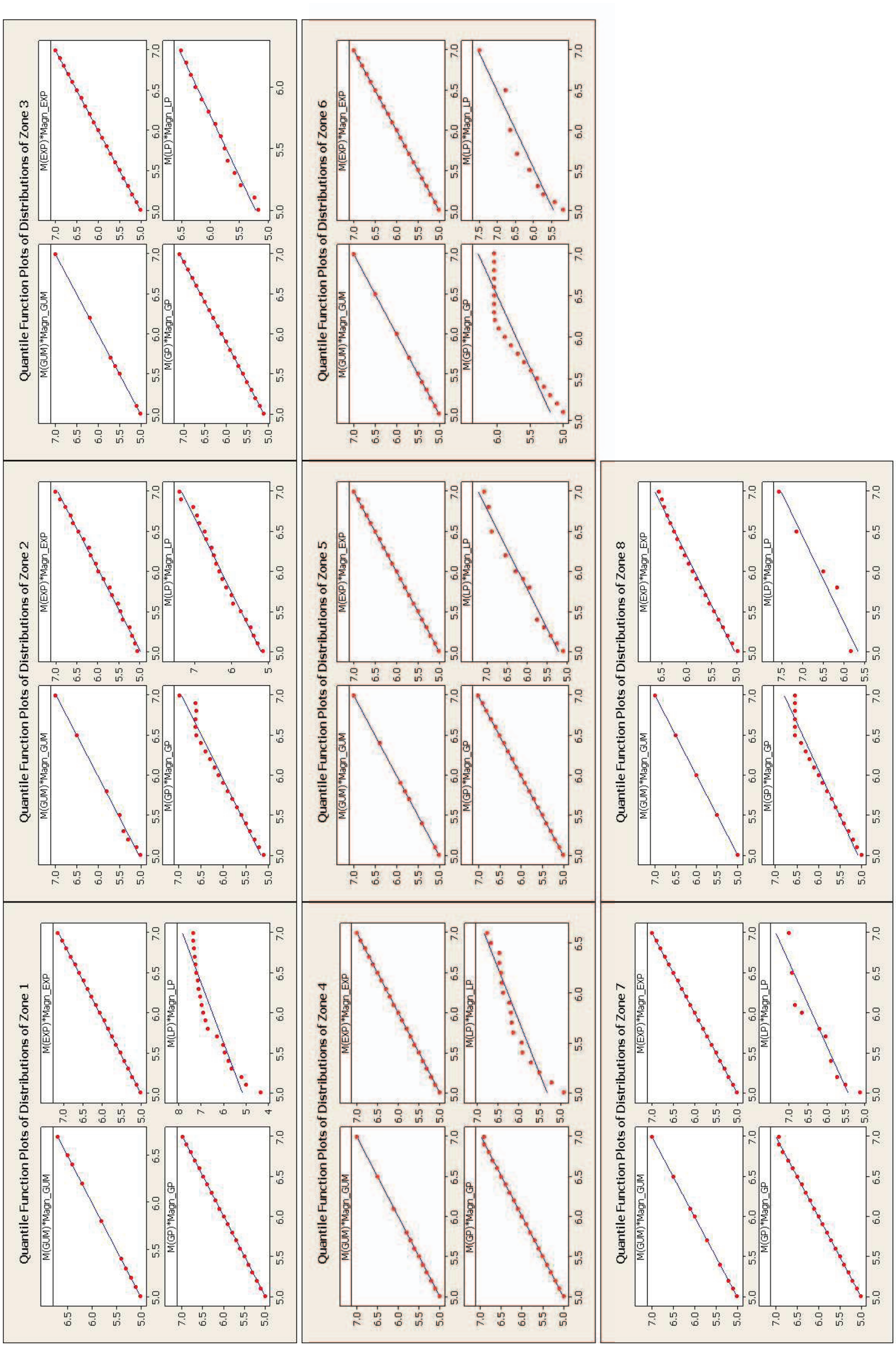

Figure 5 . Quantile function plots of distributions for each zone.

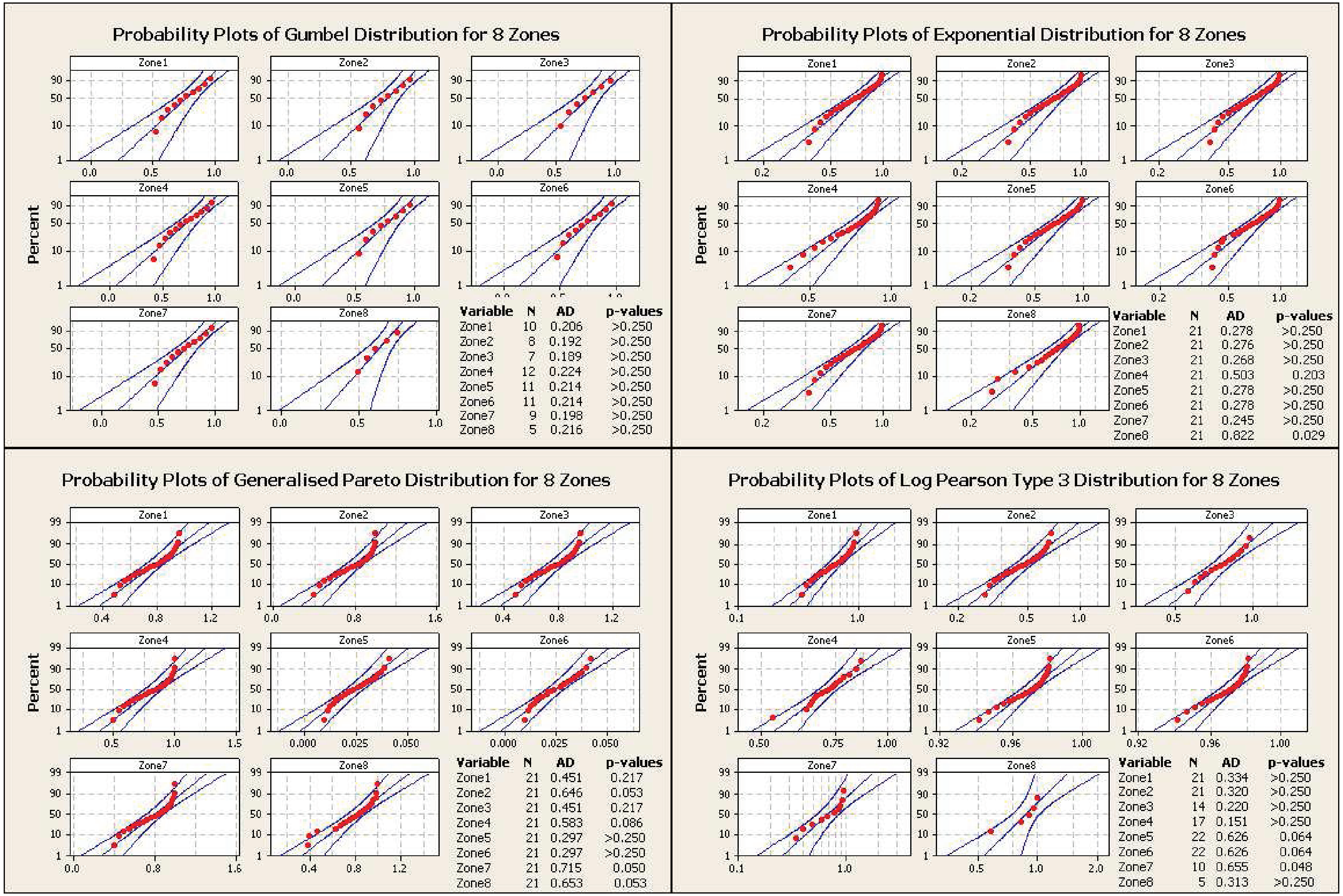

Figure 6. Probability plots of distributions for each zone.

The following steps are applicable for other distributions whose tabulated values are not directly available.

(5) Determine the coefficients when the parameters of the hypothetical distribution are unknown by equation 28 with the appropriate asymptotic estimator of the shape parameter (κ) of the distribution (Laio 2004). In the equation 28, ξ, β and η are respectively location, scale, and shape parameters.

$$\xi = 0.145\,(1 + 0.17\hat{\kappa}^{-1} + 0.333\hat{\kappa}^{-2}) \quad (28a)$$

$$\beta = 0.186\,(1 + 0.34\hat{\kappa}^{-1} + 0.30\hat{\kappa}^{-2}) \quad (28b)$$

$$\eta = 1.194\,(1 - 0.004\hat{\kappa}^{-1} + 0.12\hat{\kappa}^{-2}) \quad (28c)$$

(6) Find ω from equation 29 by the coefficients when the parameters of the hypothetical distribution are known (Laio 2004).

$$\omega = \beta_0 \left(\frac{Q_u^2 - \xi_u}{\beta_u} \right)^{\frac{\eta_u}{\eta_0}} + \xi_0 \;, \qquad 1.2\xi_u \le Q_u^2 \quad (29)$$

where ξ_u, β_u and η_u are respectively coefficients of location, scale and shape parameters when parameters of hypothetical distribution are completely or partially unspecified. ξ_0, β_0 and η_0 are respectively coefficients of location, scale, and shape parameters when parameters of the hypothetical distribution are completely specified.

With the transformation z=F(x), the general quadratic statistic Q^2 in equation 29 is $Q^2 = n\int_0^1 (F_n(z) - z)^2\,\psi(z)dF(z)$ where $F_n(z)$ is the empirical distribution function (EDF) of the variable z. Q^2_u is a quadratic statistic when parameters of the hypothetical distribution are completely or partially unspecified.

(7) Compare ω with the appropriate percentage points for the selected significance level.

(8) The null hypothesis is rejected if the ω is greater than the critical value.

Anderson-Darling test results for each distribution as shown in Figure 6 and Table 7 give brief information about the hypothesis testing process. Anderson-Darling test statistics (A^2) are given in column 3 and 9 for each zone and each distribution in Table 7. These results (col. 3 and 9) have been obtained using equation 27. Column 4 and 10 give the critical values for each distribution. Critical values for EXP and GUM and LP3 have been obtained from D'Agostino & Stephens (1986). Due to the transformation on LP3, a single table is sufficient for carrying out the test containing the values of the three coefficients of the transformation for any distribution of interest (Laio 2004). Only for LP3, ω is compared with critical values. Tabulated values for the statistics A^2 of GP are obtained according toin Choulakian & Stephens (2001).

Results and Comparisons of the Distributions of Earthquake Magnitudes

In this study the linear least square method has been applied to obtain *a* and *b* parameters for each zone shown in Figure 3 using the earthquakes of $M_s \geq 4.0$ occurred from 1.1.1900 to the end of 2006.

According to Table 1 results, in two of the highest and lowest *b* values are calculated as 0.757 (± 0.07), 0.722 (± 0.08) for Zone 7, 2 and as 0.404 (± 0.09), 0.496 (± 0.08) for Zone 8, 4, respectively. As it is well-known, a high *b* value implies that the high seismic activity had rolled in that region. In general, low *b* values are related to high stress-drop; high *b* values are related to high heterogeneity of material and crack density (Weeks *et al*. 1978; Urbancic *et al*. 1992; Wiemer & Katsumada 1999).

Figure 4 shows the magnitude-frequency relations. The estimated relation is given by the confidence interval band of 95%, S (standard deviation) and R^2 (determination coefficient), respectively. Zones 1, 2, and 7 have better magnitude-frequency relations than the other zones.

According to the seismic risk estimations in Table 2, the highest earthquake occurrence probability of $M_s \geq 7.0$ in the next 100 years is 32.2% (S= 0.184, R^2= 66) for Zone 8 and 30.2% (S= 0.2633, R^2= 54) for Zone 5. Return periods for the earthquakes with the same magnitude have been found as 257 and 277 years in these zones, respectively.

The quantile function plots (in Figure 5) show that GP has better convergences properties than the other distributions for each zone. In Figure 6, the peaks over the threshold series plots and annual maximum series plots are graphically shown, and AD values quantitatively confirm that the GP pdf fits the peaks over the threshold series

Table 7. Hypothesis testing results for each distribution and each zone.

		Anderson-Darling (A^2)	Critical Values ($\alpha = 0.05$)	p-values	Decision			Anderson-Darling (A^2)	Critical Values ($\alpha = 0.05$)	p-values	Decision
	EXP	0.278	1.321[a]	> 0.250	Fail to reject H_0		EXP	0.276	1.321[a]	> 0.250	Fail to reject H_0
Zone 1	GUM	0.206	0.757[b]	> 0.250	Fail to reject H_0	Zone 2	GUM	0.192	0.757[b]	> 0.250	Fail to reject H_0
	LP3	0.334	2.492[c]	>0.250	Fail to reject H_0		LP3	0.320	2.492[c]	>0.250	Fail to reject H_0
		($\omega = 0.403$)			($\omega = 0.403 < 2.492$)			($\omega = 0.318$)			($\omega = 0.318 < 2.492$)
	GP	0.451	2.492[d]	0.217	Fail to reject H_0		GP	0.646	2.492[d]	0.053	Fail to reject H_0
	EXP	0.268	1.321[a]	> 0.250	Fail to reject H_0		EXP	0.503	1.321[a]	0.203	Fail to reject H_0
Zone 3	GUM	0.189	0.757[b]	> 0.250	Fail to reject H_0	Zone 4	GUM	0.224	0.757[b]	> 0.250	Fail to reject H_0
	LP3	0.220	2.492[c]	>0.250	Fail to reject H_0		LP3	0.151	2.492[c]	>0.250	Fail to reject H_0
		($\omega = 0.051$)			($\omega = 0.051 < 2.492$)			($\omega = 0.204$)			($\omega = 0.204 < 2.492$)
	GP	0.451	1.651[d]	> 0.217	Fail to reject H_0		GP	0.583	2.492[d]	0.086	Fail to reject H_0
	EXP	0.278	1.321[a]	> 0.250	Fail to reject H_0		EXP	0.278	1.321[a]	> 0.250	Fail to reject H_0
Zone 5	GUM	0.214	0.757[b]	> 0.250	Fail to reject H_0	Zone 6	GUM	0.214	0.757[b]	> 0.250	Fail to reject H_0
	LP3	0.626	2.492[c]	0.064	Fail to reject H_0		LP3	0.626	2.492[c]	0.064	Fail to reject H_0
		($\omega = 0.339$)			($\omega = 0.339 < 2.492$)			($\omega = 0{,}271$)			($\omega = 0.271 < 2.492$)
	GP	0.297	1.321[d]	> 0.250	Fail to reject H_0		GP	0.297	2.492[d]	> 0.250	Fail to reject H_0
	EXP	0.245	1.321[a]	> 0.250	Fail to reject H_0		EXP	0.822	1.321[a]	0.059	Fail to reject H_0
Zone 7	GUM	0.198	0.757[b]	> 0.250	Fail to reject H_0	Zone 8	GUM	0.216	0.757[b]	> 0.250	Fail to reject H_0
	LP3	0.655	2.492[c]	0.048	Reject H_0		LP3	0.313	2.492[c]	>0.250	Fail to reject H_0
		($\omega = 2.574$)			($\omega = 2.574 > 2.492$)			($\omega = 0.083$)			($\omega = 0.083 < 2.492$)
	GP	0.715	1.651[d]	0.050	Fail to reject H_0		GP	0.653	2.492[d]	0.053	Fail to reject H_0

Tabulated values from D'Agostino & Stephens (1986) [a] Table 4.14, p. 138, [b] Table 4.17, p. 146, [c] Table 4.2, p.105
Tabulated values from Choulakian & Stephens (2001) [d] Table 1, p.480

observations better than the EXP, and GUM pdf fits the annual maximum series better than LP3 pdf for each zone. Two distributions of four advanced distributions performed poorly. It is found that the LP3 distribution fits the low magnitude earthquake data well.

According to hypothesis testing results in Table 7, the null hypotheses can not be rejected with large p-values of EXP distribution for each zone except for Zone 8. Since the p-value is very near 0.05, this distribution may not be appropriate for Zone 8. The hypotheses that the data come from the GUM distribution cannot be rejected with large p-values for each zone. The general rule is that a small p-value is evidence against the null hypothesis whilst a large p-value means little or no evidence against the null hypothesis. Hence we have little or no evidence against the null hypothesis for the GUM distribution. Zones 1, 2, 3, 4, and 8 have great p-values for LP3 distribution while Zones 5 and 6 have p-values very close to 0.05. The LP3 distribution does not fit statistically in Zone 7 as well and similar results can be interpreted for GP. Zones 1, 3, 5, and 6 have large p-values for the GP distribution while Zones 2, 4, 7, and 6 have p-values very close to 0.05. These results indicate that the GP distribution may be inappropriate for these zones.

From the results of both the quantitative goodness-of-fit test and the quantile estimation, in comparison with POT and AM series, the GP distribution can be concluded as the best distribution for describing POT series in western Anatolia. GP shows very good descriptive and predictive abilities and is robust for estimating the quantiles and this is apparent in the box plots of exceedance probabilities whilst LP3 underestimates the exceedance probabilities.

Figure 7 shows the return periods of various earthquake magnitudes for each zone. According to this figure; the most realistic result belongs to GP model. Figure 8 shows box plots of exceedance probabilities computed from the four distributions for each zone. As shown in Figure 8, GP is more convenient for large magnitude earthquake data. The variation of the GP distribution of the earthquake data is smaller than the others.

Model parameter estimation results and exceedance probabilities are tabulated in Table 3 for EXP, in Table 4 for GP, in Table 5 for GUM, and in Table 6 for LP3. According to the GP result for Zone 1, the probability of occurrence earthquake greater than 7.0 is 0.03 while EXP is 0.07; GUM is 0.06 and LP3 is 0.03. The same earthquake magnitude in Zone 1 can return approximately in 20 years while EXP is in 27, GUM is in 80, and LP3 is in 32.

Conclusion

The field of earthquake frequency analysis has evolved from the usage of the EXP and GUM models of peaks over the threshold and annual maximum series, respectively. Although the limitations of the GR and EXP models have been discussed at length in the seismic literature, there are still few alternative probabilistic approaches that have been advanced. The recent developments in regional earthquake frequency analysis are shown in this study. The behaviour of the upper tail of the distribution is poorly resolved in seismology because there are few records of large events. Further, physical models predict an upper bound for earthquakes, which is inconsistent with more parsimonious and commonly used probability distributions that are unbounded above (i.e. GR, EXP and LP3). GP and GUM can exhibit an upper bound.

Probability plots and probability plot hypothesis tests (Anderson-Darling) are employed to document that the GP pdf and the GUM pdf provide a much better fit to observed peaks over the threshold earthquake series and annual maximum earthquake series than either EXP or LP3 pdfs, which are still in common use.

This study describes the estimation of parameters using the four probabilistic models for several sets of data. One of the most important results of this study is that the probability of occurrence of a great earthquake in some zones reaches 0.13 for the time period of 49 years in the western Anatolia. It is concluded that the probabilities of some zones being hit by a great earthquake in the foreseeable future are high. For western Anatolia, time has already passed for the occurrence of a great earthquake with high probability, and it is inferred that it may be due to the frequent occurrences of moderate size earthquakes in the area.

Further research, physical models of the mean or maximum earthquake, could be used to constrain GP and GUM pdfs. The same models of seismic moment rate that are used to truncate the GR model can be applied to the distributions introduced here.

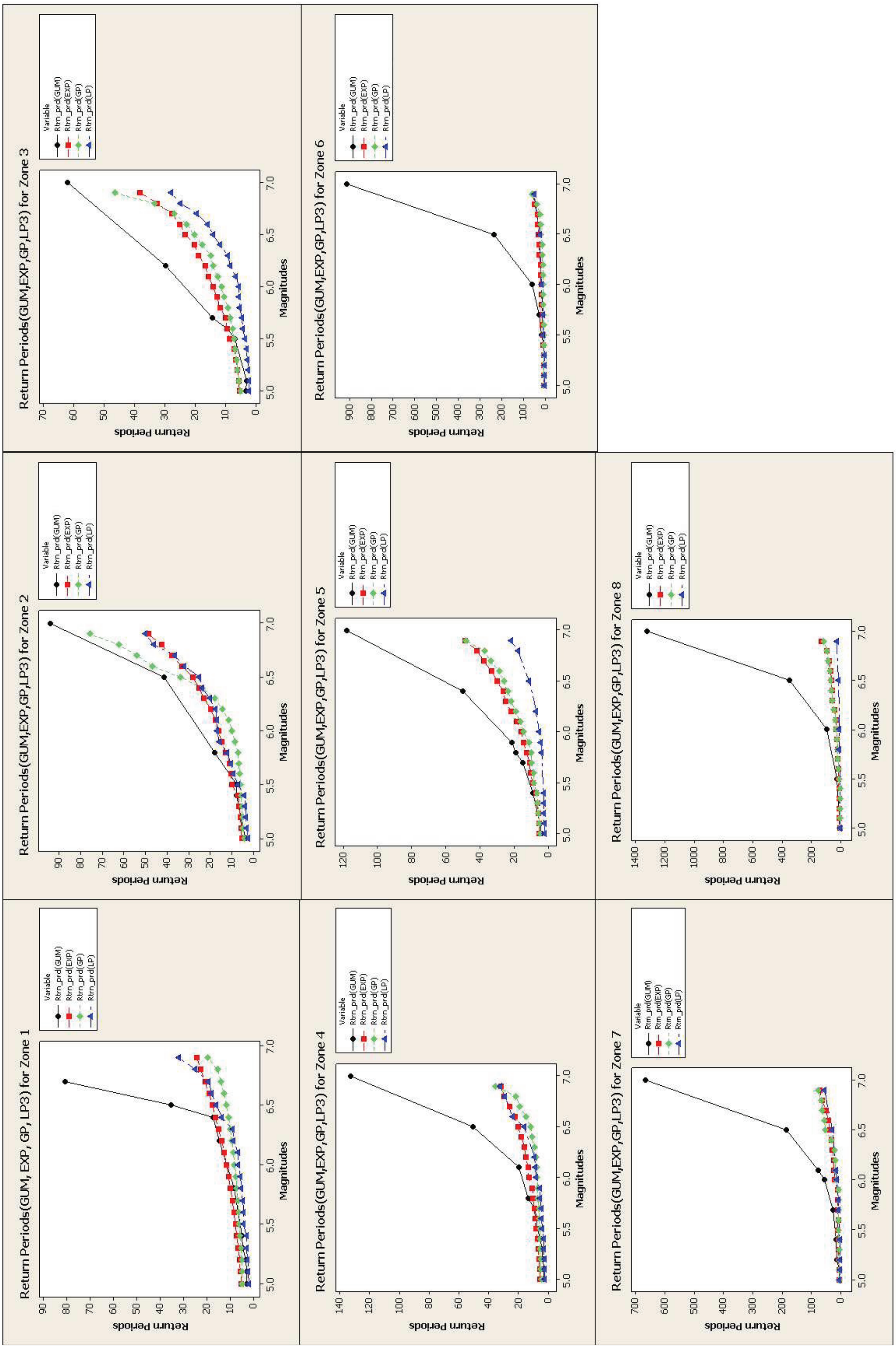

Figure 7. Return periods of distributions for each zone.

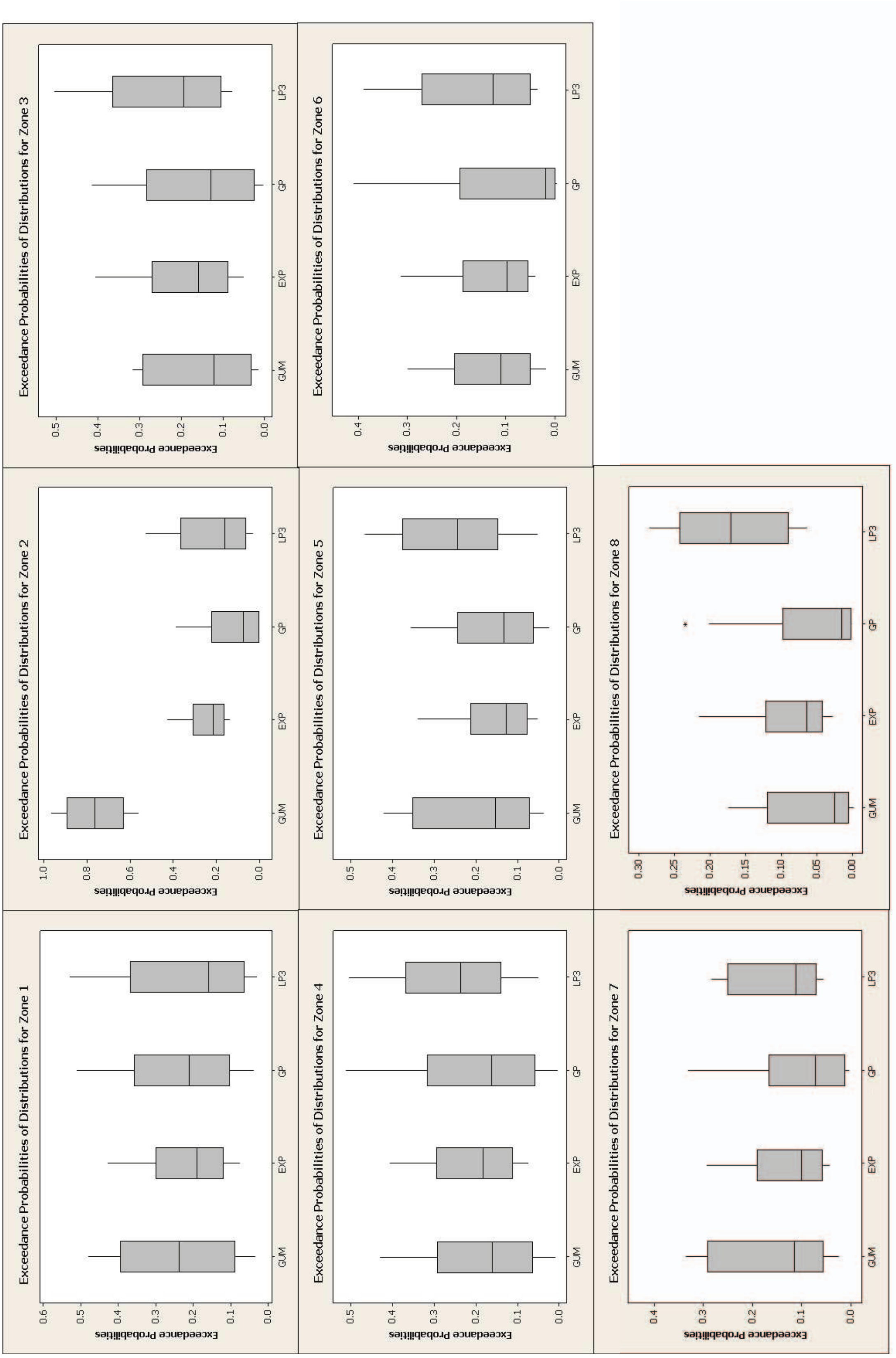

Figure 8. Exceedance probabilities of distributions each zone.

Acknowledgements

The author would like to thank two reviewers, Statis Stiros and Eric Thompson, for their careful reading of the manuscript and their valuable suggestions. The author thanks Hasan Sözbilir for his valuable contribution in technical drawing of Figure 2. The author is grateful to Orhan Polat who helped in preparation of Figure 3 and his cooperation in zonation of western Anatolia. Melanie Brooks and John A. Piper edited the English of the final text.

References

Akyol, N., Zhu, L., Mitchell, B.J., Sözbilir, H. & Kekovali, K. 2006. Crustal structure and local seismicity in western Anatolia. *Geophysical Journal International* **166**, 1259–1269.

Barka, A.A. & Reilinger, R. 1997. Active tectonics of the Mediterranean region: deduced from GPS, neotectonic and seismicity data. *Annali di Geophisica* **XI**, 587–610.

Bozkurt, E. 2001. Neotectonics of Turkey–a synthesis. *Geodinamica Acta* **14**, 3–30

Bozkurt, E. & Sözbilir, H. 2004. Tectonic evolution of the Gediz Graben: field evidence for an episodic, two-stage extension in western Turkey. *Geological Magazine* **141**, 63–79.

Choulakian, V. & Stephens, M.A. 2001. Goodness-of-fit tests for the generalized pareto distribution. *Technometrics* **43**, 478–484.

D'agostino, R.B. & Stephens, M.A. 1986. *Goodness-of-Fit Techniques.* Marcel Dekker, New York.

Dargahi-Noubary, G.R. 1986. A method for predicting future large earthquakes using extreme order statistics. *Physics of the Earth and Planetary International* **42**, 241–245.

Emre, Ö., Özalp, S., Doğan, A., Özaksoy, V., Yildirim, C. & Göktaş, F. 2005. *Active Faults in Around İzmir and Potential of Earthquake Occurrence*. Mineral Research and Exploration Institute (MTA) of Turkey Report no. **10754** [in Turkish, unpublished].

Gringortens, I.I. 1963. A plotting rule for extreme probability paper. *Journal of Geophysical Researh* **68**, 813–814.

Gutenberg, B. & Richter, C.F. 1954. *Seismicity of the Earth*. Princeton University Press, Princeton, New Jersy.

Hosking, J.R.M. 1990. L-Moments: analysis and estimation of distributions using linear combinations of order Statistics. *Journal of the Royal Statistical Society Series B* **52**,105–124.

Hosking, J.R.M. & Wallis, J.R. 1997. *Regional Frequency Analysis – An Approach Based on L-Moments*. Cambridge University Press.

Kagan, Y.Y. 1993. Statistics of characteristic earthquakes. *Bulletin of Seismological Society of America* **83**, 7–24.

Kiratzi, A.A. & Louvari, E. K. 2001. On the active tectonics of the Aegean Sea and the surrounding lands. *In*: Taymaz, T. (ed), *Proceedings of Symposia on Seismotectonics of the North-Western Anatolia-Aegean and Recent Turkish Earthquakes*. İstanbul Technical University, Turkey, 88–95.

Kottegoda, N.T. & Rosso, R. 1997. *Statistics, Probability and Reliability for Civil and Environmental Engineers.* McGraw-Hill.

Laio, F. 2004. Cramer-von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters. *Water Resources Research* **40**, W09308, doi: 10.1029/2004WR003204

Main, I. 1996. Statistical physics seismogenesis, and seismic hazard. *Review of Geophysics* **34**, 433–462.

Ocakoğlu, N., Demirbağ, E. & Kuşçu, İ. 2004. Neotectonic structures in the area offshore of Alaçatı, Doğanbey and Kuşadası (western Turkey): evidence of strike-slip faulting in Aegean Province. *Tectonophsyics* **391**, 67–83.

Ocakoğlu, N., Demirbağ, E. & Kuşçu, İ. 2005. Neotectonic structures in İzmir Gulf and surrounding regions (western Turkey): evidences of strike-slip faulting with compression in the Aegean extensional regime. *Marine Geology* **219**, 155–171.

Öncel, A.O. & Wilson, T. 2004. Correlation of seismotectonic variables and GPS strain measurements in western Turkey. *Journal of Geophsical Research* **109**, B11306.

Papazachos, B.C. 1990. Seismicity of the Aegean and surrounding area. *Tectonophysics* **178**, 287–308.

Papazachos, B.C. & Kiratzi, A.A. 1996. A detailed study of the active crustal deformation in the Aegean and surrounding area. *Tectonophysics* **253**, 129–153.

Rosbjerg, D., Madsen, H. & Rasmussen, P.F. 1992. Prediction in partial duration series with generalized pareto-distributed exceedances. *Water Resources Research* **28**, 3001–3010.

Sari, C. & Şalk, M. 2006. Sediment thicknesses of the western Anatolia graben structures determined by 2D and 3D analysis using gravity data. *Journal of Asian Earth Sciences* **26**, 39–48.

Şaroğlu, E., Emre, Ö. & Kuşçu, İ. 1992. *Active Fault Map of Turkey*. Mineral Research and Exploration Institute (MTA) of Turkey Publications.

Sayil, N. & Osmanşahin, İ. 2007. An investigation of seismicity for western Anatolia. *Natural Hazards.* DOI 10.1007/s11069-007-9141-2.

Şengör, A.M.C., Görür, N. & Şaroğlu, F. 1985. Strike-slip faulting and related basin formation in zones of tectonic escape: Turkey as a case study. *In*: Biddle, K. & Christie-Blick, N. (eds), *Strike-Slip Deformation, Basin Formation and Sedimentation*. Society of Economic Paleontologists and Mineralogists, Special Publications **37**, 227–264.

Stedinger, J.R., Vogel, R.M. & Foufoula-Georgiou, E. 1993. Frequency analysis of extreme events. *In*: Maidment, D.A. (ed), *Handbook of Hydrology.* Chapter 18, Mcgraw-Hill, New York.

STEPHENS, M.A. 1974. EDF statistics for goodness-of-fit and some comparisons. *Journal of the American Statistical Association* **69**, 730–737.

THOMPSON, M.E., BAISE, G.L. & VOGEL, R.M. 2007. A global index earthquake approach to probabilistic assessment of extremes. *Journal of Geophysical Research* **112**, B06314.

URBANCIC, T.I., TRIFU, C.I. & LONG, J.M. 1992. Space-time correlations of b value with stress changes associated with the 1999 November 12, Düzce (Turkey) eartquake (Mw= 7.1). *Geophysical Journal International* **153**, 229–241.

UTSU, T. 1999. Representation and analysis of the earthquake size distribution: a historical review and some new approaches. *Pure and Applied Geophysics* **155**, 509–535.

WEEKS, J., LOCKNER, D. & BYERLEE, J. 1978. Change in b-values during movement on cut surfaces in granite. *Bulletin of Seismological Society of America* **68**, 333–341.

WESSEL, P. & SMITH, W.H.F. 1995. New version of the generic mapping tools (GMT). *EOS Transactions* **76**, p. 329

WIEMER, S. & KATSUMADA, K. 1999. Spatial variability of seismicity parameters in aftershock zones. *Journal of Geophysical Research* **104**, 13135–13151.

Received 18 January 2007; revised typescript received 11 January 2008; accepted 29 January 2008